EAI/Springer Innovations in Communication and Computing

Zoltán Balogh *Editor*

The 16th International Scientific Conference on Distance Learning in Applied Informatics

DIVAI 2026

EAI
RESEARCH MEETS INNOVATION

Springer

EAI/Springer Innovations in Communication and Computing

Series Editor

Imrich Chlamtac, *European Alliance for Innovation, Ghent, Belgium*

The impact of information technologies is creating a new world yet not fully understood. The extent and speed of economic, life style and social changes already perceived in everyday life is hard to estimate without understanding the technological driving forces behind it. This series presents contributed volumes featuring the latest research and development in the various information engineering technologies that play a key role in this process. The range of topics, focusing primarily on communications and computing engineering include, but are not limited to, wireless networks; mobile communication; design and learning; gaming; interaction; e-health and pervasive healthcare; energy management; smart grids; internet of things; cognitive radio networks; computation; cloud computing; ubiquitous connectivity, and in mode general smart living, smart cities, Internet of Things and more. The series publishes a combination of expanded papers selected from hosted and sponsored European Alliance for Innovation (EAI) conferences that present cutting edge, global research as well as provide new perspectives on traditional related engineering fields. This content, complemented with open calls for contribution of book titles and individual chapters, together maintain Springer's and EAI's high standards of academic excellence. The audience for the books consists of researchers, industry professionals, advanced level students as well as practitioners in related fields of activity include information and communication specialists, security experts, economists, urban planners, doctors, and in general representatives in all those walks of life affected ad contributing to the information revolution.

Indexing: This series is indexed in Scopus, Ei Compendex, and zbMATH.

About EAI - EAI is a grassroots member organization initiated through cooperation between businesses, public, private and government organizations to address the global challenges of Europe's future competitiveness and link the European Research community with its counterparts around the globe. EAI reaches out to hundreds of thousands of individual subscribers on all continents and collaborates with an institutional member base including Fortune 500 companies, government organizations, and educational institutions, provide a free research and innovation platform. Through its open free membership model EAI promotes a new research and innovation culture based on collaboration, connectivity and recognition of excellence by community.

Zoltán Balogh
Editor

The 16th International Scientific Conference on Distance Learning in Applied Informatics

DIVAI 2026

Springer

Editor
Zoltán Balogh
Department of Informatics
Constantine the Philosopher University
Nitra, Slovakia

ISSN 2522-8595 ISSN 2522-8609 (electronic)
EAI/Springer Innovations in Communication and Computing
ISBN 978-3-032-32103-9 ISBN 978-3-032-32104-6 (eBook)
https://doi.org/10.1007/978-3-032-32104-6

© European Alliance for Innovation 2027

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Unless otherwise expressly agreed by the Publisher in writing, the Publisher does not authorise use or reproduction of all or any part of this work in any manner for the purpose of training, fine-tuning, evaluating, or developing artificial intelligence models, technologies, or systems. In accordance with Article 4(3) of the EU Directive 2019/790 on copyright and related rights in the Digital Single Market, the Publisher expressly reserves all rights with respect to this work under the text and data mining exception of the directive.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

We are delighted to introduce the proceedings of the 16th International Scientific Conference on Distance Learning in Applied Informatics (DiVAI 2026). This conference, organized by the European Alliance for Innovation (EAI) in cooperation with the Department of Informatics, Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, Slovakia, and Óbuda University in Budapest, Hungary, was held during May 4–6, 2026, at the Vadas Thermal Resort in Štúrovo, Slovakia. DiVAI has brought together researchers, educators, developers and practitioners from across Europe and beyond who are leveraging modern information and communication technologies in education, intelligent systems, and applied informatics.

The conference received over 66 submissions, of which 55 full papers were accepted, including 2 invited keynote papers, all of which were presented during the conference sessions. The technical program of DiVAI 2026 was organized into four main conference sessions and two special sessions. The main conference tracks covered two broad thematic areas: Educational and Learning Technologies, Intelligent Systems, Data, and Applied Informatics. The four main sessions were chaired by Prof. Boris Aberšek, Dr. Gergely Sebestyén, Prof. Attila Kővári, and Dr. Jan Francisti. In addition, the conference featured two special sessions: Smart Biosignal AnalysisIoT, Sensors and Machine Learning Applications, chaired by Prof. Zoltán Balogh, and Algorithmic Trading and Computational Intelligence in Finance, chaired by Prof. Martin Drozda. Apart from the high-quality technical paper presentations, the technical program also featured two keynote speeches. The first keynote was delivered by Prof. Viera Stopjaková from the Slovak University of Technology in Bratislava, Slovakia, on the topic of "Path to Energy-Autonomous Systems-On-Chip Through Ultra-Low-Voltage Design and Energy Harvesting." The second keynote was presented by Prof. Tomáš Pitner from Masaryk University, Czech Republic, on "Scaffolding Data-Centric AI: Engineering Patterns and Failure Modes in Undergraduate Projects."

Coordination with the steering chair, Prof. Imrich Chlamtac, and the conference chair, Prof. Milan Turčáni, was essential for the success of the conference. We sincerely appreciate their constant support and guidance. It was also a great pleasure to work with such an excellent organizing committee team for their hard work in organizing and supporting the conference. In particular, the Technical Program Committee, who completed the peer-review process of all submitted papers and ensured a high-quality technical program. We are also grateful to the EAI conference management team for their support, and to all the authors who submitted their papers to the DiVAI 2026 conference.

We strongly believe that the DiVAI conference provides a valuable forum for all researchers, educators, developers and practitioners to discuss science and technology aspects relevant to distance learning, educational technologies, intelligent systems,

and applied informatics. We also expect that the future DiVAI conferences will be as successful and stimulating as indicated by the contributions presented in this volume.

Zoltán Balogh

Conference Organization

Steering Committee

Imrich Chlamtac	European Alliance for Innovation (EAI)
Milan Turčáni	OZ Distance Learning in Applied Informatics (DiVAI) in Nitra, Slovakia

Organizing Committee

Conference Chair

Milan Turčáni	OZ Distance Learning in Applied Informatics (DiVAI) in Nitra, Slovakia

Scientific Chair

Zoltán Balogh	Constantine the Philosopher University in Nitra, Nitra, Slovakia

Conference Patrons

František Petrovič	Constantine the Philosopher University in Nitra, Slovakia
Levente Kovács	Óbuda University Budapest, Hungary

Technical Program Committee Chair

Ľubomír Benko	Constantine the Philosopher University in Nitra, Slovakia

Special Session Chairs

Zoltán Balogh	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Martin Drozda	Slovak University of Technology Bratislava, Slovakia

Technical Chairs

Tibor Tóth	Constantine the Philosopher University in Nitra, Slovakia
Katarína Látečková	Constantine the Philosopher University in Nitra, Slovakia

Web Chair

Jan Francisti	Constantine the Philosopher University in Nitra, Slovakia

Conference Coordinators

Milan Turčáni	OZ Distance Learning in Applied Informatics (DiVAI) in Nitra, Slovakia
Zoltán Balogh	Constantine the Philosopher University in Nitra, Slovakia
Ľubomír Benko	Constantine the Philosopher University in Nitra, Slovakia
Jan Francisti	Constantine the Philosopher University in Nitra, Slovakia
Katarína Látečková	Constantine the Philosopher University in Nitra, Slovakia
Martin Vozár	Constantine the Philosopher University in Nitra, Slovakia

Technical Program Committee

Boris Aberšek	University of Maribor Maribor, Slovenia
Agnaldo Arroio	University of São Paulo São Paulo, Brazil
Abzetdin Adamov	Kafkaz University Baku, Azerbaijan
Zoltán Balogh	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Peter Baranyi	Széchenyi István University Győr, Hungary
Štefan Beňuš	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Martin Bílek	Charles University Praha, Czech Republic
Martin Boltižiar	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Andris Broks	University of Latvia Riga, Latvia
Jana Burgerová	University of Prešov in Prešov Prešov, Slovakia

Douglas Butler	ICT Training Centre Oundle, UK
Klára Císařová	Technical University of Liberec Liberec, Czech Republic
Soňa Čeretková	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Zsolt Čonka	Óbuda University Budapest, Hungary
Martin Drlík	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Martin Drozda	Slovak University of Technology Bratislava, Slovakia
Ludvík Eger	University of West Bohemia Plzeň, Czech Republic
Ladislav Főző	Technical University of Košice, Košice, Slovakia
Jan Francisti	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Mikuláš Gangur	University of West Bohemia Plzeň, Czech Republic
Sue Greener	Brighton Business School Brighton, UK
Claudio Guarnaccia	University of Salerno Salerno, Italy
Hashim Habiballa	Ostrava University, Czech Republic
Milan Houška	Czech University of Life Sciences Prague, Prague, Czech Republic
Mikuláš Hub	Slovak University of Technology Bratislava, Slovakia
Marie Hubálovská	University of Hradec Králové, Hradec Králové, Czech Republic
Štěpán Hubálovský	University of Hradec Králové, Hradec Králové, Czech Republic
Josef Hynek	University of Hradec Králové, Hradec Králové, Czech Republic
Mirjana Ivanović	University of Novi Sad, Novi Sad, Serbia
Tomáš Javorčík	University of Ostrava Ostrava, Czech Republic
Gabriel Juhás	Pan-European University Bratislava, Slovakia
Jozef Kapusta	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Cyril Klimeš	Mendel University in Brno, Brno, Czech Republic
József Kopják	Óbuda University, Budapest, Hungary
Štefan Koprda	Constantine the Philosopher University in Nitra, Slovakia
Peter Korba	Technical University of Košice, Košice, Slovakia
Kateřina Kostolányová	University of Ostrava Ostrava, Czech Republic
Levente Kovács	Óbuda University, Budapest, Hungary
Attila Kővári	Eszterházy Károly Catholic University, Hungary

Jan Kříž	University of Hradec Králové, Hradec Králové, Czech Republic
Vincentas Lamanauskas	Vilnius University Šiauliai, Lithuania
Jan Lojda	ČADUV, Praha, Czech Republic
Arno Louw	University of Johannesburg, Johannesburg, South Africa
Gabriela Lovászová	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Josef Malach	Ostrava University, Czech Republic
Martin Magdin	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Peter Mikulecký	University of Hradec Králové Hradec Králové Czech Republic
Eva Milková	University of Hradec Králové, Hradec Králové, Czech Republic
György Molnár	Óbuda University, Budapest, Hungary
Nataliia Morze	Borys Grinchenko Kyiv University, Ukraine
Renate Motschnig	University of Vienna, Vienna, Austria
Arnošt Motyčka	Mendel University in Brno, Brno, Czech Republic
Michal Munk	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Daša Munková	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Michal Musílek	University of Hradec Králové, Hradec Králové, Czech Republic
István Németh	Óbuda University, Budapest, Hungary
František Petrovič	Constantine the Philosopher University in Nitra, Nitra, Slovakia
Vladimír Pisarský	University of Prešov in Prešov, Prešov, Slovakia
Tomáš Pitner	Masaryk University, Brno, Czech Republic
Miroslav Plevný	University of West Bohemia, Plzeň, Czech Republic
Petra Poulová	University of Hradec Králové, Hradec Králové, Czech Republic
Ivana Rábová	Mendel University in Brno, Brno, Czech Republic
Lucie Rohlíková	University of West Bohemia in Pilsen, Plzeň, Czech Republic
Ivana Šimonová	Jan Evangelista Purkyně University Ústí nad Labem, Czech Republic
Jiří Šťastný	Mendel University in Brno, Brno, Czech Republic
Antonín Slabý	University of Hradec Králové, Hradec Králové, Czech Republic

Viera Stopjaková	Slovak University of Technology Bratislava, Slovakia
Róbert Szabolcsi	Óbuda University, Budapest, Hungary
Marta Takács	University of Novi Sad, Novi Sad, Serbia
Darina Tóthová	EUNIS Slovakia, Slovakia
Tomáš Tóth	Slovak University of Agriculture in Nitra, Nitra, Slovakia
Pavel Trojovský	University of Hradec Králové, Hradec Králové, Czech Republic
Eugenia Smyrnova-Trybulska	University of Silesia Katowice, Poland
Milan Turčáni	University of Ostrava, Ostrava, Czech Republic
Andreas Ulovec	University of Vienna, Vienna, Austria
Mateja Ploj Virtič	University of Maribor, Maribor, Slovenia
Wei-Chi Yang	Radford University, Radford, Virginia, USA
Ján Záhorec	Comenius University in Bratislava, Bratislava, Slovakia

Contents

Transforming Education, Artificial Intelligence and Learning Innovation

Path to Energy-Autonomous Systems-On-Chip Through Ultra-Low-Voltage Design and Energy Harvesting 3
Viera Stopjaková

Scaffolding Data-Centric AI: Engineering Patterns and Problem Categories in Undergraduate Projects 15
Miroslav Jaroš, Tomáš Pitner, and Oldřich Trenz

From Disruption to Evolution: The Next Era of Educational Transformation . . . 27
Boris Aberšek, Metka Kordigel Aberšek, and Maja Kerneža

Comparison of Human and GenAI Feedback Generation for Programming Assignments 39
Ľubomír Benko, Janka Pecuchová, and Lucia Benková

Integrating AI and Formative E-Assessment in Geometry Education 52
Katerina Biernatova and Lilla Korenova

Interaction Between Real and Virtual Environments in Chemistry Education Using the Example of Redox Reactions 63
Hana Hlavsova, Veronika Machkova, and Martin Bilek

An Indicator-Based Framework for Adaptive Distance Learning: Design of a Learner-Needs Model and Expert Validation 75
Iryna Bilous, Viktoriia Homotiuk, Oksana Gomotiyk, Azeta Tartaraj, Beata Akimjakova, and Oto Mucka

Experiences with AI, Attitudes Towards Its Use and Ethical Concerns of Teaching Students: Correlation Analysis 85
Jana Burgerová and Vladimír Piskura

STEM in Education: A Comparative Analysis of the Influence of Physical Hardware and Simulation on the Development of Algorithmic Thinking 98
Petr Coufal and Eva Milková

Critical Thinking and Media Use in Technical Education 110
Elod Gogh and Attila Kovari

Student Perceptions of School Support for Creativity and Innovation in Technical Education 119
Elod Gogh and Attila Kovari

Inquiry-Based Learning in Secondary Education Through Scavenger Applications 129
Imrich Jakab, Zuzana Pucherová, Martina Agarski, and Dana Kollárová

Didactic Potential of 3D Pen Activities Designed by Pre-service Teachers in the Context of the New National Curriculum: A Qualitative Content Analysis 141
Maria Anna Jedina, Lilla Korenova, Angelika Schmid, and Katarina Zilkova

Students' Intrinsic Motivation When Using Moodle in Informatics Classes 153
Ladislav Kalous and Miloslava Cerna

Reading and Writing as Enabling Practices in Technology-Enhanced Learning for Students with Special Educational Needs 163
Maja Kerneža, Maja Vičič Krabonja, and Andrej Flogie

AI Literacy of Teachers in the Czech Republic in the Context of DigComp 3.0 175
Kristyna Kilianova, Petra Kockova, and Katerina Kostolanyova

The Use of Guilford's Structural Model of Intelligence in the Design of Adaptive Learning Systems 187
Ernest Kováč, Jozef Belko, Jaroslav Kováč, Ľubomír Pekarčík, Miroslav Tvrdoň, and Filip Makáň

Use of Artificial Intelligence: University Students' Positions in Lithuania and Romania 199
Vincentas Lamanauskas, Gabriel Gorghiu, and Costin Pribeanu

Digital Inclusive Technologies in the Literacy Education of Primary School Pupils 211
Filip Makan, Andriana Shyshak, Rostyslav Rudenskyi, Silvia Jakabová, Miroslav Tvrdoň, and Beáta Akimjaková

Hierarchical Object Detection in Historical Handwritten Substitution Cipher Keys Using YOLOv11 222
Stanislav Marochok and Michal Balogh

Generative AI in Academia: Student Perspectives 235
David Martinčík, Olga Martinčíková Sojková, Václav Procházka, and Jan Tlučhoř

A High-Fidelity VR Framework for Immersive Performance Analytics and Seamless User Feedback in Construction 4.0 Environments 248
Mariana Moreiras, Leonel Deusdado, and Luís Afonso

Developing Spatial Imagination and STEM Skills Using Unmanned Technologies 260
Radek Nemec

Kandó Cloud Campus: A Project-Based, Distributed-Resource Higher-Education Ecosystem 272
István Németh, Gergely Sebestyén, Dávid Sik, and György Molnár

Intelligent Database Systems for Sustainable Development: Real-Time ESG Monitoring in African FMCG Operations 285
Chinyere Chidera Okechukwu, Petra Poulová, and Pavel Bachmann

Integrating Generative Artificial Intelligence as a Learning Assistant in Software Engineering: Evaluating Developer Productivity with Cost-Benefit Analysis 300
Marta Palade, Maria Elena Boatca, Andra Diaconescu, and Anca Draghici

Transforming Higher Education: A Case Study on the Integration of Coursera at the University of Hradec Králové 312
Petra Poulová and Miloslava Černá

Digital Education as a Tool for Shaping Modern Human Capital: A Comparative Analysis of Eastern European Countries 322
Viktoriia Savitska, Oksana Gomotiuk, Mykola Karmaz, Tomáš Lengyelfalusy, Ervín Weiss, and Filip Makan

Mapping Citizen Science Platforms for Environmental Education in Slovakia 337
Katarína Skokanová and Imrich Jakab

Bridging the Policy–Practice Gap in Primary Informatics Education: Ukrainian Teachers' Practices and the Polish Curricular Framework 349
Nataliia Morze and Eugenia Smyrnova-Trybulska

Reproducible Synthetic Tabular Data Generation for Education with Gemini 362
Olga Martinčíková Sojková, Miroslav Adámek, and Mikuláš Gangur

Analysis of the Pilot Verification of the Questionnaire on the Use of Digital Teaching Aids by Teachers in the School Education Process 372
Júlia Sudorová, Ján Záhorec, and Michal Munk

Data-Driven Analysis of EV Charging Behavior for Priority-Oriented Load Management Systems 385
István Szűcs, József Kopják, and Gergely Sebestyén

Digital Green Campus Interventions: Conditions Under Which AI and IoT Generate Educational Value Alongside Operational Sustainability 395
Imre Tóbel

AI-Generated Single-Use Educational Applications: The Process of Creation, Publication and Quality Verification 405
Petr Voborník

Teaching Art History with the Help of a Three-Dimensional Virtual Model 420
Jan Vopršal and Marie Hubálovská

Voice-Controlled Game Interaction 432
Patrik Voštinár and Radoslav Proft

Feedback and Assessment as a Comprehensive Learning Support System in Online Teacher Education Courses 445
Jana Vejvodová and Lucie Rohlíková

PLECS for Resource-Constrained and Distance-Friendly Power Electronics Education: An Education-Focused Comparison with MATLAB/Simulink 457
Mark Wendler, Sebestyén Gergely, and József Kopják

Assessing Cognitive Load in Technology-Enhanced Learning Environments: An Interface-Based Evaluation Framework 469
Iryna Yaroshchuk, Oksana Gomotiyk, Oleksandr Kolesnyk, Lubomír Pekarcik, Tomáš Lengyelfalusy, and Filip Makan

Modeling Cyber Security in Smart Cities: A Knowledge-Based Approach 482
Andrea Žváčková, Daniela Ponce, Peter Mikulecký, Pavel Čech, Karel Mls, Tereza Otčenášková, and Petr Tučník

Smart Biosignal Analysis: IoT, Sensors and Machine Learning Applications

Security Evaluation of Communication Protocols in Wireless Sensor Networks for Smart Building Environments 495
Zoltán Balogh, Zsolt Čonka, Martin Magdin, Norbert Ádám, and Shkëlqim Fortuzi

Multimodal Analysis of VR Training Scenarios for Industry 5.0: An Interdisciplinary Approach to Physiological, Behavioral, and Technological Aspects 508
Denis Ďuriač and Zoltán Balogh

Machine Learning Classification of Soil Properties for Discriminating Regenerative and Conventional Agricultural Management Regimes Using IoT Sensor 523
Kristián Fodor and Zoltán Balogh

Analysis of Security Risks Associated with Cookies in University Information Systems 538
Jan Francisti, Marián Hlavačka, Azeta Tartaraj, and Zoltán Balogh

Human vs. LLM Risk Assessment: A Comparative Study of ISO 27001 Cyber Risk Classification 548
Marek Hrabčák and Zoltán Balogh

Development of a Robot Function to Support Children's Movement Coordination: A NAO-Based Approach 561
Enikő Nagy and György Molnár

Low-Cost Flight Data Recording and Cloud-Based Analysis Platform for General Aviation Training 571
Martin Vozár, Ladislav Főző, Peter Korba, and Zoltán Balogh

Algorithmic Trading and Computational Intelligence in Finance

Gamified Financial Education: Human vs. Deep Learning Trading Agents 587
Maros Cavojsky, Filip Hodon, and Martin Drozda

Cumulative True Range for Benchmarking of Trading Algorithms 599
Martin Drozda and Maros Cavojsky

Kagi Chart Representations as Structured Inputs for Deep Learning Models 611
Filip Hodon, Maros Cavojsky, and Martin Drozda

The Impact of BinaryTarget Engineering on Sector-Wise Class Differentiation in Next-Day Return Prediction 624
Lívia Kelebercová, Michal Munk, and Jakub Marták

Geometric and Topological Perspectives on Financial Market Dynamics 635
Jakub Marták, Michal Munk, Daša Munková, and Lívia Kelebercová

Enhancing Time Series Forecasting Using Persistent Homology 649
Patrik Sebes, Martin Drozda, and Maros Cavojsky

Portfolio Optimization with TD3 Algorithm Using Technical Indicators 662
Teodora Šimon, Peter Likavec, Maros Cavojsky, and Martin Drozda

Author Index 677

About the Editor

Prof. Ing. Zoltán Balogh, PhD. received his engineering degree (Ing.) in 2000 in the study program of Electrical Engineering, Automation and Informatics in Agriculture at the Faculty of Mechanical Engineering, Slovak University of Agriculture in Nitra, Slovakia. He received his PhD degree in 2004 at the Slovak University of Agriculture in Nitra. In 2012, he habilitated in the field of System Engineering and Informatics at the Faculty of Informatics and Management, University of Hradec Králové, Czech Republic, and in 2022, he received his inauguration (professorship) in the same field at the same institution. Currently he is a Professor at the Department of Informatics, Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, Slovakia, where he leads the research team on Modeling, Simulation and Development of Systems for the Internet of Things (IoT). His research interests include process modeling using Petri nets, IoT sensor networks, biosignal analysis, machine learning for data-driven systems, and educational technology. He has authored 70 publications indexed in Web of Science and 80 publications indexed in Scopus, including journal articles in Expert Systems with Applications, Scientific Reports, Results in Engineering, Sustainable Futures, Smart Learning Environments, and Acta Polytechnica Hungarica (Elsevier, Springer, and Nature Portfolio), as well as numerous contributions in Springer and IEEE publication series. His work has received over 600 citations, with an H-index of 13 in both Web of Science and Scopus. He has led or co-investigated multiple national and international research projects (VEGA, KEGA, APVV, Erasmus+). Zoltán Balogh is a member of the Scientific Council of the Faculty of Informatics and Management, University of Hradec Králové, an evaluator of the Slovak Accreditation Agency for Higher Education, and a member of IEEE, INSTICC, and IFAC. He is a member of several conference program committees, a co-founder of the DiVAI conference and serves as its Scientific Chair and volume editor since 2024.

Transforming Education, Artificial Intelligence and Learning Innovation

Path to Energy-Autonomous Systems-On-Chip Through Ultra-Low-Voltage Design and Energy Harvesting

Viera Stopjaková(✉)

Department of IC Design & Test, Institute of Electronics and Photonics, Slovak University of Technology, Ilkovičova 3, 841 04 Bratislava, Slovakia
viera.stopjakova@stuba.sk

Abstract. Advances in semiconductor technology enable the development of high-performance electronic systems with incredible computing power; however, these advances also bring new parasitic effects, reduced supply voltage, significant fluctuations in process parameters, and other phenomena typical of new, emerging technologies. All these technological limitations, combined with strict requirements for low power consumption and compact dimensions, significantly complicate the design of systems-on-chip. Therefore, low-voltage and low-power integrated circuits, particularly analog ones, are designed using unconventional design techniques based either on non-standard devices and structures or on new circuit topologies and transistors operating in the subthreshold region. This article provides an overview of design methods for (ultra) low-voltage and low-power digital and analog circuits, as each requires a specific design approach. We then focus on energy harvesting systems that convert alternative energy (available in the environment) into electrical energy. Finally, we present an example of a developed on-chip photovoltaic energy harvesting system capable of self-powering, which was fabricated using a standard 65 nm CMOS process.

Keywords: Low Power · Ultra-Low-Voltage · Energy Harvesting · IC Design · Energy-Efficient Systems · IoT · Systems-on-chip

1 Introduction

Research and development in the field of electronics is largely driven by the real needs of society, but it is constrained by currently available materials and technologies. Advances in semiconductor technology are closely linked not only to the performance of electronics but also to its energy efficiency and certain limitations when speaking about microelectronic circuit design. Reducing physical dimensions of components has positive effect on the power consumption, while increasing the component integration density brings more computing power for the same chip area [1].

These advances have paved the way for the emergence of smart and wireless electronics, such as the Internet of Things (IoT) and wearable electronics [2, 3], where power dissipation has become a major concern, since the battery-powered devices and systems

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 3–14, 2027.
https://doi.org/10.1007/978-3-032-32104-6_1

penetrate to diverse domains of industry and social life (e.g. smart sensors, automotive industry, space systems, sensor node networks, medical equipment, wearable electronics, etc.).

Recently, providing the required performance features and enough of computing power in a small area is no longer issue thanks to continuous development of semiconductor technologies. Advanced technologies offer high-performance circuit components (e.g. logic gate propagation delays range from tens to single-digit picoseconds), which is very desirable for implementation of high-speed circuits. On the other hand, there are several factors and issues associated with nanoscale technologies, which make design of integrated circuits (IC) and systems-on-chip (SoC) rather difficult and challenging, especially when efforts are being made to simultaneously reduce power consumption and chip size while still targeting the high performance. Although the emerging technologies (e.g. 3 nm FinFET) inherently offer transistors working with low voltages and are optimized for low-power operation, these technologies lack quality electronic components vital for reliable analog and/or mixed-signal design (resistor, capacitor, etc.). Therefore, as a rule, only (selected) digital circuits are manufactured using cutting-edge technologies.

As already mentioned, advanced technologies and device shrinking are linked with increased integration density of circuit components, and low-voltage and low-power operation of developed ICs. Thus, mainly design of analog and mixed-signal integrated circuits (or systems) face serious limitations in advanced technologies. One of the biggest challenges in analog IC design is the reduction in the power supply voltage (which scales down much faster than the MOS transistor threshold voltage). As a result, the available voltage headroom is drastically reduced; consequently, for many existing and widely used circuit topologies, this drop in power supply voltage means that their use is either limited or even impossible. Thus, new circuit topologies and design approaches must be developed and used for low-power and ultra-low-voltage ICs.

Nevertheless, in the case of new nanotechnologies, circuit parameters (such as immunity to interference, maximum current densities, parasitic effects, process variations, aging effects, and other parameters) are inferior to those of proven CMOS processes using planar transistors. This is the main reason why analog and mixed-signal circuits ICs are still being designed in matured technology nodes.

In wearable electronics or IoT devices, which are typically battery-powered, battery life depends on the energy consumption of the electronic system itself and on the battery capacity. In most applications, recharging or replacing the battery is easy to do. However, in hard-to-reach places or in implantable medical devices, such maintenance may be either difficult or impossible. From this perspective, using alternative energy sources from the environment and reducing energy consumption are very important. The first way to meet this requirement is to develop integrated circuits and systems capable of operating at very low supply voltages and exhibiting low power consumption. The second approach involves using circuits that are self-powered (at least partially).

So-called energy harvesters (EH), generating electrical energy from ambient environment, represent a promising solution towards enhancing the battery lifetime or avoiding the necessity of battery change. Energy harvesting systems have been developed for virtually all sorts of ambient energy sources. However, not every harvester type

is suitable and feasible for full or partial integration onto silicon. The recent development of on-chip energy harvesters mainly uses MEMS piezoelectric, radiofrequency (RF), thermal or solar energy sources employing the maximum power point tracking (MPPT) algorithms and further energy storage management [4–12]. Nevertheless, for some energy harvesters, power management circuitry (consisting of voltage converters, voltage comparators, etc.) are needed [13, 14].

2 Technology Development vs Design Challenges

As already mentioned, the integration of complex electronic systems on a single chip die represents a perspective challenge in the field of design, development and implementation of ICs and SoCs. The possibility of high integration is supported by ongoing development of fabrication technologies, which allow implementation of heterogenous SoC, consisting of analog and digital circuits but also MEMS structures and wireless sensors with RF circuitry within one chip/package. The high-level integration significantly reduces fabrication costs, and moreover, it contributes to tremendous benefits in terms of signal integrity and overall performance (e.g. lower propagation delay, reduced number of input/output ports, less external components, etc.). On the other hand, high-level integration increases the complexity of integrated systems and brings demands on the energy accumulator that supplies the whole system. Consequently, the current (energy) consumption and the efficiency of energy management of conversion systems and being implemented directly on a chip, become extremely important research objectives. Thus, energy systems supplying individual parts of a SoC represent an important issue that the scientific community, as well as the industry, must deal with. The development of technologies, their resolution and minimal technology dimensions are shrinking by the factor of 0.7 every two years, which doubles the number of circuit elements integrated on the same area every two years as well, as predicted by ITRS [15]. Furthermore, the power consumption decreases with VDD square, which correlates with the trend shown in Fig. 1 (left). On the other hand, lowering the dimensions of structures and interconnections between them introduces new parasitic phenomena, which cause difficulties and obstacles in design of mixed-signal IC in nanoscale technologies. Latest technologies are perfectly suitable mainly for digital circuits such as processors, memories and others, where we can fully benefit and profit from integrating much more transistors on the same area. Digital circuits fabricated in nanotechnologies can be clocked at higher frequencies that also increases the data processing rate. Another important fact is that digital circuits are generally less susceptible to parasitic phenomena due to the discrete voltage levels at which they operate. These phenomena hinder progress in the design of analog integrated circuits using new technologies. When designing complex SoCs, there is a growing need to integrate analog circuits into advanced SoCs. These can be used, for example, as units that process signals from sensors (external or integrated MEMS structures).

From the IC design point of view, one of the main problems caused by a lowered VDD is the reduction of useful voltage range for existing and standard circuit topologies, since the transistor threshold voltage (V_{TH}) cannot follow the trend of the supply voltage level decrease due to substantial leakage currents. Decreasing the threshold voltage and

simultaneously reducing thickness of the gate oxide of transistors cause steep rising of the subthreshold leakage current that is typical for nanotechnologies. These reasons do limit the further decrease of the threshold voltage. Figure 1 (right) depicts the dependency of the VDD level and the threshold voltage on the technology node for last several years. Reduction of the VDD – V_{TH} gap drastically affects the available voltage room, and mainly analog ICs are suffering mostly from this drawback. Therefore, for analog blocks and cores, providing the low power consumption together with other requirements for high performance SoC is rather difficult in nanotechnologies.

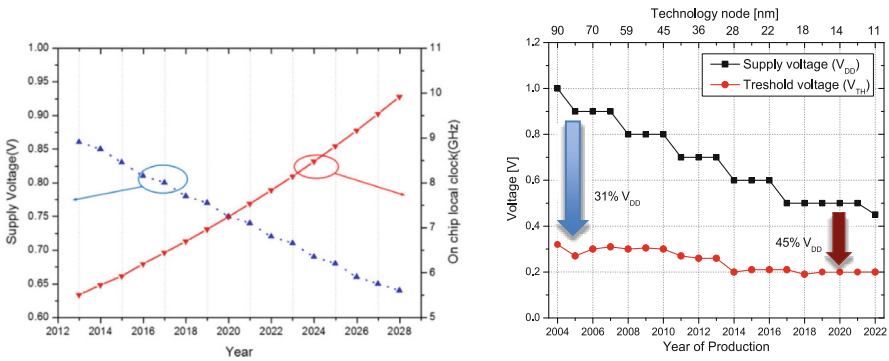


Fig. 1. Trends in VDD and the clock frequency (left), and in the VDD – V_{TH} gap (right) with respect to technology node.

The rising integration scale also exhibits disadvantages such as new limiting undesired effects affecting the robustness and reliability of ICs, e.g. significant variation of technological parameters, specific physical defects and phenomena like NBTI (Negative Bias Temperature Instability) or electromigration, elevated subthreshold leakage current, gate leakage current due to hot electron injection, gate oxide breakdown and so on. These parasitic effects degenerate characteristic of electrical parameters such as threshold voltage of the transistors or transconductance (g_m) and output conductance (g_{ds}). These factors also negatively affect other important characteristics of integrated systems, such as reliability or power consumption in idle mode. Another problem caused by the miniaturization of electronic components is increased sensitivity to variations in manufacturing processes and temperature changes, particularly in the case of small-sized transistors. Together with VDD variations, these are known as PVT (Process-Voltage-Temperature) variations. In addition, short-channel transistors exhibit higher offset, subthreshold leakage current, and lower intrinsic transconductance compared to long-channel transistors. Low supply voltage also reduces the maximum signal swing, which significantly reduces the dynamic range and results in a poorer signal-to-noise ratio (SNR) for analog integrated circuits. As a result, SoC manufacturing yield is reduced, aging occurs more rapidly, and reliability is also affected.

Energy-efficient SoCs, designed primarily for use in portable, miniature, battery-powered devices, are expanding the scope of application for smart integrated solutions across various sectors of industry and society. For these reasons, there is a growing

need to develop new methods and methodologies for designing integrated circuits with extremely low VDD voltages in nanotechnology. This need is a natural response to the already unsatisfactory properties of standard topologies of basic building blocks of integrated circuits and the conventional design methodologies used (this applies mainly to analog and mixed-signal ICs).

3 Low-Voltage / Low-Power Design Approaches

Reducing power consumption and supply voltage are among the most important requirements in the development of modern integrated systems, which necessitates energy efficiency in the analog and digital circuits that make up the entire system. To ensure the proper functioning of analog and digital circuits, it is first necessary to use different voltage domains within a single chip so that they can operate together but with minimum interfering with one another. The nature of a circuit also determines the methods used in their low-power and low-voltage design.

3.1 Low-Power Digital Design

As already mentioned, the development of new semiconductor technologies brings benefits particularly for digital circuits, which are typically the first to be produced in a new technology node. Given these advantages, it can generally be said that achieving high performance in digital integrated circuits is relatively straightforward; the challenge, however, lies in energy efficiency. The main fraction of the total power of digital circuits (in CMOS process) is represented by the dynamic power given by (1):

$$P_{dynamic} = \alpha\, C_L\, f\, VDD^2, \tag{1}$$

where α is activity factor, C_L represents the load capacitance, f is the clock frequency and VDD is the power supply voltage. One can observe that the supply voltage plays a decisive role in this case, as dynamic power consumption increases quadratically with it. Clock frequency f and load capacitance C_L may also affect the power consumption, however, these parameters are determined mainly by performance requirement and circuit complexity. In digital circuits, the activity factor α has also significant impact on the dynamic power, as it represents the fraction of digital nodes switching from logic 0 to logic 1, and this parameter can be reduced by selection of appropriate circuit topology.

The most used low-power digital design approaches include:

- Power gating
- Clock gating
- Multi-voltage domain
- Logic optimization

The power gating technique helps reduce power consumption by turning off the power to parts of the circuit that are not currently in use (or inactive). This can be achieved in two ways: fine power-off or coarse power-off, depending on the location of the switching transistors. The principle behind clock gating is to disable the clock of

those sequential elements whose data is not toggling or elements that are not currently in use. Multi-voltage domain is another low-power design technique a system-on-chip is partitioned into different power domains, each running at optimized voltage level, which allows high-performance blocks to use a higher voltage, while non-critical or idle blocks operate at lower voltages. The last approach is designing logic paths to minimize glitches (unintended temporary switching before the circuit stabilizes) that contribute to higher average activity factors.

All these techniques significantly reduce either dynamic power dissipation or leakage.

3.2 Analog Low-Voltage and Low-Power Design

There have been several design techniques developed for low-voltage (LV) and low-power analog ICs. Very common and useful approach in LV analog design is using the differential or balanced structures. The most used approaches include [16, 17]:

- Floating-gate MOS transistors
- Self-cascode topologies
- Level shifters
- Bulk-driven MOS transistors
- Dynamic-threshold MOS transistors

In general, the design techniques described above have certain performance limitations, such as degradation of the gain bandwidth (GBW), increased input noise and impedance, and decreased overall transconductance. The main characteristics of the introduced design techniques for analog circuits are summarized and compared in Table 1 [16]. As low-voltage circuit design techniques can be used for various technologies, design methods involving MOS transistors operating in the sub-threshold region such as bulk-driven and dynamic-threshold transistors are beneficial. These techniques do not necessitate alterations to the standard CMOS manufacturing process. Furthermore, using these two low-voltage methods enables the designed circuits to achieve low power consumption.

In the case of a conventionally connected MOSFET, the gate is used as an input electrode to modulate the transistor's channel and control the output current. However, in the bulk-driven design technique, the bulk electrode of the MOS transistor acts as the signal input, controlling the drain current, as the name of the technique suggests (see Fig. 2 on the left). This technique effectively addresses the issue of threshold voltage at the MOS transistor input, thereby increasing the voltage headroom. The dynamic-threshold approach is similar to the bulk-driven technique, with one key difference: the MOS transistor gate and the bulk electrodes are interconnected, and the threshold voltage is dynamically reduced in response to the input signal swing (see Fig. 2 on the right).

The fact that these two low-voltage design techniques are fully compatible with standard CMOS technology and do not require any additional process steps makes them highly advantageous in terms of cost. They are therefore the most widely used in the design of cost-effective, low-voltage, and low-power analog and mixed-signal ICs.

Examples of integrated systems developed using LV design techniques can be found in references [14, 16–21]. In addition to the LV design approaches mentioned above,

Table 1. Comparison of main parameters of analog LV design approaches.

Design method	GBW	Power consumption	Technology
Floating gate MOS	Medium	High	Double Polysilicon
Self-cascode	Medium	Medium	Multi-threshold
Bulk-driven MOS	Low	Low	Standard (triple-well)
Dynamic threshold MOS	Medium	Low	Standard (triple-well)
Lever shifter	High	Medium	Standard

Fig. 2. A bulk-driven (left) and a dynamic-threshold (right) MOS transistors

various calibration techniques are employed in analog IC design to compensate for the undesirable effects of PVT fluctuations (e.g. input voltage offset compensation in amplifiers) [14, 22].

4 On-Chip Energy Harvesting

As mentioned in Sect. 1, enhanced computational power and reduced energy consumption pave the way for the development of sensor systems, wearable electronics, IoT devices and implantable medical devices. A key energy-related component of these systems is typically the power supply, which is usually a battery. Its ability to provide energy decreases over time and is strongly influenced by the electrical load and environmental conditions. Using energy harvesters (EH) can extend battery life and, in certain applications, eliminate the need for batteries altogether. One of the major advantages of energy harvesting for low-power electronic devices is the wide variety of energy conversion methods available, such as piezoelectric, photovoltaic, thermoelectric and radio-frequency (RF) harvesting, as well as biofuel cells. Furthermore, combining multiple converter technologies in hybrid energy harvesting systems can enhance efficiency and reliability even further.

Although energy harvesters can be used in many electronic systems, the type of EH selected always depends on the intended application. Important factors to consider include the installation environment, the availability of ambient energy sources that can be converted into electrical energy, the achievable output voltage and power density, and, finally, the device's physical size, which is particularly important for on-chip

implementations. Table 2 [23] compares the parameters of various energy converter types.

Table 2. Comparison of energy converters (EC) in terms of main parameters

EC type	Output voltage	Power density	Efficiency (%)
Photovoltaic	~600 mV	100 mW/cm^2	15–25
Thermoelectric	<100 mV	50–100 uW/cm^2	1–17
Piezoelectric	<150 V	15 W	50–90
RF	1–4 V	< 1 W	1–90
Biofuel cell	<300 mV	12 uW/cm^2	<12

For applications where physical dimensions are particularly important, such as integrated electronics or systems-on-chip, the commonly used energy converter is the solar cell. Although the efficiency of PV energy converters is comparatively low, their achievable power density can reach several hundred mW/cm^2, which is adequate for powering a wide range of low-power electronic applications. A single solar cell typically produces an output voltage of a few hundred millivolts, whereas higher voltage levels can be achieved by connecting several cells in series. As output voltage depends on the number of interconnected solar cells and current lighting conditions, a DC-DC step-up voltage converter is typically placed between the solar cell and load/energy storage to ensure stable operation of the EH system.

Additionally, the power-voltage characteristic of a solar cell is inherently nonlinear and typically exhibits a single maximum power point (MPP), as shown in Fig. 3. As the system is in the MPP, it can deliver the maximum target power at the output. In other words, the maximum output power can be extracted from the PV energy converter in this point.

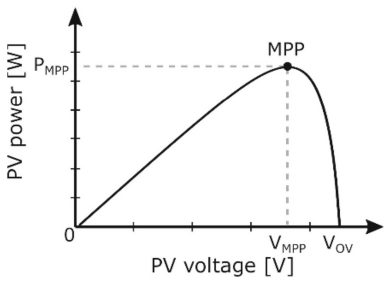


Fig. 3. P-V curve of a solar cell with MPP [23].

The location of the operating point is influenced by factors such as the intensity of incident radiation and the connected load. As these conditions vary dynamically over time, the operating point may shift unpredictably. This can result in reduced energy

harvesting efficiency and, in extreme cases, cause the entire system to malfunction. Thus, for the maximum power extraction from the energy converter device (e.g. PV cell), a control loop implementing maximum power point tracking (MPPT) algorithm is usually used within the energy harvesting system.

There are several MPPT algorithms, which can be categorized into two groups based on the type of inputs and the method of tuning the operating point: indirect and direct algorithms. Indirect MPPT algorithms primarily work with static variables. Using the obtained PV cell characterization data, it is possible to determine the operating point of the energy converter for maximum energy extraction. However, they cannot dynamically respond to changes in input conditions caused by factors such as changes in illumination, partial shading of the PV cell and material degradation due to ageing. Unlike indirect MPPT algorithms, direct MPPT algorithms do not require precise knowledge of the I-V and P-V curves of the solar cell. Instead, they can dynamically adapt to changes in input light conditions, temperature and material degradation. Further details on MPPT algorithms are presented in [23].

As a real example of EH systems, we present a developed on-chip PV-based energy harvester that converts solar energy into electrical energy. It is a relatively complex system on a chip (SoC), where the goal was to achieve the highest possible level of integration using standard 65 nm CMOS technology (excluding the PV cells). A simplified block diagram of the PV-based EH system is shown in Fig. 4. This figure illustrates the parts that have been integrated onto the chip as well as the external discrete components that have been used for the prototyped energy harvester [23].

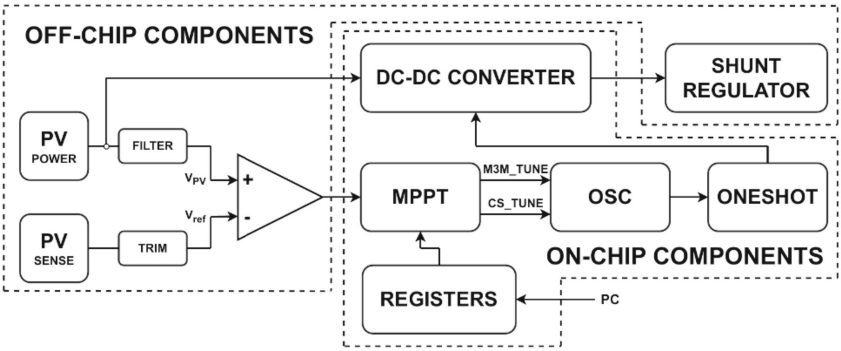


Fig. 4. Block diagram of PV-based EH system with a MPPT control loop [23].

The main component of the energy harvesting (EH) system is a DC-DC voltage converter utilizing a custom-designed fully integrated inductor, DC-DC converter is based on the Conventional Boost Converter (CBC) architecture. The primary objective was to achieve a stable output voltage of 1.5 V from two photovoltaic (PV) cells connected in series, which under optimal conditions (in the MPP) generate approximately 0.56 V at the energy converter output. This requirement naturally led to the implementation of a step-up (boost) converter topology.

To simplify the design and minimize power losses, the first EH prototype employed the most straightforward converter configuration possible. The intention was to reduce energy dissipation by limiting the number of power transistors and minimizing the complexity and consumption of the associated control circuitry. Key elements of the system, including the power switches, inductor, and part of the input capacitor, were fully integrated on-chip.

The converter's low-side power switch is driven by a dedicated control loop that regulates the input impedance of the DC-DC converter, thereby adjusting the EH operating point to maximize energy extraction under varying conditions. The control loop of the DC-DC converter is composed of REGISTERS, which enable parameter adjustments and corrections through a computer interface, MPPT circuitry generating a pulse-frequency-modulated signal for the oscillator, a tunable relaxation oscillator (OSC) with adjustable frequency controlled by capacitor and current-source banks, and a ONE-SHOT circuit responsible for producing pulses with predefined pulse width.

At this development stage, most of functional blocks are integrated directly on the chip, while the photovoltaic cells, shunt regulator, and measurement circuitry remain external components (Fig. 5).

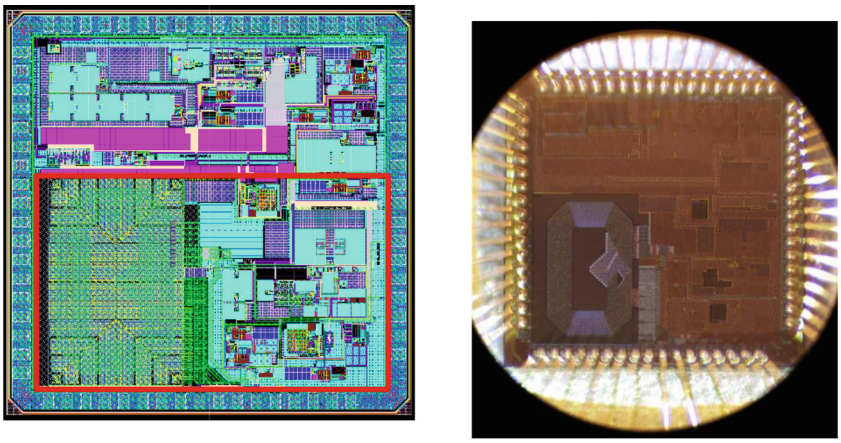


Fig. 5. Layout and a photo of the fabricated EH chip

5 Conclusion

As the use of standard circuit topologies and conventional design techniques is relatively limited under (ultra-) low-voltage and low-power conditions, designers must seek new circuit topologies and design methods that can handle a range of parasitic effects typical of nanoscale technologies and are resistant to PVT variations. The commonly used design techniques for low-voltage and low-power digital and analog ICs were discussed. This was followed by a discussion of power converters and energy harvesting systems, which can extend battery life or even replace it entirely. Finally, a developed energy harvesting

chip with implemented MPPT algorithms was presented as an example. It is believed that ultra-low-voltage and low-power design approaches, which go hand in hand with on-chip energy harvesting, will pave the way for energy-autonomous electronic systems-on-chip in the future.

Acknowledgement. This work was supported in part by grant VEGA 1/0572/25, and the Slovak Research and Development Agency under grants APVV-23-0071 and VV-MVP-24-0313.

References

1. Randall, J.N., Owen, J.H., et al.: Digital atomic scale fabrication an inverse Moore's law—a path to atomically precise manufacturing. Micro Nano Eng. **1**, 1–14 (2018)
2. Ahmad, S., Alam, N., Hasan, M.: Robust TFET SRAM cell for ultralow power IoT applications. AEU-Int. J. Electron. C. **89**, 70–76 (2018)
3. Leoni, A., Pantoli, L., Stornelli, V., Ferri, G., Russo, M., Solic, P.: A combined 90/900 MHz IC architecture for power-assisting in IoT applications. J. Commun. Softw. Syst. **14**(1), 27–32 (2018)
4. Bizon, N., Tabatabaei, N.M., Blaabjerg, F., Kurt, E.: Energy harvesting and energy efficiency. In: Technology, Methods, and Applications, p. 37. Springer, Berlin/Heidelberg, Germany (2017)
5. Cho, K.Y., Kim, J., et al.: Shinsung solar energy high efficiency commercial crystalline Si solar cells. In: 38th Proceedings on IEEE Photovoltaic Specialists Conference, pp. 002145–002148. IEEE, Austin, TX, USA (2012)
6. Ryu, B., Chung, J., et al.: Best thermoelectric efficiency of ever-explored materials. Iscience. **26**, 106494 (2023)
7. Vikhor, L., Kotsur, M.: Evaluation of efficiency for miniscale thermoelectric converter under the influence of electrical and thermal resistance of contacts. Energies. **16**, 4082 (2023)
8. Seo, G.S., Shin, J.W., Cho, B.H.: A magnetic component-less series resonant converter using a piezoelectric transducer for low profile application. In: International Proceedings on Power Electronics Conference-ECCE ASIA, pp. 2810–2814, Sapporo, Japan (2010)
9. Hasanuddin, M.O., Hafizh, A.Z., Prakoso, G.A., Izzuddin, A.: Design and implementation of RF to DC converter for low power IoT device using RF energy harvesting. In: Proceedings of the International Symposium on Electronics and Smart Devices, pp. 1–5, Bandung, Indonesia (2022)
10. Gu, X., Hemour, S., Guo, L., Wu, K.: Integrated cooperative ambient power harvester collecting ubiquitous radio frequency and kinetic energy. IEEE Trans. Microwave Theory Techn. **66**(9), 4178–4190 (2018)
11. Seok, S., Lefeuvre, E., Wang, C., Park, J.: Electrical characterization of textile-based enzymatic biofuel cell for energy harvesting Interface circuit. In: Proceedings on Symposium on Design Test, Integration & Packaging of MEMS and MOEMS, pp. 1–4. France, Paris (2019)
12. Kobayashi, A., Hayashi, K., et al.: A solar-cell-assisted, 99.66% biofuel cell area reduced, biofuel-cell-powered wireless biosensing system in 65-nm CMOS for continuous glucose monitoring contact lenses. In: 26th Proceedings on International Conference on Electronics Circ. Syst., pp. 61–64. IEEE, Genova, Italy (2019)
13. Stopjakova, V., Kovac, M., Potocny, M.: On-chip energy harvesting for implantable medical devices. Radioengineering. **29**(1) (2020)
14. Nagy, L., Potocny, M., Ondica, R., Hudec, A., Stopjakova, V.: A novel ultra low-voltage/low-power rail-to-rail comparator topology in nanoscale CMOS technology. AEU-Int. J. Electron. C. **166**, art. no. 154651 (2023)

15. International Technology Roadmap for Semiconductors 2013 Edition, (ITRS-2013)
16. Stopjakova, V. et al.: Ultra-low voltage analog IC design: challenges, method and examples. Radioengineering. **27**(1), 171–185 (2018)
17. Stopjaková, V., Kováč, M., Arbet, D., Nagy, L.: Towards energy-autonomous integrated systems through ultra-low voltage analog IC design. In: 26th International Conference "Mixed Design of Integrated Circuits and Systems", pp. 38–45. IEEE, Rzeszow, Poland (2019)
18. Khateb, F., Kumngern, M., Kulej, T., Stopjakova, V., Psychalinos, C.: 0.3-V, 357.4-nW voltage-mode first-order analog filter using a multiple-input VDDDA. IEEE Access. **11**, 9663–9664 (2025)
19. Ravasz, R., Stopjakova, V., Maljar, D., Arbet, D., Nagy, L., Kovac, M.: Development of slope detection ASIC for on-chip current sensing in voltage converters. Integr. VLSI J. **104** (2025)
20. Kulej, T., Khateb, F., Arbet, D., Stopjakova, V.: a 0.3-V 62-dB bulk-driven OTA with hybrid frequency compensation in 65 nm CMOS. Int. J. Circ. Theory Appl. **54**, 2987–2998 (2025)
21. Kulej, T., Khateb, F., Arbet, D., Stopjakova, V.: A 0.3-V high linear rail-to-rail bulk-driven OTA in 0.13 μm CMOS. IEEE Trans Circuits Syst II Express Briefs. **69**(4), 2046–2050 (2022)
22. Maljar, D., Sovcik, M., Potocny, M., Ondica, R., Arbet, D., Stopjakova, V.: Digital calibration of input offset voltage and its implementation in FDDA circuits. Electronics. **12**(22), 4615 (2023)
23. Hudec, A., Stopjakova, V., Ondica, R., Potocny, M., Nagy, L.: The impact of maximum power point tracking algorithms on properties of on-chip PV-based energy harvester for IoT devices. Sensors. **27**(3) (2026)

Scaffolding Data-Centric AI: Engineering Patterns and Problem Categories in Undergraduate Projects

Miroslav Jaroš(✉), Tomáš Pitner, and Oldřich Trenz

Faculty of Business and Economics, Department of Informatics, Mendel University in Brno, Brno-sever, Czechia
{miroslav.jaros,tomas.pitner,oldrich.trenz}@mendelu.cz

Abstract. Teaching data-centric AI requires more than training models: students must also learn to collect, document, validate, and iteratively improve datasets. This paper reports an exploratory cohort study embedded in a semester-long undergraduate assignment in which students individually created object-detection datasets end to end, from video capture and annotation in CVAT to COCO export and baseline detector fine-tuning. We analyze 89 structured student reports and a post-project reflection questionnaire from 37 students to identify recurring breakdowns in dataset-engineering workflows and gaps in reproducibility-related reporting. The findings show a contrast between tool-coupled deliverables, which were typically reported consistently, and reproducibility-critical metadata, which was often missing from submitted documentation. Qualitative coding of reflective fields further identified recurrent problem categories related to occlusion, lighting diversity, export semantics, and leakage risks. Rather than claiming causal evaluation of the proposed interventions, the paper contributes a reusable teaching playbook that maps observed problem categories to evidence-first scaffolding measures, such as split manifests, validation scripts, and targeted audit checkpoints. We position the work as a systematic experience report with transferable design implications for computing-education settings teaching data-centric computer vision.

Keywords: Data-centric AI · dataset engineering · computing education · problem categories · scaffolding

1 Introduction

Computer-vision models in practice are often constrained by the availability, coverage, and quality of training data, motivating a shift toward data-centric approaches that treat dataset iteration as a first-class engineering activity [1]. However, "data work" remains comparatively under-incentivized and less systematically taught than model-building workflows [2]. Students frequently learn how to train models before they learn how to specify, validate, and document the datasets those models rely on.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 15–26, 2027.
https://doi.org/10.1007/978-3-032-32104-6_2

In this paper, we treat dataset work as an engineering activity with recurring problem categories and corresponding scaffolding needs. Rather than evaluating a single model or a single dataset, we focus on recurring breakdowns in end-to-end student pipelines and on how course design can shift those breakdowns from late-stage debugging to early-stage prevention.

In project-based undergraduate settings [3], students often experience the full lifecycle of model training but lack structured support for the earlier, more error-prone stages of data collection and packaging. Recent experience reports on teaching end-to-end ML engineering (MLOps) similarly note that standard curricula overemphasize model optimization "in the lab" while under-supporting upstream and pipeline practices, motivating project-based designs with tool-aligned scaffolding [4].

In this paper, we examine a full-cycle object-detection assignment designed to make dataset work explicit, auditable, and teachable. The paper should be read as an exploratory, course-embedded study rather than as a controlled evaluation of instructional effectiveness. Our goal is not to claim discovery of novel computer-vision problem categories, but to identify which known data-quality and reproducibility problems recur in an undergraduate end-to-end assignment, how these problems become visible in student-produced evidence, and how they can be translated into explicit instructional scaffolds.

We analyze artifacts from 89 students to address two research questions:

- **RQ1 (Pipeline Evidence Gaps):** Which stages of the dataset-engineering pipeline are most prone to omission or under-documentation in submitted student reports?
- **RQ2 (Problem Categories and Remediation):** What recurring problem categories appear in student reflections on technical difficulties, and what specific scaffolding actions can be derived to address them?

The contribution of the paper is therefore threefold: (i) a report-based empirical characterization of which evidence elements were consistently documented and which were frequently omitted in one undergraduate cohort, (ii) an exploratory categorization of recurrent student-reported dataset-engineering problems, and (iii) a reusable teaching playbook that maps these observed patterns to concrete, evidence-first scaffolding measures.

2 Related Work

2.1 Data-Centric AI and Dataset Engineering

Data-centric AI argues that many practical gains in machine learning come from iterating on data—requirements, coverage, consistency, and documentation—rather than solely from changing model architectures [1]. In this view, datasets are engineered artifacts: they are specified, validated, versioned, and maintained. In computer vision, object detection pipelines are particularly sensitive to coverage gaps (e.g., occlusion and lighting), label noise, and train/test leakage induced by frame extraction and splitting practices [5–9].

A growing body of work also foregrounds the reality of "data work": dataset creation and maintenance are labor-intensive, under-incentivized, and often treated as secondary to model development, despite being central to system performance and reliability [2].

For high-stakes systems, small upstream issues can cascade into downstream failures ("data cascades"), motivating practices that surface assumptions early and make data pipelines auditable [10].

2.2 Documentation Standards for Reproducibility and Accountability

A second strand of work proposes documentation standards that make datasets and models more transparent and reproducible. Datasheets for Datasets and Model Cards formalize the idea that engineers should record provenance, intended use, collection procedures, and key statistics so that artifacts remain interpretable beyond the original development context [10, 11]. While such standards are increasingly referenced in professional ML practice, they are rarely assessed as first-class deliverables in undergraduate coursework; students may therefore learn how to run training scripts without learning how to specify and validate the dataset that the scripts consume.

2.3 Tooling and Assisted Annotation

Practical dataset engineering depends on toolchains. Annotation platforms such as CVAT reduce friction but do not guarantee semantic quality: label definitions, consistency checks, and coverage audits remain necessary. Recently, foundation-model-based tooling has made annotation more interactive; SAM-style segmentation assistance can accelerate mask and box creation, but it also introduces new risks (over-trusting suggestions, inconsistent "edge" decisions) that must be managed through guidelines and audits [12].

2.4 Scaffolding and Project-Based Computing Education

In computing and engineering education, active and project-based approaches are frequently used to teach complex, end-to-end workflows, but they require carefully designed scaffolding to prevent students from getting stuck in tooling friction and late-stage debugging [3, 13]. Our contribution sits at the intersection of these areas: we provide an exploratory analysis of evidence gaps and recurring problem categories of dataset-engineering pipelines in an undergraduate object-detection assignment and translate the observed problem categories into a compact "Teaching Playbook" of concrete, tool-aligned scaffolds.

3 Teaching Dataset Engineering in the GenAI Era

Generative AI changes what is easy for students to produce: polished text descriptions, boilerplate code, and superficially plausible reports can be generated quickly. This creates a risk of **false mastery**, where the final artifact looks complete while the underlying dataset decisions remain opaque or unverified. As a result, courses that aim to teach data-centric practice increasingly need assessment designs that privilege **evidence of work** over narrative fluency [14, 15].

We position "executable scaffolding" as a pragmatic response: instead of asking students to describe what they did, the course requires machine-checkable evidence (e.g.,

validated COCO exports, split manifests, dataset statistics, and audit samples). This shifts the burden of proof from prose to verifiable artifacts and aligns with a broader move toward process-oriented assessment and transparency in AI-enabled learning contexts [16].

Importantly, the goal is not to ban GenAI tools but to integrate them responsibly. Students can benefit from AI assistance for debugging, refactoring, or documentation drafting—if course design makes key dataset engineering decisions explicit and auditable (what was collected, how it was sampled, how labels were defined, and how leakage was prevented). In this framing, GenAI becomes a productivity aid, while the course still evaluates the student's engineering judgment through reproducible evidence (Table 1).

Table 1. Competency map for data-centric object detection coursework (M1–M6), aligned with evidence-first assessment.

Milestone	Competency	Required evidence artifact	Common AI-era failure/temptation	Scaffold
M1	Requirements & coverage planning	Capture plan + example frames	Generic plan text without coverage rationale	Checklist + peer review of capture plan
M2	Sampling & leakage control	Split manifest (by video/source) + pilot audit samples	Random shuffling that leaks near-duplicate frames	Source-based splitting rule + validation script
M3	Label definition & consistency	Label guide + consistency spot-check	Inconsistent edge cases hidden by fluent report text	Box-tightness audit + guideline examples
M4	Packaging & dataset validation	COCO export + schema/statistics report	"Looks fine" narrative while JSON is invalid	Automated sanity checks (IDs, flags, schema)
M5	Training control & traceability	Config snapshot + seeds + logs	Untraceable hyperparameter changes	Baseline presets + run template
M6	Evaluation & error analysis	Metrics + curated failure cases	Over-claiming based on a single metric	Structured error taxonomy + remediation prompts

4 Course Context and Instructional Design

The course design was implemented as a semester-long project in an undergraduate applied informatics course ($N = 89$ students). Students worked from a shared high-level specification but selected their own application domains. The workflow followed a standard data-centric pipeline:

- **Data Collection:** Recording original videos of target objects in diverse domains.
- **Annotation:** Extracting frames and annotating bounding boxes in CVAT (Computer Vision Annotation Tool), with optional use of assisted annotation tools like SAM [12].
- **Dataset Packaging:** Exporting annotations in COCO format and creating train/validation/test splits, aiming for documentation standards like Datasheets for Datasets [10].
- **Model Training:** Fine-tuning a baseline object detector (RF-DETR) and evaluating mAP (mean Average Precision).

Scaffolding included milestones (M1—M6), report templates aligned with a data dictionary, and technical demonstrations for tooling (Fig. 1; Table 2).

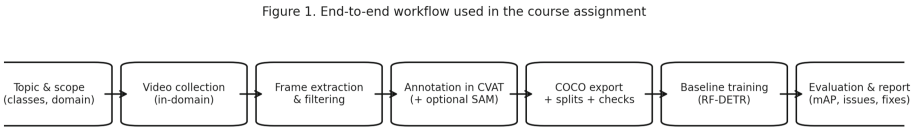


Fig. 1. End-to-end workflow used in the assignment (data collection → annotation → packaging → training → evaluation), with milestones M1—M6 mapped to artifacts.

Table 2. Instructional setting and required tools

Item	Value
Course format	Semester-long individual project (undergraduate applied informatics)
Participants	$N = 89$ students
Core tools	CVAT (annotation), COCO export, baseline training scripts
Optional tool	SAM-assisted annotation [12]
Primary outputs	COCO dataset package + report fields for reproducibility

5 Methodology

5.1 Study Type and Scope

This study is an exploratory analysis of course-generated artifacts from a single undergraduate cohort. It combines descriptive analysis of structured report fields with qualitative coding of student reflections. The purpose is not to estimate the full prevalence of student practices independently of reporting behavior, nor to test the causal effect of the proposed scaffolds, but to derive design-relevant insights for evidence-first teaching of data-centric AI workflows.

5.2 Research Procedure

We analyze project artifacts from the full cohort ($N = 89$) and a post-project technical reflection questionnaire from a subset of participants ($n = 37$). The report dataset includes design choices (domain, classes), procedural information (annotation tooling, export issues), and outcome-related fields (reported *mAP* and qualitative descriptions of difficulties and fixes). Structured fields were extracted into a single analysis table, while reflective fields were retained for qualitative analysis.

The unit of analysis is the student report (one report per student). For RQ1, the outcome is report-level evidence presence rather than externally verified execution of a practice. Accordingly, when an item is missing from a report, we interpret this as missing reproducibility evidence in the submitted artifact, not necessarily as proof that the corresponding step was not performed. This distinction is central to our pedagogical argument: if a dataset property is not made explicit in a reusable form, it cannot be reliably reviewed, reproduced, or audited.

To address RQ1 (Pipeline Evidence Gaps), we quantified the presence of key reporting elements required for reproducibility (e.g., split sizes, annotation counts, training details) and contrasted them with elements more directly coupled to tool execution (e.g., export format, architecture declaration).

To address RQ2 (Problem Categories and Remediation), we applied inductive thematic coding to the reflective fields `issues_taxonomy` and `fixes_applied`. Coding proceeded in three passes. In the first pass, an initial set of issue labels was generated directly from student wording. In the second pass, related labels were consolidated into broader categories representing recurrent dataset-engineering problems. In the third pass, a randomly selected subset of reports was revisited to check category stability, boundary consistency, and fit to the original text. Ambiguous cases were resolved conservatively by returning to the full report context and preferring broader categories when the evidence did not support finer distinctions. We report the resulting categories as exploratory analytic constructs intended to inform instructional design rather than as coder-independent ground truth.

Taken together, RQ1 and RQ2 separate two common breakdown types: (i) missing reproducibility-critical evidence in submitted reports and (ii) recurring technical problem categories described by students during training and evaluation. The teaching playbook derived from these analyses is therefore presented as a design output grounded in course evidence, not as an experimentally validated intervention model.

6 Data Sources and Availability

The study draws on two data sources produced as part of regular course delivery. First, final project reports were collected from 89 students. Each report followed a structured template and contained both declared evidence fields (e.g., split sizes, export format, model details) and free-text reflections on issues encountered and fixes applied. For analysis, structured fields were extracted into a tabular dataset, while reflective fields were retained for qualitative coding. Second, a separate post-project reflection questionnaire was completed by 37 students and was used to contextualize selected observations about tooling, difficulty, and annotation strategy.

Due to institutional privacy constraints and the possibility of re-identification from free-text responses, neither the raw reports nor the raw questionnaire responses are publicly released. However, the student assignment description and the report-extraction data dictionary used in this study are publicly available as supplementary materials.

7 Results

7.1 RQ1: Pipeline Evidence Gaps

The analysis revealed a clear contrast between evidence elements closely coupled to tool execution and those primarily serving reproducibility and later auditability. Tool-coupled elements were reported relatively consistently: 75 of 89 students (84.3%) explicitly reported the COCO export format, and 81 (91.0%) reported the model architecture, both of which were directly tied to the provided workflow and training scripts.

By contrast, reproducibility-critical metadata that were less directly enforced by the toolchain were frequently absent from submitted reports. As shown in Table 3, only 22 students (24.7%) reported exact split sizes for the training, validation, and test sets, and only 20 students (22.5%) reported the total annotation count. This pattern is consistent with the interpretation that students more reliably document what is necessary to keep the pipeline running than what is necessary to make dataset construction transparent and reproducible. Because our unit of analysis is the report, these omissions should be interpreted as missing submitted evidence rather than definitive absence of practice. Pedagogically, however, such omissions remain consequential: if key dataset properties are not explicitly documented, they cannot be reliably reviewed, reproduced, or reused.

Table 3. Pipeline completion indicators (share of students reporting each evidence element; $N = 89$)

Stage	Indicator	Students reporting
Splits	All split sizes provided (`n_train, n_val, n_test`)	22/89 (24.7%)
Annotation	Total annotations reported	20/89 (22.5%)
Packaging	COCO export format reported	75/89 (84.3%)
Packaging	Export issues reported	70/89 (78.7%)
Training	Model architecture reported	81/89 (91.0%)
Evaluation	mAP@50 reported	65/89 (73.0%)

7.2 RQ2: Recurring Problem Categories

In the reflective fields, students recurrently described a limited set of issue types concentrated in data quality and packaging rather than in model architecture. We report these as recurring problem categories observed in student reflections, not as a comprehensive taxonomy of object-detection issues.

Survey responses from the subset questionnaire (n = 37) further suggest that students did not simply accept assisted annotation outputs uncritically. When asked about their strategy for annotating difficult objects, such as blurred or partially occluded instances, 26 of 37 respondents (70.3%) reported reverting to manual correction for greater precision, whereas only 4 of 37 (10.8%) reported accepting SAM-assisted outputs without modification. In this limited sense, the responses indicate that many students adopted a human-in-the-loop approach when automation became unreliable.

Across the reflective fields, the most frequently observed categories were occlusion and complex scenes (22 reports), low contrast or lighting limitations (14 reports), and COCO export or `iscrowd` semantics (13 reports). These categories point to a recurring tension between apparently successful pipeline execution and underlying data-quality fragility. The fixes students reported were often reactive and local, such as manually correcting boxes, editing export settings, or adjusting training stability parameters. This pattern supports the need for earlier, more explicit scaffolding that surfaces predictable problems before model training begins.

8 Derived Teaching Playbook for Evidence-First Scaffolding

Based on the recurrent issue categories observed in RQ2, we derive a compact teaching playbook that links common symptoms to plausible causes and concrete scaffolding responses. The purpose of this playbook is not to claim validated intervention effectiveness within the present study, but to convert observed course-level patterns into reusable instructional design knowledge. In this sense, the playbook functions as a design artifact grounded in report-based evidence from one cohort (Table 4).

Table 4. Teaching playbook: mapping observed problems to causes and scaffolding interventions

Problem Category	Likely Cause	Teaching Scaffold (Intervention)
Occlusion / Complex Scenes	Insufficient coverage of occluded cases; ambiguous "object vs. background" boundaries.	Hard Negative Audit: Mandatory M2 pilot check specifically for occluded instances; peer cross-review of 50 samples for box tightness.
Low Contrast / Lighting	Data collected in narrow lighting conditions (single session).	Lighting Matrix: Requirement to record source video in at least 3 distinct lighting conditions; frame quality screening step.
COCO Export / IDs	Incorrect export flags; mismatch between class list and COCO category IDs.	Automated Sanity Check: Provide a script for M4 that validates JSON schema, `iscrowd` flags, and ID sequences before submission.

(*continued*)

Table 4. (*continued*)

Problem Category	Likely Cause	Teaching Scaffold (Intervention)
Frame Leakage	Frame extraction from contiguous video segments split randomly (train/test overlap).	Source-Based Splitting: Enforce split-by-video rules (e.g., Video 1 = Train, Video 2 = Test) rather than random frame shuffling.
Hardware Limits	Over-ambitious image size or batch size causing OOM errors.	Resource Budget: Provide safe baseline presets (batch/size) and a troubleshooting decision tree for resource limits.

9 Discussion and Conclusion

The findings suggest that undergraduate students can navigate the mechanics of an end-to-end object-detection workflow while still under-documenting the dataset properties needed for reproducibility and later audit. In our cohort, tool-coupled deliverables were much more consistently reported than evidence elements such as split sizes or annotation counts. This does not demonstrate that students failed to perform these steps; rather, it highlights a course-design problem in which important dataset decisions remain insufficiently visible in submitted evidence.

A second observation is that the recurrent issue categories identified in student reflections were concentrated in predictable data-centric problem areas, including occlusion, lighting variation, export semantics, and leakage risks. These issue types are not novel at the level of computer-vision research. Their contribution here is pedagogical: the study shows how such problems surface in undergraduate end-to-end projects, how they appear in student-produced reports and reflections, and how they can be translated into explicit scaffolding actions embedded in milestones and assessment requirements.

We therefore frame the main contribution of the paper as design knowledge for computing education rather than algorithmic novelty. More specifically, the paper operationalizes evidence-first dataset instruction by distinguishing between narrative completion and machine-checkable evidence, and by mapping recurrent student-reported problems to concrete preventive scaffolds such as split manifests, validation scripts, and targeted audit checkpoints. This pedagogical operationalization is, in our view, the main value of the study.

9.1 Design Principles for Evidence-First Dataset Engineering Instruction

From the course experience, we derive five design principles that may transfer to similar data-centric AI settings:

- *Grade evidence, not prose.* Require machine-checkable artifacts, such as manifests, statistics, and validation outputs, that cannot be satisfied by fluent narrative alone.

- *Shift checks left.* Place pilot audits early so predictable quality issues are discovered before training.
- *Automate the brittle parts.* Validate formats, identifiers, and schema automatically to reduce time spent on procedural debugging.
- *Make leakage prevention explicit.* Treat split logic and source provenance as first-class deliverables rather than implicit assumptions.
- *Teach semantic quality through small audits.* Use repeatable checks for box quality, lighting diversity, and difficult cases to connect data decisions to downstream behavior.

These principles should be read as transferable design hypotheses derived from one course setting, not as experimentally validated prescriptions.

9.2 Research and Practice Agenda

Future work should extend this line of inquiry in three directions. First, the proposed scaffolds should be evaluated across multiple course iterations using stronger artifact-based measures and explicit learning-outcome instruments for data-centric competencies. Second, comparative studies could examine different scaffolding intensities, for example manual review versus automation-heavy designs. Third, the same evidence-first logic should be tested in other modalities, such as text, tabular, or multimodal projects, where polished documentation may similarly obscure weak process evidence.

9.3 Threats to Validity

This study has several limitations. First, it is based on a single course implementation at one institution and one task family (object detection), which limits generalizability. Second, the primary unit of analysis is the student-authored report. As a result, missing items in Table 3 should be interpreted as missing submitted evidence rather than verified absence of practice. This affects construct validity, but it is also central to the pedagogical argument: in data-centric AI instruction, undocumented dataset decisions are themselves an important educational problem.

Third, the qualitative coding of reflective fields was conducted primarily by one researcher and only partially re-checked on a subset of reports. Although we used iterative category refinement and conservative treatment of ambiguous cases, we do not claim coder-independent reproducibility or formal inter-rater reliability. The reported categories should therefore be understood as exploratory analytic constructs intended to support instructional design.

Fourth, the proposed teaching playbook is derived from observed issue patterns but is not evaluated in the present study as an intervention. We therefore do not claim that the suggested scaffolds have been causally shown here to improve learning outcomes or project quality. Finally, some findings depend on the specific toolchain used in the course (CVAT, COCO export, and provided training scripts). While several issue categories are likely to transfer more broadly, some procedural details may be format- or tool-specific.

Declaration of Ethical Conduct. No video or audio recordings of participants were made for research purposes. The study uses course artifacts (final reports) and an optional post-project reflection questionnaire collected as part of regular teaching practice. For analysis, data were de-identified at the point of processing; only aggregated results are reported, and no attempt is made to link published outcomes to individual participants.

Raw artifacts are stored on institution-managed systems with access restricted to the course staff and researchers involved in this study. The raw datasets are not publicly released due to privacy constraints and the risk of re-identification from free-text content. The report-extraction data dictionary and report template used for structured extraction are publicly available (see Supplementary Materials).

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials. The report-extraction data dictionary used to structure the analysis is publicly available at https://doi.org/10.5281/zenodo.18482322 to support transparency and facilitate replication of the analytic procedure in comparable courses.

References

1. Whang, S.E., Roh, Y., Song, H., Lee, J.-G.: Data collection and quality challenges in deep learning: a data-centric AI perspective. VLDB J. **32**(4), 791–813 (2023)
2. Sambasivan, N., Kapania, S., Highfill, H., andorng, V.A.D., Batlajery, P.: 'Everyone wants to do the model work, not the data work': data cascades in high-stakes AI. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–15 (2021)
3. Krajcik, J.S., Blumenfeld, P.C.: Project-based learning. In: The Cambridge Handbook of the Learning Sciences, pp. 317–333. Cambridge University Press (2006)
4. Lanubile, F., Martínez-Fernández, S., Quaranta, L.: Teaching MLOps in higher education through project-based learning. In: 2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET), pp. 95–100 (2023). https://doi.org/10.1109/ICSE-SEET58685.2023.00015
5. Saleh, K., Szénási, S., Vámossy, Z.: Occlusion handling in generic object detection: a review (2021). https://doi.org/10.1109/SAMI50585.2021.9378657.
6. Premakumara, N., Jalaian, B., Suri, N., Samani, H.: Enhancing object detection robustness: a synthetic and natural perturbation approach. https://arxiv.org/abs/2304.10622. (2023). https://doi.org/10.48550/arXiv.2304.10622
7. Xia, Q., Hu, C., Lee, F., Chen, Q.: Gradient optimization for object detection in learning with noisy labels. Appl. Intell. **54**, 4248–4259 (2024). https://doi.org/10.1007/s10489-024-05357-6
8. Glazner, N., Tsfaty, N., Shalev, S., Weizman, A.: Find the leak, fix the split: cluster-based method to prevent leakage in video-derived datasets. arXiv:2511.13944. (2025) [Online]. https://arxiv.org/html/2511.13944
9. Ramos, P., Ramos, R., Garcia, N.: Data Leakage in Visual Datasets. [Online]. https://openaccess.thecvf.com/content/ICCV2025W/Findings/papers/Ramos_Data_Leakage_in_Visual_Datasets_ICCVW_2025_paper.pdf (2025)
10. Gebru, T., et al.: Datasheets for datasets. Commun. ACM. **64**(12), 86–92 (2021)
11. Mitchell, M., et al.: Model cards for model reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT '19), pp. 220–229, New York, NY, USA (2019). https://doi.org/10.1145/3287560.3287596

12. Kirillov, A., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026 (2023)
13. Prince, M.: Does active learning work? A review of the research. J. Eng. Educ. **93**(3), 223–231 (2004). https://doi.org/10.1002/j.2168-9830.2004.tb00809.x
14. Scarfe, P., Watcham, K., Clarke, A., Roesch, E.: A real-world test of artificial intelligence infiltration of a university examinations system: a 'Turing Test' case study. PLoS One. **19**(6), e0305354 (2024). https://doi.org/10.1371/journal.pone.0305354
15. OECD: OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education. OECD Publishing, Paris (2026)
16. Miao, F., Holmes, W.: Guidance for Generative AI in Education and Research. United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, France (2023)

From Disruption to Evolution: The Next Era of Educational Transformation

Boris Aberšek[1](✉), Metka Kordigel Aberšek[2], and Maja Kerneža[2]

[1] Faculty of Natural Science and Mathematics, University of Maribor, Maribor, Slovenia
Boris.abersek@um.si
[2] Faculty of Education, University of Maribor, Maribor, Slovenia

Abstract. In the last decades the educational landscape is undergoing a radical transformation, primarily driven by the emergence of Artificial Intelligence (AI) and the profound societal shifts it heralds. Human societies are moving from a singular, human-centric understanding of the world and intelligence to a dual reality encompassing both physical and cyber realms parallel to human and artificial intelligences. This necessitates a fundamental rethinking of education, moving beyond incremental improvements to embrace a period of intense societal and educational evolution, akin to waves of technological disruption. The contribution outlines six technological waves of change and several educational revolutions, from Education 2.0 and 3.0, driven by Web 2.0 and 3.0, to Education 4.0, which emphasizes holistic development and social transformation. The current sixth wave of change points towards Education 5.0, characterized by "generative pedagogy" inspired by Noam Chomsky's generativism. This new paradigm focuses on co-designing and co-implementing learning experiences in collaboration with Generative AI (GenAI), transitioning to a symbiotic relationship between human and artificial intelligence.

Contribution concludes with thesis that the future of education (Education 5.0) lies in leveraging GenAI to foster a human-centric, collaborative learning ecosystem for a Society 5.0, a super-smart society.

Keywords: Educational Transformation · Generative Artificial intelligence · Creative destruction · Society 5.0 · Technological impact

1 Introduction

Digital technologies have been part of the educational landscape in various forms for many decades. In the last decade, however, there have been intense paradigm shifts, mainly due to the emergence of AI. However, there are two fundamental post-2022 lessons that generate these shifts (Aberšek, 2024a, 2024c):

- the realization that the *world is no longer one, there are two worlds*. Ours still exists, and the physical world is evolving in parallel with the digital, cyber, artificial intelligence world, and

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 27–38, 2027.
https://doi.org/10.1007/978-3-032-32104-6_3

- the realization that intelligence is no longer one, there are diametrically two intelligences. In addition to our human intelligence, artificial intelligence is also being developed more and more intelligence, not to mention all the transient improved hybrid forms of intelligence, which are enabled to us, for example, by neuroscientists and create a new type of beings/entities - cyborgs.

As a result, therefore, we can no longer speak of a system of continuous evolutionary improvements, but of an intense transformation of society, as well as of education. Technology has appeared and still appear in a kind of waves in modern society, which is described in more detail by Mustafa Suleyman in his book *The Coming Wave* (Suleyman & Bhaskar, 2023). Thus, Suleyman points out that the world is not ready for the wave of powerful new artificial intelligence technologies and has identified the "problem of containment" and negation of reality as a challenge for our age. Suleyman was far from the first to use this thesis to describe *the upheaval* following the introduction of disruptive new technology in society. The wave engulfs both spectrums of diametrical possibilities, proponents and opponents: for some, it indicates purification and renewal, a time of new beginnings, and for others, it suggests that we are being overwhelmed by a powerful force that will overwhelm us, enslave us, and ultimately drown us in it. We are certainly entering a time of drastic change, which is most easily described as chaos. *While chaos is a great state for creating change, it's important to know exactly what we want, because the time and possibilities for corrections is not unlimited.* Let us try to briefly analyze this using the Kübler-Ross curve of change (Kübler-Ross, 1996).

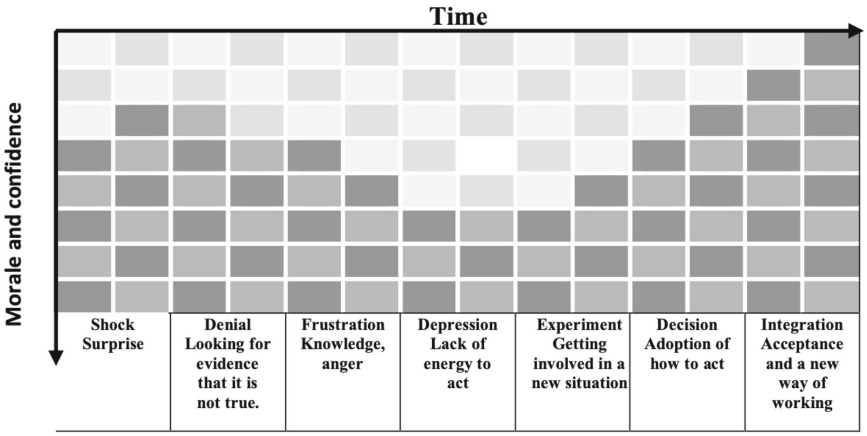


Fig. 1: Kübler-Ross change curve (adapted from Kübler-Ross, 1969)

In 1969, Elisabeth Kübler-Ross presented a curve of five stages of change in the occurrence of various changes (losses, shocks or crises), or simply the "5 stages" which is symbolically shown in Fig. 1. These stages—*denial, anger, bargaining, depression, and acceptance*—reflect shared emotional responses to change, loss, or shock. Overall, the Kübler-Ross change curve in Fig. 1 provides a framework for understanding emotional progression through change (Aberšek, 2025). He points out that the process can be

complex and cyclical, but each stage is part of an ongoing journey to acceptance and adaptation. And surely we are all the only ones who believe that the phenomenon of AI, especially the phenomenon of the GenAI, is really a change that brings drastic and radical changes to our thinking and actions.

Here, too, as a starting point for further reflections, we can once again repeat the oft-heard saying: *After 2022, according to our general awareness of the existence of large language models - LLMs (Large Language Models) and GenAIs, nothing will be the same again!*

1.1 Creative Destruction

Joseph Schumpeter used the term "*creative destruction*" to describe the dynamics of technological disruption that have driven the so-called waves of innovation (technological or industrial revolutions) in our modern era (McKraw, 2007). Each wave in the series is characterized by a set of technologies that were leading technologies of their time, briefly presented in Table 1. With each new wave of technological change, the dynamics *of creative destruction brought* with it new ways of working, living and, of course, educating as an integral part of society (Suleyman, Bhaskar, 2023).

2 Technological Transformation of Contemporary Education

In the last decade, we have experienced two upheavals, first the pandemic and then the realization that humans are not the only intelligence. We are again at a crossroads where we will have to replace what we know well for what is foreign to us and start the K-R curve of change again from the beginning. We have entered the sixth wave of change (see Table 1), we have entered the *third educational revolution*, which requires us to radically rethink, to radically transform everything we have known so far.

And the problem we face is that we don't have a lot of time to achieve these changes. Dozens of previous changes have only been transformed into years, or even months. The world began to rotate at such a speed that it is difficult for a person's conception of time to rationally follow these changes. The modern development of selection is shown in Fig. 2.

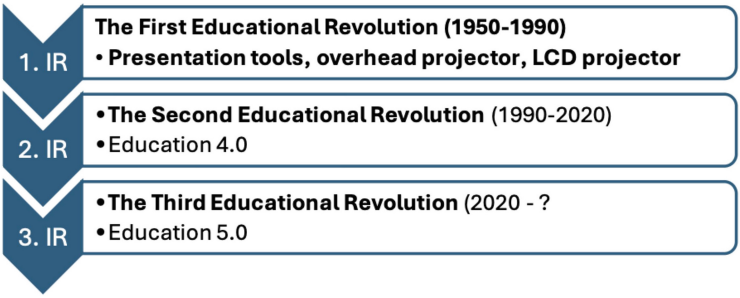


Fig. 2: Revolutions in education

Table 1: Economic, technological and social development

Economic Development (Capital)	**Technological Development**		**Social development Education**
	Pre-technological period Development of tools, weapons		
Prehistoric evolution	Science (curiosity) as a driver of development Technologies at the service of science A telescope, a microscope, Technologies in the service of warfare gunpowder, weapons		Educational evolution Didactics approx. 500 BC First educators
			Renaissance Foundations of education School as an institution
Artisanal production develops into manufacturing	Use of energy resources and related technologies Windmills and water sources,...		Modern society
The Industrial Revolution (1700-1830)	1st Technological Wave (1785-1845)		
2. The Industrial Revolution (1800 - 1900)	2nd Technological Wave (1849 - 1900) 3rd Technological Wave (1900-1950)		1. Educational revolution (1950 - 1990
3. Industrial Revolution (1950 -2010)	4th Technological wave (1950-1990)		1. Educational revolution (1950 - 1990)
4. The Industrial Revolution (2010 -	5th Technological wave (1990 - 2020		2. Educational revolution (1990 - 2020)
	6th Technological Wave (2020 - The birth of GenAI and the embryos of Artificial General Intelligence (AGI) Development of AGI		3. Educational Revolution (2020 -
Profit ←	Direction of agents	The direction of conscious-sentient AI (AGI)	→ **Welfare of society**

2.1 Educational Revolution

Technology began to permeate the educational process in the new millennium, and both students and teachers began to use technology in fundamental ways (also known as Education 2.0). Education 3.0 emerged with the advancement of technology, especially with the widespread adoption of Web 2.0, but even more so in the era of user-generated Web 3.0. This gave students access to their own sources of information, the possibility of e-learning, and a platform to communicate with teachers and other students. Education became more networked, and pupils had their direct connection to a range of different sources of knowledge, rather than being focused on exchanges between pupils and teachers.

Forerunner of Contemporary Education Paradigm–First Educational Revolution: If we want to create proper human-AI communication, humans need to be able to express themselves in a language that AI will understand. That is why it is extremely important to harmonize at the beginning of the learning process (Suppes, 1966). Many years ago, Lande postponed the use of algorithms and algorithmic thinking, which allows us to express our thoughts unambiguously and systematically (Landa, 1976). Cyber pedagogy

was also based on algorithmic thinking. Let's take a brief look at its starting points (Aberšek, Borstner, Bregant, 2014).

Second Educational Revolution:

Education 4.0 is the use of new technologies and new way of thinking to provide more humanized teaching, with an emphasis on the social and emotional development of students and solutions that improve life in society.

Table 2: Generativism

Original Chomsky	Education 5.0
Key concepts of generativism:	**Key concepts of generativism:**
1. Innate language competence: Chomsky argued that people are biologically inclined to learn a language because they have an innate mental grammar.	*1. Innate competence:* We start from the thesis that people are biologically inclined to learn because they have an innate mental structure and grammar.
2. Universal grammar: This refers to fundamental, universal principles that are common to all languages.	*2. Universal Features:* This refers to the fundamental, universal principles that are common to all knowledge acquisition.
3. Transformation-generative grammar (TGG): This model explains how users of a language can create an infinite number of grammatical sentences from a finite set of rules.	*3. Transformational-generative grammar:* This model explains how users can create an infinite number of grammatical sentences from a finite set of rules.
4. Deep Structure vs. Surface Structure: TGG distinguishes between the basic meaning (deep structure) and grammatical form (surface structure) of a sentence.	*4. Deep Structure vs. Surface Structure:* TGG distinguishes between the basic meaning (deep structure, understanding) and the superficial form of knowledge.
5. Competence vs. Performance: Chomsky focused on language competence (knowledge of grammar) rather than performance (actual use of the language).	*5. Competence vs. Performance:* Let's focus on competence (knowledge) and not on performance (actual use of knowledge).
Generative pedagogy:	**Generative pedagogy:**
1. Focus on the rules: Generative pedagogy emphasizes understanding and applying basic grammatical rules, not just memorizing them.	*1. Focus on the rules:* Generative pedagogy emphasizes understanding and applying basic rules and basic knowledge, not just memorizing them.
2. Emphasis on creativity: It encourages students to create new sentences based on their understanding of the rules, which fosters a sense of linguistic creativity.	*2. Emphasis on creativity:* It encourages students to create new knowledge based on their understanding of the rules, which fosters a sense of creativity.
3. Learner-centered approach: It recognizes the active role of the learner in the development of their own language skills.	*3. Learner-centered approach:* It recognizes the active role of the student in the formation of his own knowledge.
4. Contrast with Behavior: Unlike behavioral approaches that focus on responding to stimuli, generative pedagogy emphasizes the thought processes involved in language learning.	*4. Contrast with Behavior:* Unlike behavioral approaches that focus on responding to stimuli (behaviorism), generative pedagogy emphasizes the thought processes involved in learning.
How Generativism Affects Pedagogy:	**How Generativism Affects Pedagogy:**
1. Understanding Language Learning: Generative theory offers insight into how students acquire language and informs teaching methods.	*1. Understanding Learning:* Generative theory offers insight into how students acquire knowledge and informs about teaching methods.
2. Development of effective materials: Directs the development of language learning materials that focus on basic grammatical structures and rules.	*2. Development of effective materials:* It guides the development of learning materials that focus on fundamental structures and rules.
3. Fostering a deeper understanding: With an emphasis on the generative capacity of language, it encourages students to go beyond memorization by heart and develop a deeper understanding of language.	*3. Fostering a deeper understanding:* With an emphasis on the generative capacity of language, it encourages students to go beyond memorization by heart and develop a deeper understanding of the acquired knowledge.

Building on the lessons learned over this decade, the pandemic has shown that people need to be prepared for adversity, emotionally intelligent people who know *how to turn*

digital transformation into a tool for social transformation. And that's exactly what Education 4.0 is: a chain that connects digital and technological knowledge with human social and emotional skills to promote well-being. Let us highlight only two components:

1. *The Role of Education in Holistic Development*: Education 4.0 emphasizes the development of an individual's holistic personality, not only through cognitive skills, but also emotional intelligence, social awareness, and ethical values.
2. *The Impact of Society on Integral Development*: The goal of Society 5.0 to improve the general well-being of people encourages education systems to address broader societal challenges. This means that Education 4.0 curricula also include topics related to sustainability, ethics, social justice, and responsible use of technology, in line with the principles of Society 5.0.

If the fifth wave of changes in the field of education was marked by the philosophy of Education 4.0, based primarily on digital pedagogy, the sixth wave has also begun in education, which requires a comprehensive transformation of education. It requires both reflections on the role of technology and changes in how to impart knowledge by incorporating these technological possibilities, as well as methods and strategies of knowledge transfer. We need to talk about Education 5.0, which must be based on new pedagogical platforms, digital pedagogy must be replaced by a new, generative pedagogy, which can also be called universal generativism in short.

The term generativism is a well-known term that was introduced into the linguistic sciences by Naom Chomsky, back in the 1950s (Chomsky, 1966). Chomsky's generativism also known as generative grammar, is a language theory that proposes that language is governed by a set of innate, universal rules that allow people to create an infinite number of grammatical sentences. This theory, developed by Noam Chomsky, shows that people are born with a "language-learning device" and a universal grammar that allows them to learn a language naturally, and not just through imitation or environmental input. Generative grammar is a special approach to language modelling developed by Noam Chomsky that uses rules to create or build sentences based on a grammar model that is as simple as possible while still taking into account all the complexity of the language.

Chomsky's generativism, specifically transformational-generative grammar (TGG), revolutionized linguistics by proposing that language learning was driven by innate, universal grammar. The generative pedagogy inspired by this theory emphasizes understanding the fundamental linguistic structures and rules that govern language, rather than memorizing grammatical rules.

It focuses on how students internalize and use these rules to create new sentences, with an emphasis on the creative and generative capacity of language. In short, Chomsky's generativism, we can transform it into Education 5.0. This comparison is shown in Table 2. In this way, we could identify three key areas,

1. key concepts of generativism;
2. generative pedagogy and
3. how Generativism affects pedagogy.

3 Integration of Technology in Education 5.0

When we talk about technological integration into educational processes, we can point out:

- *Education and technology*: Education 5.0 takes advantage of technological advances to personalize learning experiences, offer interactive content, and facilitate lifelong learning. This reflects the Society 5.0's vision of technology as a factor that enables a better quality of life and social progress.
- *Society and Technology*: The technological approach of Society 5.0 requires a population that is trained to use and innovate with emerging technologies. Education 5.0 plays a key role in equipping individuals with the digital literacy and technical knowledge needed to actively participate and contribute to the progress of Society 5.0.

To sum up, Education 5.0 and Society 5.0 are intertwined in their efforts to shape a future that combines technological innovation, human well-being and sustainable development. Education 5.0 provides individuals with the skills and mindset necessary to actively participate in the technologically enhanced ecosystem of Society 5.0, while Society 5.0 provides a context in which these skills can thrive and contribute to a more successful and inclusive society, Education 5.0 from a global, especially social and economic point of view, while from a slightly more local, intra-social and which emphasizes above all holistic, the social aspect of development. It is not tied to economic development, but to how to achieve this economic development through a knowledge-based society (Aberšek, Kordigel Aberšek, 2025).

3.1 The AI Education Process: From Digital Pedagogy to Generativism

As leading technologies and learning environments change, so do the methods of work and learning content, as well as the learning process itself, methods and strategies of learning and teaching. By replacing a network of static web pages with information flowing in one direction, users in Web 2.0 can collaborate and share information, comment, create websites, send opinions, publish blogs, and communicate. Unfortunately, however, many educators still operate within operating models that could be called Education 1.0 (using the philosophy of Web 1.0), that is, more analogous behaviourist education than Education 2.0, digital, constructivist and connectivistic education (Flogie, Aberšek, Pesek, 2025; Flogie, Aberšek, 2019; Aberšek, 2025).

3.2 Digital Pedagogy

Effective teaching must always be based on pedagogy. As we mentioned the learning process must be influenced by several factors, including content knowledge, organization of knowledge, competence development, intentional goal-oriented feedback, self-directed learning, learning environment and of course motivation for learning. *Digital pedagogy*, as the foundation of Education 4.0, is primarily oriented on how to use digital technologies to support and facilitate the learning and teaching process (Ambrose et al., 2010; Sharples, 2019; WEF, 2023; Aberšek, 2025).

Connectivism was presented by Siemens and Downes in two articles, "*Connectivism: Learning as Networking* Creation" (Siemens, 2005a, 2005b) and "*Introduction to Connective Knowledge*" (Downes, 2005). Connectivism is based on the idea that humans learn when we make connections between different "nodes" of information and create connections by forming knowledge. It highlights the important role that technology plays in the learning process and the ease with which technology allows learners to access information in the digital age. Connectivism is therefore oriented to student, as it learning responsibilities shift from teacher to student, so that students can latter understand the information they discover in the digital world. Such interaction in the digital world means that connectivism also treats learning as a social process, not just as a process, that is purely internal and individual. Connectivism treats this process as something that is distributed and dynamic. Connectivism has therefore presents the foundations for approaches that can also be called *education 4.0* enabled by artificial intelligence. In 1966, Patrick Suppes (Suppes, 1966), a pioneer of computer learning, envisioned a future in which students would have access to unlimited quantities of knowledge with the help of computer tutors (Aberšek, Flogie, Pesek, 2023; Aberšek, 2025).

3.3 Generativism

Today, it is more and more clearly that the future of education must be defined also by cooperation with the GenAI in all areas, and that there will have to be an intense collaboration between both types of intelligence, human and artificial to create some new social order including education (Aberšek, 2025; Aberšek, 2024b; Pratschke, 2024). GenAI is ready to change practice, including education, in all areas. As this practice changes and GenAI becomes part of the research and development process, it will also transform our society with education at the beginning. But what means this in practice? How can we work with the GenAI to shape learning and teaching? *Generativism describes a symbiotic approach to designing and delivering learning in collaboration with the GenAI.* Generativism is based on *the principle of learning as a process* and on some of the most important and influential learning theories and approaches of our time (Aberšek, 2025):

- *constructivism,* which point out the need for active learning to construct meaning to emphasises the role of the digital network, in which learners are active actors in this learning;
- *social learning*, emphasizing the importance of cooperation;
- *experiential learning (experiance based learning)*, which emphasizes the role of learning by doing and thinking about that experience; and *generative learning*, which presents learning as a process of creating meaning through the use of schemas.

The implementation of the GenAI in generativist education process is a defining feature of generativism as a practice, whereby knowledge will be generated while working in collaboration with the GenAI, through learning activities that are co-designed, facilitated and evaluated by the GenAI. Generativism can be summarized as (Aberšek, 2025):

- co-design of the learning experience in collaboration with the GenAI;
- co-implementation of learning activities and assessment in cooperation with the GenAI; and

- assessment of learning as a process in collaboration with the GenAI.

Generativism uses established digital frameworks as a foundation that provides the methodological approach, we need to design the learning process in and for the learning environments supported by the GenAI. Students must be able to work with the AI, not only as a tool for effectiveness in the workplace, but as a means of research and knowledge creation both in their field of study and in everyday life. We may be asked a series of questions, such as (Aberšek, 2025):

- How can we design educational experiences that foster the critical meaning-making that is needed when working with this emerging and still deeply flawed technology?
- How do you leverage the generative and social privileges of the GenAI while still instilling core societal values alongside the skills of critical thinking, creativity, and exploration?

Digital and generative transformation are undoubtedly changing the skill set needed to succeed in the world. The standard digital ecosystem of a physical institution (schools, colleges) includes learning management systems (LMS), which is sometimes referred to as a *virtual learning environment* (Aberšek, Borstner, Bregant, 2014). It also includes any related external tools to which the institution subscribes.

However, if we start from the definitions of Education 4.0 and the Taxonomy for Teaching in the Future (WEF 2023), we could use this as the starting point for the Education 5.0 taxonomy (Aberšek, 2025).

3.4 Taxonomy of Learning for Education 5.0

In the proposed taxonomy of Education 5.0, the core levels 1 and 2 remain identical to those of Education 4.0, with a slight modification of Level 3, which covers the concepts that must follow the society of the 6th wave of technological change and must be understood as the capabilities that Education 5.0 seeks to build, as shown in Fig. 3.

Overall, however, Level 4 is added, which represents an actualization of the changes that have and will be, or should be, in the education system, especially at the strategic/paradigmatic level in the next few years.

4 Conclusion

When we talk about generative GenAI tools for education, they can be divided into two groups:

1. the first group includes GenAI tools that increase *efficiency and are focused on the teacher*. Their fundamental goal is to help the teacher and save time by automating tasks, as in the case of generators for contents and learning analytics, and
2. the second group is related to effectiveness, which is focused on how *to use GenAI tools to teach more effectively* and achieve better learning outcomes for students (generativism).

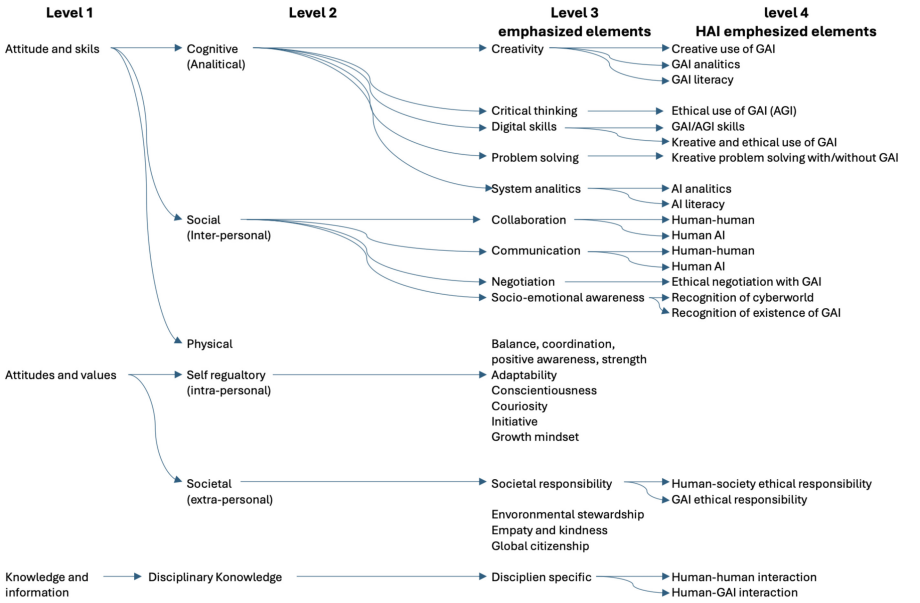


Fig. 3: Education Taxonomy 5.0 (Aberšek, 2025)

When it comes to the efficiency of administrative work, AI assistants and agents have a lot to offer. From writing to editing, from creating presentations to analyzing data. An AI (GenAI) helps also with the tasks that teachers perform as part of their work process. Certainly, however, the existence of people in the modern world will undergo major changes, just as we will have to find the place and role of the GenAI in modern society, we will also have to redefine the role of man in it (Aberšek, 2025). Therefore, we need to approach GenAI integration differently than we did in the past, mainly because the AI ecosystem not only improves traditional education but allows us to create something completely new. *And whether we want it or not is no longer a question at all. The era of shock, frustration, and depression is long gone.* But, nevertheless, the consideration of effectiveness, which suggests that GenAI tools can optimize the learning process, needs to be supplemented with some considerations. Research shows that while AI assistants may be attractive, they may not aid learning and, conversely, learners' competencies may regress rather than progress. In this context, surely the first thing to mention is *the cognitive laziness*[1] that will surely follow when AI shares space – both physical and virtual – with classmates (humans). However, once teachers acquire didactic competencies related to the use of artificial intelligence based on modern pedagogy (generativism), the disruption in education that many have predicted will be much less likely (Aberšek, 2025).

[1] *Cognitive laziness* forces us to avoid looking for possible solutions (staying on Kahneman's intuitive System 1, and not turning on System 2, an in-depth, statistical way of thinking) and thus closing the door to creativity. Cognitive laziness causes mild cognitive impairment, including problems with memory, language, and judgment.

GenAI is still an emerging technology in the early stages of development, which means that it needs to be handled with care. These foundations can be used as a starting point to build a new model of education that is human-centric but defined by collaboration with generative AI. It offers us the potential to increase our own intelligence in previously unimaginable ways, while awakening our deepest fears of becoming obsolete. It is the only promising transition to Society 5.0, to a super-smart society (Flogie, Aberšek, Pesek, 2025).

The implications for education will be profound. The teacher will no longer be the sole bearer of knowledge, a written assignment as proof of what has been learned is no longer feasible, and the classroom is no longer the center of activity (Aberšek, 2025).

And finally, ChatGPT opinion about future of Education:

The future of education can be summarized as: Harnessing the potential of Generative AI to create dynamic, collaborative learning experiences that emphasize critical thinking and holistic development (ChatGPT).

References

Aberšek, B.: The process of education with AI: from digital to generative pedagogy. J. Balt. Sci. Educ. **24**(6), 1012–1016 (2025)

Aberšek, B., Kordigel Aberšek, M.: Filozofska analiza mehanizmov kognitivnega modeliranja in umetne inteligence v izobraževanju, [Philosophical Analysis of Mechanisms of Cognitive Modeling and Artificial Intelligence in Education]. Univerza v Mariboru, Univerzitetna založba (2025)

Aberšek, B.: Philosophy of Mind, Artificial Intelligence and Existentialism. Cambridge Scholars Publishing (2024a)

Aberšek, B.: Transformation of society and education through time. Probl. Educ. 21st Century. **82**(6), 782–784 (2024b). https://doi.org/10.33225/pec/24.82.782

Aberšek, B.: From chaos to the absurd: existentialism for the 21st century. Philosophies. **9**(6) (2024c). https://doi.org/10.3390/philosophies9060168

Aberšek, B., Flogie, A., Pesek, I.: AI and cognitive modelling for education. Springer (2023). https://doi.org/10.1007/978-3-031-35331-4

Aberšek, B., Borstner, B., Bregant, J.: Virtual Teacher: Cognitive Approach to e-Learning Material. Cambridge Scholars Publishing (2014)

Ambrose, S.A., Bridges, M.W., DiPietro, M., Lovett, M.C., Norman, M.K.: How Learning Works: Seven Research-Based Principles for Smart Teaching. John Wiley & Sons, Inc. (2010)

Chomsky, N.: Topics in the Theory of Generative Grammar. Mouton (1966)

Downes, S.: An Introduction to Connective Knowledge. https://www.researchgate.net/publication/248290359 (2005)

Flogie, A., Aberšek, B., Pesek, I.: Education strategy for the net generation. Information. **2025**(16), 756 (2025). https://doi.org/10.3390/info16090756

Flogie, A., Aberšek, B.: The Impact of Innovative ICT Education and AI on the Pedagogical Paradigm. Cambridge Scholars Publishing (2019)

Kübler-Ross, E.: On Death and Dying: What the Dying Have to Teach Doctors, Nurses, Clergy and their Own Families. The Macmillan Company (1996)

McKraw, T.K.: Prophet of Innovation: Joseph Schumpeter and Creative Destruction. Harvard University Press (2007)

Landa, L.: Instructional Regulation and Control: Cybernetics, Algorithmization, and Heuristics in Education. Educational Technology Publications (1976)

Pratschke, B.M. (2024). Generative UI and Education. Digital Pedagogies, Teachinf Innovation and Learning Design. Springer

Sharples, M.: Practical Pedagogy: 40 New Ways to Teach and Learn. Taylor & Francis Group (2019)

Siemens, G.: Connectivism: a learning theory for the digital age. Int. J. Instr. Technol. Distance Learn. **2**(1) (2005a)

Siemens, G.: Connectivism: learning as network-creation. ASTD Learn. News. **10**(1) (2005b)

Suppes, P.: The uses of computers in education. Sci. Am. **215**(3), 206–223 (1966)

Suleyman, M., Bhaskar, M.: The Coming Wave. Vintage (2023)

WEF: Defining Education 4.0: A Taxonomy of Learning. World Economic Forum (2023)

Comparison of Human and GenAI Feedback Generation for Programming Assignments

Ľubomír Benko(✉), Janka Pecuchová, and Lucia Benková

Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, Nitra 949 01, SK, Slovakia
{lbenko,jpecuchova,lbenkova}@ukf.sk

Abstract. The paper focuses on comparing teacher feedback with Large Language Model (LLM) feedback for programming assignments in Java. Fifty students completed two object-oriented programming assignments and received three feedback texts per assignment. One feedback was provided by the teacher and two generated by API using GPT-4.1 and GPT-5 under restricted prompting and three feedback strategies (structured, metacognitive, reflexive). Students rated each feedback on a 5-point Likert scale evaluating actionability, specificity, and perceived correctness without source knowledge. Finally, they selected the most useful feedback from their point of view. Results showed significant source differences for actionability and specificity, with both LLM models rated higher than teacher feedback. Perceived correctness was high for all sources and statistically significant differences were not identified. Students favored GPT-4.1 as the most useful feedback, but the preference distribution did not significantly differ from equal distribution. However, the evaluated perceived correctness was not objectively verified correctness of feedback and poses a limitation of the research. Results suggest that LLMs can provide short-format feedback perceived as highly usable, though limitations of a small experimental sample urge caution in the interpretation of the results and aim for further research.

Keywords: Feedback · Generative Artificial Intelligence · Large Language Model · Programming Assignments

1 Introduction

The use of Generative Artificial Intelligence (GenAI) by students to find the right answers to questions, or to check grammar or summarize teaching texts is currently a common thing. On the other hand, teachers use GenAI to create interesting assignments for students. One of the main issues for teachers is timely and effective feedback on student solutions. In the case of high-enrollment or online courses used for distance education, such feedback is not available. This leads to the exploration of various options for automating feedback that would benefit students in these courses and at the same time relieve the workload of teachers, offer various functions and can be integrated into the educational process.

Traditional approaches are usually based on predefined test cases [1, 2], static analysis or rule-based methods to evaluate the code and offer some kind of feedback. However,

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 39–51, 2027.
https://doi.org/10.1007/978-3-032-32104-6_4

in most cases, the feedback is limited to errors or compilation of information. This does not allow the students to understand where in code they make mistakes and how to improve their understanding of programming language. On the other hand, students would obtain enough information to independently find the solution [2, 3]. GenAI models have already been integrated into various domains as automated feedback generators [4–9]. This paper is aimed at examining the integration of GenAI models into the process of feedback generation for students' programming assignment solutions.

The structure of this paper is as follows. The related work section focuses on the integration of GenAI models into feedback generation. The methods section describes the chosen approach, description of the dataset and formulating of research questions. The subsequent section focuses on research results. The discussion and conclusion sections summarize the findings.

2 Related Work

Each student is unique, as is their learning style. To achieve the best possible results, it is important that the student receives effective feedback that can guide them in the right direction. Manual feedback is time-consuming, so automatic feedback could be used. However, it only offers simple feedback that informs the student whether the solution is correct or not, specifically for coding tasks [10–12]. According to Hattie & Timperley [13], effective feedback should answer the questions "Where am I going" (feed up), "How am I going" (feedback), and "Where next" (feed forward).

After the arrival of Large Language Models (LLM), artificial intelligence techniques (such as machine learning and deep learning) were also integrated into the feedback generation process, with one field where LLM has found application as a feedback generator being coding tasks. The complexity of generating formative programming feedback lies mainly in the complexity of the syntax of the programming language, as well as in the number of different but correct solutions. GenAI and models like GPT-4 allow for the automation of personalized feedback and code analysis, which greatly streamlines programming education on a mass scale. The integration of these tools into educational processes is the subject of various research. According to Sahai et al. [14], LLMs are a more efficient approach for generating code fixes in Python than traditional automated program repair (APR). The integration of LLMs in the feedback process can take several forms, as LLM-assisted feedback systems or LLM-generated feedback [14–19].

Sahai et al. [14], distinguishes four approaches to generating feedback in programming tasks. The first approach is the use of test cases for error detection and diagnostic feedback. This approach is used, for example, by the Priscilla platform [2]. The disadvantage of this approach is that it cannot guide the student to the correct solution or specify the cause of the program failure. Correcting erroneous code itself is one way to learn programming, however, logical issues can be caused by the students' lack of understanding of the procedure. The second approach is the use of APR to generate corrections with manual feedback. This approach provides the student with hints in the form of partial corrections and the teacher with full corrections for assessment purposes. The third approach is the use of test cases for error detection and LLM to generate feedback

[14]. Finally, the fourth approach, which seems to be the most effective for both student and the teacher, is one where LLMs are used not only to generate corrections but also to generate feedback.

According to Pădurean et al. [16], current methods of generating feedback using LLM fail to meet the diverse needs of students, as they rely on prompt engineering techniques that can lead to rigid and limited responses. The authors argue that fine-tuning models on data collected from students produces feedback that better matches the style of feedback written by students. Moreover, such fine-tuning of models improves accuracy compared to feedback generated through prompt engineering alone.

One way to ensure that feedback reflects the real needs and experiences of users is to systematically collect feedback from users and implement it in Personalized and Adaptive Learning in computer programming. The question of whether and to what extent it is possible to control the feedback generated by LLM was addressed by Lohr et al. The authors propose a taxonomy that can capture the specifics of the generated feedback, such as randomness, uncertainty, and degree of variation. Na Nongkhai et al. [20], point to the suitability of using these systems. The adaptive system ensures the appropriate selection of additional tasks, given the needs of the student, while GenAI provides useful feedback. Kiesler et al. [19], argue that feedback generated by LLMs (specifically using ChatGPT 3.5) can help students solve simple programming tasks. However, they point out the need for feedback from the teacher, as the generated feedback can contain misleading information. The authors also point out the limits of such feedback, which are logical and semantic errors, as well as solutions with multiple errors. Research by Tang et al. [21], confirms the potential of LLMs (specifically ChatGPT) to generate practical and meaningful feedback for programming students. However, the relatively long time to generate feedback can be a problem, especially in classrooms with many students. In addition, the authors point out the importance of diversity in the tone of feedback, as some respondents preferred a positive tone, while others sought a more critical assessment. Respondents perceived the lack of clarity of feedback and the absence of code examples as shortcomings. From the perspective of a teacher, according to Tang et al. [21], an intuitive system interface is important, which allows for easier control and more effective use of AI-generated content in an educational context.

3 Methods

The students of applied informatics were offered the option to participate in the experiment to evaluate the feedback quality on their two assignments on Object Technologies subject. All of the student information was anonymized, and each participant has an option to recall the participation in the experiment at any time.

The students had two programming assignments focused on object-oriented programming principles. The participation was voluntary, however, the students had a restricted time window of 90 minutes to work on the assignment. The students did not use any GenAI models to help them create the codes. After submitting the assignment, teacher feedback was created for each submission and GenAI feedback was produced separately, using two models (GPT-4.1 and GPT-5) via API. The feedback was produced in Slovak language. Integrating different styles of feedback, three strategies for feedback generation were applied:

- structured,
- metacognitive,
- reflexive.

Students were offered three versions of feedback for their solution, one feedback was created by the teacher and two feedback were generated using the GenAI model. The GPT-4.1 and GPT-5 models were used to generate the feedback. Each student received one feedback from each of the examined models, with the feedback based on one of the strategies. The applied feedback strategies were selected at random but student was not assigned the same strategies in those two generated feedback. Students were informed in advance that the feedback was generated using GenAI but were disclosed which is which. The students' task was to rate each feedback using a 5-point Likert scale (1: completely disagree; 2: tend to disagree; 3: neither agree nor disagree; 4: tend to agree; 5: completely agree) to answer three questions:

1. The feedback helped me to understand what I can fix in the code right away.
2. The feedback focuses on specific errors in my code
3. I think the feedback is factually correct.

Finally, students had to choose the version of feedback that suited them best for each assignment.

After the second assignment, the students took a mid-term test, which topics were related to the two assignments and the scores were recorded. Students that submitted inadequate solutions or did not evaluate the feedback were excluded from the dataset.

This way a dataset containing 267 rows with 10 variables, representing 50 unique students, was created. For each student, three feedback items were recorded (human, GPT-4.1, GPT-5) and each feedback source contributed an equal number of feedback items ($n = 89$ per source).

OpenAI models were chosen as GenAI representatives, where the feedback was obtained from two different models using API. GPT-4.1 was selected as a non-reasoning model and GPT-5 was a reasoning model. The feedback was generated using both models during October and November 2025. Both models were prompted using the same prompt, that followed similar rules. The feedback had to be in Slovak language, without scores and grades, produced as one continuous paragraph without bullet points. The models were prohibited to include code blocks or inline code snippets and to reveal solutions. They were permitted to reference the class or method where issues were raised. The feedback length was restricted to 90-120 words. Input variables contained feedback strategy, assignment text, and the student code. The model settings are described in Table 1, where GPT-4.1 and GPT-5 use different output mechanism that are directly not comparable (temperature and reasoning effort).

The aim of the experiment was to evaluate the potential of GenAI models to create relevant and appropriate feedback to coding assignments. The use of GenAI models would reduce the time constraints placed on teachers by assessing the assignments and the need to create textual feedback. The following research questions were examined:

RQ1: *Do students rate feedback quality differently across sources (human, GPT-4.1, GPT-5) with respect to actionability, specificity, and perceived correctness?*

Table 1. GenAI Model Settings.

GPT-4.1		GPT-5	
Parameter	Value	Parameter	Value
Text format	Text	Text format	Text
Temperature	0.20	Reasoning effort	High
Max tokens	2048	Verbosity	Medium
Top P	1.00	Summary	None

RQ2: *Does the feedback source influence which feedback students select as the most useful?*
RQ3: *How do feedback texts differ across sources and feedback strategies (structured, metacognitive, reflexive) in linguistic and structural properties?*

4 Results

Students' evaluation of the feedback was examined. The three categories were scored by students using the 5-point Likert scale. All examined variables were not normally distributed ($p < 0.001$) which was tested using the Shapiro-Wilk test.

Overall, students evaluated GenAI feedback with higher scores than human feedback. In the case of the first category (actionability), statistically significant differences were identified using the Friedman test between the feedback sources ($p < 0.01$). GPT-4.1 obtained the highest average score from students (4.24), followed by GPT-5 (4.09) and human feedback obtained the lowest average score (3.71). The second category (specificity) showed rather similar scores for both GPT models. Similarly, the highest evaluated score was achieved by GPT-4.1 (4.27), followed by GPT-5 (4.22). Statistically significant differences were identified between the sources in this category ($p < 0.001$), while human feedback obtained the lowest average scores (3.47). The last category (correctness) showed similar results with all sources obtaining scores higher than 4 on average. Statistically significant differences were identified between the sources in the last category ($p < 0.05$) with similar performance as in previous categories (GPT-4.1 = 4.36; GPT-5 = 4.30; human = 4.08). Figure 1 depicts the distribution of scores for each feedback source per specific category.

Students rated GenAI feedback higher than human feedback most clearly on specificity, and to a lesser extent on actionability. Wilcoxon paired test with Bonferroni correction was used to compare the student evaluation. Table 2 shows that statistically significant differences for the examined categories were identified mostly between the human and GenAI feedback, while no significant differences were identified between the GenAI models ($p > 0.05$). All the feedback sources produced correct feedback from the students' point of view ($p > 0.05$).

Across the 89 student feedback items (Fig. 2), GPT-4.1 was selected as the most useful feedback source 36 times (40.45%), followed by GPT-5 with 32 selections (35.96%), and human feedback with 21 selections (23.60%). However, the overall preference distribution did not differ significantly from equal preference ($p = 0.131$). Even

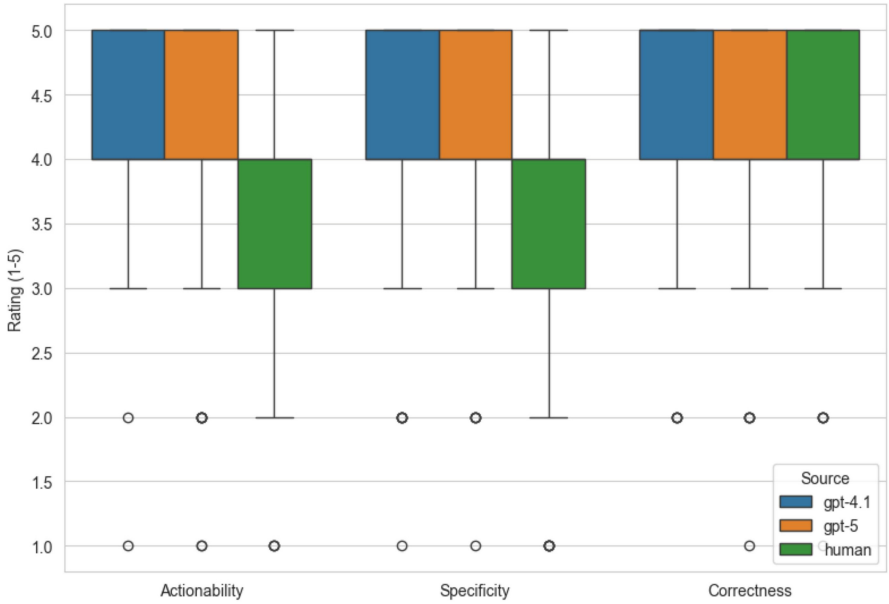


Fig. 1. Likert Scale Ratings by Feedback Source.

Table 2. Pairwise comparisons using Wilcoxon paired test with Bonferroni correction.

	GPT-4.1 × GPT-5	GPT-4.1 × human	GPT-5 × human
Actionability p value	0.5801	0.0016**	0.0600
Specificity p value	1.0000	0.0001***	0.0002***
Correctness p value	1.0000	0.0843	0.3124

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

though the source preference was not significant, the selected most useful feedback had obtained higher ratings than the non-chosen feedback on all three evaluated categories ($p < 0.001$ for each category). This indicates that students' selection was dependent on their perceived usefulness and correctness ratings. However, LLM is known to generate hallucinations, therefore perceived correctness cannot be interpreted as factual correctness.

As the student's preference leaned more towards GenAI generated feedback, a more detailed insight into which type of generated feedback strategy was better valued, was investigated. Within the preferred GenAI feedback ($n = 68$), there were no statistically significant differences identified between the GenAI source and feedback strategy ($p = 0.691$). In terms of test scores that followed up the two feedback assignments, feedback

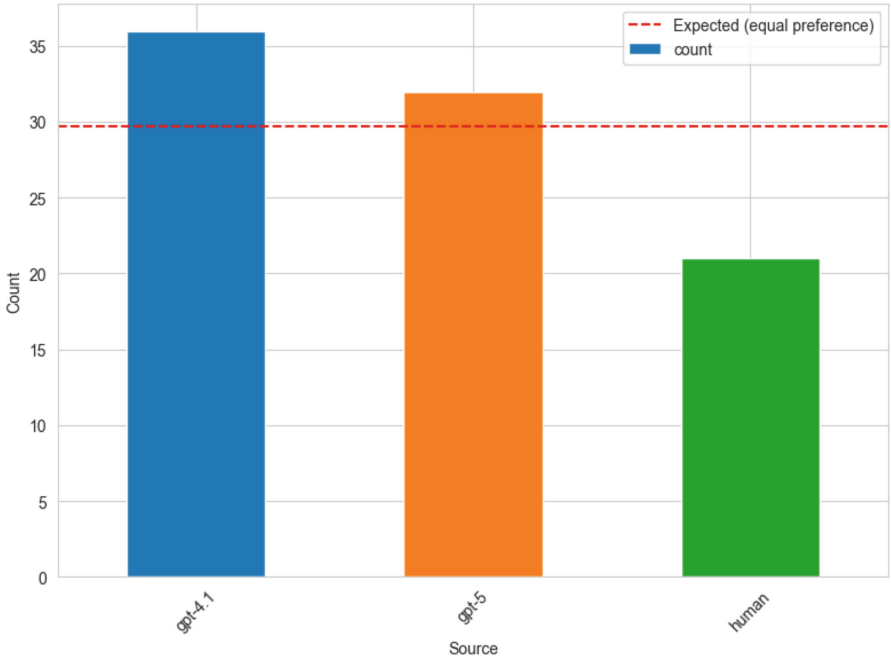


Fig. 2. Frequency of Selected Feedback Source.

strategies were not statistically significant (Kruskal-Wallis $H = 0.83$; $p = 0.661$). However, the means were slightly higher for metacognitive and reflexive feedback strategies, compared to structured feedback (Figs. 3 and 4). Overall, the test scores were not influenced by the feedback source ($p = 1.000$) and no relationship between the selected feedback and associated test score ($p = 1.000$) was identified. On the other hand, comparison of test scores of students that participated in the feedback experiment and those who did not showed benefits of the feedback. This comparison is associative rather than causal, as more engaged students were likely to have participated in the feedback experiment, which substantially confounds the interpretation of any performance difference. Students who received feedback obtained higher test scores (34 points of 50 on average) than those who did not participate (22 points of 50 on average) with statistically significant differences ($p < 0.01$). The results support an association between feedback participation and performance but do not establish causality. The analysis of feedback strategy differences should be treated as exploratory given the small per-cell sample sizes inherent in this design. The results of obtained no statistically significant differences between strategies should not be interpreted as evidence of equivalence as the study was likely insufficient to detect strategy-level effects.

Lastly, a linguistic analysis was done to compare the differences between human and GenAI feedback. Similarity was estimated using various sentence embedding models: SBERT (sentence-transformers/paraphrase-multilingual-MiniLM-L12-v) [22], BERT (bert-base-multilingual-cased), E5 (infloat/multilingual-cased), LaBSE

(sentence-transformers/LaBSE) [23], BGE (BAAI/bgr-reranker-v2-m3), GPT (text-embedding-3-large), and Cohere (embed-v4.0). The mean similarity between the GenAI models was lower than expected and the human feedback had higher average similarity scores with GPT-4.1, than GPT-5 (Table 3). All identified differences in similarity were statistically significant ($p < 0.001$). A morphological annotation done by Stanza [24], offered a more detailed insight into the feedback differences. Proportionally, the GenAI models produced similar distribution of word classes in the feedback. The feedback consisted mainly of nouns (GPT-4.1 = 36%; GPT-5 = 37%), adpositions (GPT-4.1 = 13%; GPT-5 = 13%), verbs (GPT-4.1 = 11%; GPT-5 = 12%) and adjectives (GPT-4.1 = 10%; GPT-5 = 11%). In comparison, the human feedback also contained mainly nouns (32%), verbs (15%), adpositions (10%) and adjectives (10%).

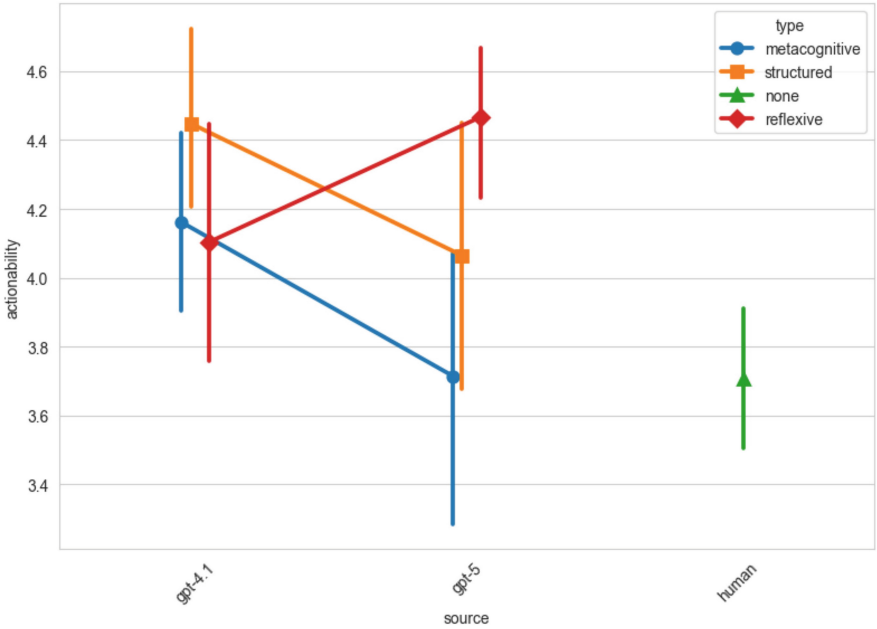


Fig. 3. Actionability ratings by feedback source and feedback strategy.

Table 3. Cosine similarities between feedback sources.

Model	Mean Human-GPT-4.1	Mean Human-GPT-5	Mean GPT-4.1-GPT-5	H-stat	p value
SBERT	0.6263	0.5159	0.6228	50.0348	0.000***
BERT	0.8969	0.8853	0.9216	71.9770	0.000***
E5	0.8864	0.8684	0.9096	172.0221	0.000***
LaBSE	0.6214	0.5419	0.7180	153.5695	0.000***

(*continued*)

Table 3. (*continued*)

Model	Mean Human-GPT-4.1	Mean Human-GPT-5	Mean GPT-4.1-GPT-5	H-stat	p value
BGE	0.6603	0.5961	0.7358	149.2524	0.000***
GPT	0.6823	0.6353	0.7499	112.3952	0.000***
Cohere	0.7330	0.6999	0.7754	61.5221	0.000***

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

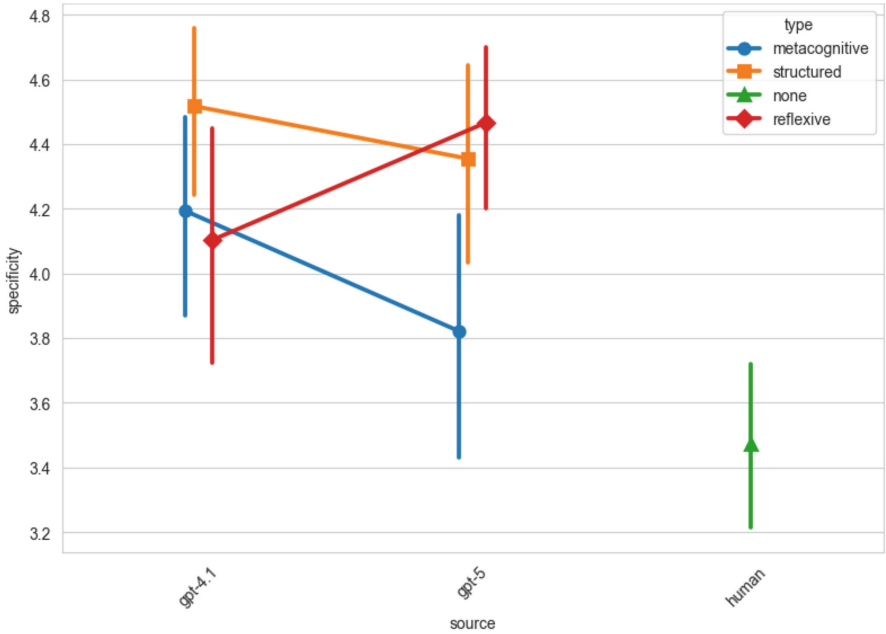


Fig. 4. Specificity ratings by feedback source and feedback strategy.

5 Discussion

In this study, students evaluated teacher feedback and two GenAI feedback sources (GPT-4.1 and GPT-5) generated under three strategies. Across two assignments, GenAI feedback received higher student ratings on actionability and specificity, while perceived correctness was high for all three sources. Preference for the best feedback tended to favor GenAI outputs, although the preference distribution did not significantly deviate from equal preference. Overall, these results suggest that, under the short-format constraint used in this study, GenAI feedback can achieve perceived usefulness comparable to teacher feedback.

RQ1: *Do students rate feedback quality differently across sources (human, GPT-4.1, GPT-5) with respect to actionability, specificity, and perceived correctness?*

Regarding the first research question, non-parametric comparisons indicated statistically significant differences between the sources for actionability and specificity, with higher average ratings for both GenAI sources and slightly higher means for GPT-4.1 than GPT-5. In the case of perceived correctness with averages higher than 4, no significant differences were identified between the sources. This could have been caused by the prompting restrictions of the GenAI models (short length, single paragraph, no code), that restricted the specificity of the produced feedback in producing directive formulations that students interpret as actionable. On the other hand, the teacher may write more naturally and assume shared context from the lectures, which can be read as less actionable even if it can be pedagogically appropriate. A key limitation is that this paper evaluated perceived correctness rather than objectively verified correctness of feedback. However, LLM feedback can be misleading, therefore perceived correctness cannot be interpreted as factual correctness.

RQ2: *Does the feedback source influence which feedback students select as the most useful?*

The second research question dealt with students selecting GenAI feedback as most useful more often than teacher feedback, with GPT-4.1 receiving the highest portion of selections. However, the differences in the selection preference were not statistically significant, which suggests that no single source dominated in perceived usefulness within this dataset. Chosen feedback had higher Likert ratings across categories, which means that students' choices aligned with their own rating criteria.

RQ3: *How do feedback texts differ across sources and feedback strategies (structured, metacognitive, reflexive) in linguistic and structural properties?*

Regarding the third research question, which is treated as exploratory given the constraints of the study design, linguistic analysis showed that feedback texts produced by GenAI were fairly similar to teacher feedback in overall word class distributions, with small differences such as a higher proportion of nouns and a lower proportion of verbs in GenAI outputs. The higher noun count in GenAI outputs may show a tendency toward information-dense phrasing, which could contribute to the higher specificity ratings GenAI feedback received. While the relative lack of verbs may partially explain why actionability ratings were somewhat more variable. However, no statistically significant differences were identified between feedback strategies in the obtained dataset. One probable reason is a limited sample, where each student evaluated only two assignment sets. Also, the short-format output restrictions may have reduced the practical differences between structured, metacognitive, and reflexive feedback strategies. Critically, the per-cell sample sizes resulting from 50 students across two assignments and three strategies were too small to create an interpretable null result. The absence of significant differences reflects insufficient statistical power rather than a confirmed lack of effect. The linguistic comparison did not show clear systematic differences between non-reasoning (GPT-4.1) and reasoning (GPT-5) model outputs, suggesting that in this restricted setting, model

type did not lead to detectable stylistic differences. However, sentence embedding similarity results indicated that human feedback was more similar to GPT-4.1 than to GPT-5. This may indicate to GPT-5's reasoning oriented generation process producing structurally more elaborate outputs. This could also lead to the marginally lower actionability and specificity ratings GPT-5 received in comparison to GPT-4.1. Overall, the linguistic findings are best taken as a descriptive secondary contribution to the paper. They provide a surface-level description of the feedback corpus and offer possibilities for future work. That could be focused on examining how morphological and structural properties of feedback relate to student perception. It should be noted that GPT-4.1 and GPT-5 were configured through fundamentally different settings (temperature of 0.20 for GPT-4.1 and high reasoning effort for GPT-5). This variation is inherent to the models' respective APIs and could not be equalized. It should be treated as a potential contributor to the observed differences between the two models and prevents direct provenance of those differences to model capability alone.

6 Conclusion

This paper dealt with examining the potential of GenAI to offer feedback to students for programming assignments. The students obtained feedback from three sources, which they evaluated in actionability, specificity and perceived correctness. Finally, they selected the most useful feedback from their point of view. One feedback was produced by the teacher and two generated using OpenAI GPT models (GPT-4.1 and GPT-5). Students' perceived feedback quality differed by source. GenAI feedback achieved higher ratings than teacher feedback on actionability and specificity, while perceived correctness was similarly high across sources. Students more frequently selected GenAI feedback as most useful, but this preference was not statistically significant. Feedback strategies did not yield reliable differences in perceived quality or test performance within this dataset. Linguistic analysis showed that GenAI models produced similar word-class distribution regardless of the used model. Limitation of the experiment was a small student sample participating in the feedback experiment and assessment of only two assignments. Another limitation was the restriction prompt for the GenAI models to produce short-format output without specific code issues and examples. Creating feedback is very time consuming for teachers and to offer an objective comparison, the generation of GenAI feedback had to approximately match the teachers' feedback. Future work should validate feedback correctness through expert annotation and connect feedback to measurable code revisions to evaluate learning impact more directly. Another research direction could be focused on not limiting the GenAI models and letting them produce feedback of any length. It would be interesting to compare the various prompting techniques and feedback strategies on a bigger sample. In conclusion, GenAI can be used as scalable supplemental feedback under teacher supervision, which could reduce the teacher's workload and offer important feedback to students in their education process.

Acknowledgements. This work was supported by EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project no. 09I03-03-V04-000596.

References

1. Skalka, J., Drlik, M., Obonya, J.: Automated assessment in learning and teaching programming languages using virtual learning environment. In: IEEE Global Engineering Education Conference, EDUCON, pp. 689–697. IEEE Computer Society (2019)
2. Skalka, J., Drlík, M.: Development of automatic source code evaluation tests using grey-box methods: a programming education case study. IEEE Access. **11**, 106772–106792 (2023). https://doi.org/10.1109/ACCESS.2023.3317694
3. Saikkonen, R., Malmi, L., Korhonen, A.: Fully automatic assessment of programming exercises. In: Proceedings of the 6th Annual Conference on Innovation and Technology in Computer Science Education, pp. 133–136. ACM, New York, NY, USA (2001)
4. Gao, R., Merzdorf, H.E., Anwar, S., et al.: Automatic assessment of text-based responses in post-secondary education: a systematic review. Comput. Educ.: Artif. Intell. **6**, 100206 (2024). https://doi.org/10.1016/j.caeai.2024.100206
5. Henkel, O., Hills, L., Roberts, B., McGrane, J.: Can LLMs grade open response reading comprehension questions? An empirical study using the ROARs dataset. Int. J. Artif. Intell. Educ. (2024). https://doi.org/10.1007/s40593-024-00431-z
6. Pecuchova, J., Drlik, M.: The role of generative artificial intelligence in the assessment of open-ended questions. In: International Conference on Interactive Collaborative Learning, pp. 394–405. Springer, Tallinn, Estonia (2025). https://doi.org/10.1007/978-3-031-85652-5_40
7. Pecuchova, J., Benko, Ľ., Drlik, M.: Automated grading of open-ended questions in higher education using GenAI models. Int. J. Artif. Intell. Educ. **35**, 3813–3846 (2025). https://doi.org/10.1007/s40593-025-00517-2
8. Pecuchová, J., Benko, Ľ.: Evaluating embedding-based similarity models for assessment of open-ended questions. In: Advanced Research in Technologies, Information, Innovation and Sustainability. ARTIIS 2025. Communications in Computer and Information Science, pp. 175–188. Springer, Cham (2026). https://doi.org/10.1007/978-3-032-16848-1_14
9. Pecuchova, J., Benko, L., Drlik, M.: Reimagining feedback through generative AI in engineering education. Comput. Educ.: Artif. Intell., 100574 (2026). https://doi.org/10.1016/j.caeai.2026.100574
10. Crow, T., Luxton-Reilly, A., Wuensche, B.: Intelligent tutoring systems for programming education. In: Proceedings of the 20th Australasian Computing Education Conference, pp. 53–62. ACM, New York, NY, USA (2018)
11. Jeuring, J., Keuning, H., Marwan, S., et al.: Towards giving timely formative feedback and hints to novice programmers. In: Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education, pp. 95–115. ACM, New York, NY, USA (2022)
12. Keuning, H., Jeuring, J., Heeren, B.: A systematic literature review of automated feedback generation for programming exercises. ACM Trans. Comput. Educ. **19**, 1–43 (2019). https://doi.org/10.1145/3231711
13. Hattie, J., Timperley, H.: The power of feedback. Rev. Educ. Res. **77**, 81–112 (2007). https://doi.org/10.3102/003465430298487
14. Sahai, S., Ahmed, U.Z., Leong, B.: Improving the coverage of GPT for automated feedback on high school programming assignments. In: NeurIPS'23 Workshop Generative AI for Education (GAIED), pp. 1–11. MIT Press, New Orleans, USA (2023)
15. Yousef, M., Mohamed, K., Medhat, W., et al.: BeGrading: large language models for enhanced feedback in programming education. Neural Comput. Appl. **37**, 1027–1040 (2025). https://doi.org/10.1007/s00521-024-10449-y

16. Pădurean, V.-A., Phung, T., Kotalwar, N., et al.: Humanizing automated programming feedback: fine-tuning generative models with student-written feedback. In: Proceedings of the 18th International Conference on Educational Data Mining, pp. 434–441. International Educational Data Mining Society, Palermo, Italy (2025)
17. Zhang, Z., Dong, Z., Shi, Y., et al.: Students' perceptions and preferences of generative artificial intelligence feedback for programming. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 23250–23258 (2024). https://doi.org/10.1609/aaai.v38i21.30372
18. Pankiewicz, M., Baker, R.S.: Large Language Models (GPT) for automating feedback on programming assignments. In: Proceedings of the 31st International Conference on Computers in Education, pp. 1–10. Asia-Pacific Society for Computers in Education, Matsue, Japan (2023)
19. Kiesler, N., Lohr, D., Keuning, H.: Exploring the potential of large language models to generate formative programming feedback. In: 2023 IEEE Frontiers in Education Conference (FIE), pp. 1–5. IEEE (2023)
20. Na Nongkhai, L., Wang, J., Mendori, T.: An adaptive learning support system based on ontology of multiple programming languages. In: International Conference on Computers in Education (2023). https://doi.org/10.58459/icce.2023.958
21. Tang, X., Wong, S., Huynh, M., et al.: SPHERE: supporting personalized feedback at scale in programming classrooms with structured review of generative AI outputs. In: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, pp. 1–17. ACM, New York, NY, USA (2025)
22. Condor, A., Litster, M., Pardos, Z.: Automatic short answer grading with SBERT on out-of-sample questions. In: Proceedings of The 14th International Conference on Educational Data Mining (EDM 2021), pp. 345–352. International Educational Data Mining Society, Paris, France (2021)
23. Feng, F., Yang, Y., Cer, D., et al.: Language-agnostic BERT sentence embedding. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 878–891. Association for Computational Linguistics, Stroudsburg, PA, USA (2022)
24. Qi, P., Zhang, Y., Zhang, Y., et al.: Stanza: a python natural language processing toolkit for many human languages. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 101–108. Association for Computational Linguistics, Stroudsburg, PA, USA (2020)

Integrating AI and Formative E-Assessment in Geometry Education

Katerina Biernatova[1](✉) and Lilla Korenova[2]

[1] Faculty of Education, University of Ostrava, Ostrava, Czech Republic
bierka30@osu.cz
[2] Faculty of Education, Comenius University in Bratislava, Bratislava, Slovakia
korenova@fedu.uniba.sk

Abstract. The rapid integration of artificial intelligence (AI) into educational contexts has created new opportunities for enhancing formative assessment, particularly in mathematics education. Despite increasing attention to AI-supported learning, its didactic potential in fostering students' conceptual understanding of geometry remains insufficiently explored. This study addresses this gap by examining how AI-driven formative assessment can support learning processes in geometry at the primary and lower secondary levels. Grounded in constructivist and socio-constructivist learning theories, as well as principles of formative assessment and self-regulated learning, the study adopts a design-based research (DBR) approach. Iterative cycles of design, implementation, and reflection were conducted in authentic classroom settings to develop and refine AI-supported assessment strategies. Data were collected through classroom observations, student artefacts, and semi-structured interviews, and analysed using a mixed-methods framework. The findings indicate that AI-supported formative assessment can enhance students' engagement, provide immediate and adaptive feedback, and support the development of deeper conceptual understanding in geometry. At the same time, the results highlight important challenges related to teachers' readiness, the risk of over-reliance on automated feedback, and the need for careful pedagogical integration. The study contributes to the field by proposing a pedagogical framework that integrates AI-driven feedback mechanisms with formative assessment practices in geometry education. The findings offer implications for the design of technology-enhanced learning environments and for the professional development of teachers in the context of digital transformation in education.

Keywords: artificial intelligence in education · formative assessment · e-assessment · geometry education · mathematics education · technology-enhanced learning · scaffolding · self-regulated learning · teacher orchestration · digital competence

1 Introduction

Geometry represents a central domain of mathematics education, fostering spatial reasoning, logical thinking, and the understanding of relationships between objects in two- and three-dimensional space. These competencies are essential not only for further mathematical study but also for everyday problem solving and various professional contexts.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 52–62, 2027.
https://doi.org/10.1007/978-3-032-32104-6_5

Despite its importance, geometry is frequently perceived by students as abstract and difficult. Learners often encounter challenges in visualising spatial relationships, interpreting formal definitions, and applying geometric concepts in meaningful situations (Hoyles & Healy, 1997; Leikin et al., 2000). Traditional instructional approaches may contribute to these difficulties. Geometry teaching often relies on static representations and procedural exercises, which can limit opportunities for exploration and active knowledge construction. As a result, students may develop fragmented understanding and reduced engagement. In response to these challenges, digital technologies have increasingly been integrated into mathematics education. Dynamic visualisations and interactive tools enable students to manipulate geometric objects, explore relationships, and connect abstract concepts with visual representations. Recent developments in artificial intelligence (AI) have further expanded the range of technological support available in educational settings. AI-supported tools offer possibilities for adaptive explanations, personalised feedback, and differentiated task design (Mustafa, 2024; Mariyono, 2025). At the same time, formative e-assessment systems provide continuous feedback and data that can inform both student reflection and instructional decisions (Bennett, 2015; JISC, 2010). However, technology alone does not ensure meaningful learning. Constructivist perspectives emphasise that understanding develops through active exploration, conjecturing, and reflection guided by purposeful pedagogical design (Carr & Kemmis, 2003). The aim of this paper is to propose and theoretically ground a pedagogical framework that integrates AI-supported tools and formative e-assessment within constructivist geometry instruction at the lower secondary level. The paper outlines the theoretical foundations of this approach and illustrates how these elements can be coherently embedded in regular classroom practice. The contribution of this paper lies in articulating a pedagogically coherent interaction model that integrates AI-supported tools and formative e-assessment within a constructivist geometry framework. By identifying design principles and structural relationships between these components, the paper shifts the focus from technological adoption to pedagogically intentional AI integration. This paper is conceptual in nature and seeks to integrate existing research into a coherent pedagogical model for AI-supported geometry instruction. Rather than reporting empirical findings, it articulates structural relationships, design principles, and pedagogical tensions that define meaningful technology integration within constructivist mathematics education.

2 Theoretical Framework

The proposed framework is grounded in three interconnected areas: artificial intelligence in education, formative e-assessment, and constructivist learning theory. Rather than treating these components separately, the framework emphasises their pedagogical complementarity in geometry instruction. Artificial intelligence has recently emerged as a significant area of interest in mathematics education. Recent research highlights both the instructional potential and pedagogical risks of generative AI systems in education, particularly in relation to scaffolding, automated feedback, and learner autonomy (Chiu, 2023). AI-supported tools can provide adaptive explanations, generate differentiated tasks, and offer immediate feedback tailored to students' needs (Mustafa, 2024; Mariyono, 2025). AI-supported tools can provide adaptive explanations, generate differentiated

tasks, and offer immediate feedback tailored to students' needs. Research on intelligent tutoring systems demonstrates that AI-supported guidance can approach the effectiveness of human tutoring when carefully aligned with pedagogical goals (VanLehn, 2011). However, effective implementation depends not merely on technological capacity but on pedagogical orchestration and instructional alignment. In this paper, AI-supported tools refer primarily to conversational AI systems, adaptive tutoring environments, and algorithmically generated scaffolding mechanisms capable of providing context-sensitive explanations and differentiated task suggestions. These tools are conceptualised as pedagogical mediators rather than autonomous instructional agents. In geometry education, such tools are particularly relevant due to the subject's reliance on visualisation and spatial reasoning. Dynamic representations and AI-assisted explanations can help students explore geometric relationships, clarify misconceptions, and approach problems from multiple perspectives. However, current research also stresses that AI should function as a supportive resource under teacher guidance rather than as an autonomous instructional agent (Klimova & Pikhart, 2025). Formative e-assessment represents a second theoretical pillar of the framework. Unlike traditional summative assessment, formative digital assessment supports learning during the instructional process by providing timely feedback and enabling continuous monitoring of understanding (Bennett, 2015). In geometry, digital assessment environments allow the inclusion of visual and interactive elements that more accurately reflect geometric reasoning than purely symbolic tasks. When thoughtfully integrated, e-assessment can promote reflection, self-regulated learning, and informed instructional decision-making (JISC, 2010). The pedagogical foundation of the framework is rooted in constructivist learning theory, which views learning as an active process of knowledge construction through exploration, problem solving, and reflection (Carr & Kemmis, 2003). In geometry education, constructivist approaches encourage students to investigate properties, formulate conjectures, and justify reasoning using multiple representations. Digital tools can support such processes by offering dynamic environments for experimentation, but meaningful learning depends on coherent task design and teacher mediation (Conole & Oliver, 2006). Taken together, these perspectives suggest that the educational value of technology in geometry lies not in its novelty, but in its integration within a structured pedagogical framework. AI-supported tools enhance flexibility and differentiation, formative e-assessment provides feedback and insight into learning processes, and constructivist principles give direction and meaning to instructional design. Their deliberate combination forms the conceptual basis of the proposed approach. Recent developments in generative AI, particularly large language models, have further expanded the possibilities of adaptive explanation and contextualised scaffolding in mathematics education (Kasneci et al., 2023).

3 Pedagogical Design of AI-Supported Geometry Instruction

The proposed pedagogical design builds on the theoretical foundations outlined above and translates them into a coherent instructional structure for geometry teaching at the lower secondary level. Its central premise is that meaningful integration of technology requires deliberate alignment between learning objectives, instructional activities, assessment strategies, and digital tools. Technology is not introduced as an additional

layer, but as an element embedded within the overall didactic structure. The design framework can be understood as a dynamic interaction between three core components: AI-supported instructional tools, formative e-assessment mechanisms, and constructivist task design. Each component serves a distinct pedagogical function while remaining interdependent with the others. AI-supported tools are primarily employed to enhance visualisation, provide adaptive explanations, and support differentiation. In geometry education, where students frequently encounter difficulties related to spatial imagination and interpretation of symbolic representations, such tools can facilitate access to multiple representations of the same concept. For instance, AI systems may generate alternative explanations, propose scaffolded hints, or suggest varied examples that address specific misconceptions. When carefully integrated, these features support cognitive flexibility and encourage students to approach geometric problems from different perspectives. Importantly, AI-supported tools are used to scaffold reasoning rather than to replace it. The pedagogical intention is to promote metacognitive engagement by prompting students to reflect on their solution strategies, justify their reasoning, and reconsider initial assumptions. In this sense, AI functions as a mediating instrument within the learning process, not as an autonomous source of authority. Formative e-assessment represents the second structural element of the design. Assessment is conceptualised as an ongoing feedback process embedded within instruction rather than as a separate evaluative event. Short digital tasks, interactive questions, or dynamic geometry exercises can be integrated at multiple stages of a lesson to monitor conceptual development. Immediate feedback supports error recognition and conceptual refinement, while aggregated results provide teachers with insight into learning patterns and potential areas of difficulty.

This continuous feedback cycle enables adaptive instructional responses. Teachers may modify task difficulty, revisit key concepts, or provide targeted clarification based on emerging patterns. In this way, formative e-assessment strengthens instructional responsiveness and supports differentiated learning pathways without fundamentally altering the curriculum structure. Constructivist task design forms the pedagogical core of the framework. Learning activities are structured to promote exploration, conjecturing, and justification rather than routine procedural repetition. Students are encouraged to manipulate geometric objects, compare alternative strategies, and articulate mathematical arguments. Digital tools provide dynamic environments for experimentation, allowing learners to observe the consequences of transformations or structural changes in real time. The coherence of the design lies in the deliberate orchestration of these elements. AI-supported tools enhance flexibility and individual support, formative assessment ensures continuous feedback and instructional awareness, and constructivist task design fosters meaningful conceptual engagement. The teacher remains central in coordinating these components, setting boundaries for tool use, and ensuring that technological support aligns with clearly defined learning goals. Through this structured yet adaptable design, technology becomes integrated within a pedagogically intentional environment that prioritises conceptual understanding, active participation, and reflective thinking in geometry education.

3.1 Conceptual Structure of the Framework

The pedagogical design can be conceptualised as a triadic interaction model composed of three mutually reinforcing components: AI-supported instructional tools, formative e-assessment, and constructivist task design. Rather than functioning as independent elements, these components operate within a dynamic and iterative system of instructional interaction. AI-supported tools primarily address cognitive and representational dimensions of learning. They provide scaffolding, alternative explanations, and access to multiple representations that help students navigate complex geometric concepts. Their purpose is not to simplify learning to a procedural level, but to lower unnecessary cognitive barriers while maintaining meaningful intellectual engagement.

Formative e-assessment serves as a regulatory mechanism within the model. It establishes continuous feedback loops between students and teachers, supporting both self-reflection and instructional adjustment. Feedback enables students to refine strategies and reconsider misconceptions, while aggregated assessment data informs teachers' decisions regarding pacing, differentiation, and task sequencing. Constructivist task design forms the epistemic core of the framework. It defines the structure and purpose of learning activities, ensuring that exploration, conjecturing, and justification remain central to geometry instruction. Digital tools and assessment strategies are therefore aligned with these epistemic goals rather than operating as isolated technological add-ons. The interaction between these components is cyclical rather than linear. Constructivist tasks initiate exploration; AI tools provide scaffolding during problem solving; formative assessment generates feedback; and instructional adaptations shape subsequent task design. This iterative cycle supports sustained conceptual development and pedagogical responsiveness. Conceptualising the framework as an interaction model emphasises coherence and intentionality in technology integration. The teacher occupies a central coordinating position within this system, orchestrating the interplay of technological and pedagogical elements to ensure alignment with clearly defined learning objectives. The interaction between the three components of the framework is illustrated in Fig. 1.

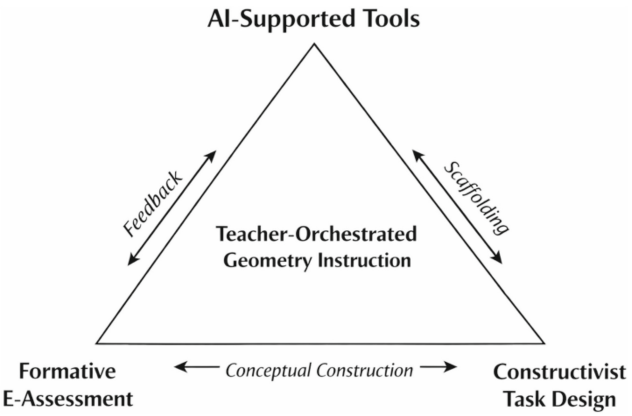


Fig. 1 Conceptual interaction model of AI-supported geometry instruction

3.2 Design Principles for AI-Supported Geometry Instruction

To operationalise the proposed framework and strengthen its pedagogical coherence, a set of design principles can be articulated. These principles do not prescribe specific technologies, but rather define conditions under which AI-supported tools and formative e-assessment can meaningfully contribute to geometry learning.

Principle 1: AI as Scaffold, Not Substitute
AI-supported tools should function as cognitive scaffolds that support reasoning without replacing it. Their purpose is to guide exploration, provide alternative representations, and offer graduated hints rather than complete solutions. By maintaining an appropriate level of productive struggle, AI can enhance conceptual understanding while preserving students' epistemic agency.

Principle 2: Feedback Targeting Reasoning, Not Only Correctness
Formative e-assessment should prioritise feedback that addresses conceptual structure and reasoning processes rather than merely indicating right or wrong answers. Geometry learning depends on justification and relational understanding; therefore, digital feedback mechanisms should prompt reflection, encourage explanation, and support error analysis.

Principle 3: Alignment with Epistemic Goals of Geometry
Technology integration must be aligned with the epistemic core of geometry: exploration, conjecturing, and justification. Digital tools and AI explanations should support multi-representational thinking and structural reasoning rather than reduce tasks to procedural execution. The primary learning goal remains conceptual insight into geometric relationships.

Principle 4: Teacher-Orchestrated Mediation
The teacher retains a central coordinating role in regulating the interaction between AI tools, assessment processes, and task design. Effective integration requires deliberate orchestration, including setting boundaries for tool use, framing inquiry tasks, and facilitating reflective discussion. AI operates within teacher-defined pedagogical intentions.

Principle 5: Iterative Feedback Cycles
The interaction between AI scaffolding, formative assessment, and task design should function as an iterative cycle. Exploration generates questions; AI provides targeted support; formative assessment produces feedback; and instructional decisions adjust subsequent learning tasks. This cyclical structure promotes sustained conceptual development rather than isolated task completion.

4 Illustrative Classroom Implementation

The proposed pedagogical design can be implemented within regular geometry lessons without requiring fundamental curricular changes. This section outlines general principles of how AI-supported tools and formative e-assessment may be integrated into

classroom practice at the lower secondary level. AI-supported tools can be incorporated during individual or small-group work to assist students in exploring geometric concepts. Such tools provide alternative explanations, guiding questions, and scaffolded hints that support students in developing their reasoning. Rather than replacing problem solving, AI functions as a cognitive scaffold that encourages students to articulate their thinking and reflect on solution strategies. In geometry education, digital environments play a crucial role in supporting visualisation and multiple representations. Students can construct and manipulate geometric objects, test transformations, and examine relationships between elements in dynamic settings. This active engagement helps bridge the gap between abstract definitions and concrete understanding while supporting the development of spatial reasoning. Formative e-assessment can be embedded throughout the lesson through short digital activities, such as interactive quizzes or quick-response tasks. These activities provide immediate feedback and allow both students and teachers to monitor understanding during the learning process. Feedback can be used to identify misconceptions and inform instructional adjustments. Constructivist task design remains central to the implementation. Learning activities are structured to promote exploration, conjecturing, and justification. Students are encouraged to engage in problem-oriented tasks and participate in collaborative discussion and reflection. The teacher maintains a guiding role by orchestrating the interaction between AI tools, assessment processes, and learning activities. Together, these elements create a structured yet flexible learning environment in which technology enhances exploration, feedback, and differentiation while maintaining pedagogical coherence.

4.1 Illustrative Application Scenario

To provide a more concrete illustration of the proposed framework, an application scenario is presented to demonstrate how the interaction model could function in a lower secondary geometry lesson. The aim is not to provide empirical validation, but to offer a practical example of pedagogical application. The scenario is situated within a geometry lesson focused on exploring relationships between geometric objects through inquiry-oriented tasks. Students are expected to formulate conjectures, test their ideas using dynamic representations, and justify their reasoning. During individual and small-group work, AI-supported tools (e.g., conversational AI systems such as ChatGPT or similar tools) can be used to provide alternative explanations, guiding questions, and scaffolded hints. These tools support students in reflecting on their reasoning and comparing different solution strategies rather than supplying complete answers.

Formative e-assessment is integrated through short digital activities embedded at key moments of the lesson. Interactive quizzes or quick-response tasks provide immediate feedback and allow the teacher to monitor students' understanding. Based on this feedback, the teacher can adjust instruction, revisit key concepts, or provide targeted support. From a pedagogical perspective, this type of implementation may enhance student engagement, support conceptual understanding through multiple representations, and encourage iterative refinement of solutions. At the same time, it highlights potential challenges, such as the risk of overreliance on AI-generated hints, which underscores the importance of teacher mediation. Although this scenario is illustrative, it demonstrates the practical applicability of the proposed framework and clarifies the interaction

between AI support, formative assessment, and constructivist task design in geometry instruction.

5 Discussion

The proposed framework suggests that AI-supported formative e-assessment can enhance students' conceptual understanding of geometry, particularly by supporting iterative learning processes and providing immediate feedback. From a constructivist perspective, learning is understood as an active process of knowledge construction, where students refine their understanding through exploration, reflection, and feedback. In this context, AI-supported feedback can function as a form of scaffolding that supports students in developing more sophisticated problem-solving strategies. The integration of formative e-assessment further contributes to instructional coherence by embedding feedback directly into the learning process. Immediate and adaptive feedback may support students in identifying misconceptions, revising their reasoning, and engaging in iterative refinement of their solutions. This aligns with previous research on formative assessment, which emphasises the importance of timely feedback in fostering deeper learning and conceptual understanding in mathematics. At the same time, the effectiveness of AI-supported formative assessment depends on its pedagogical integration. Feedback mechanisms must be aligned with tasks that promote reasoning, justification, and conceptual understanding rather than procedural correctness. Without such alignment, there is a risk that digital assessment replicates traditional testing practices without contributing to meaningful learning. The domain of geometry plays a specific role in this context. Its strong connection to visualisation and spatial reasoning makes it particularly suitable for AI-supported and dynamic learning environments. However, effective implementation requires carefully designed tasks that encourage exploration, conjecturing, and the use of multiple representations. Despite its potential, the integration of AI-supported tools into geometry instruction also presents several challenges. Effective use of such technologies requires not only access to appropriate digital infrastructure, but also sufficient levels of teacher digital competence and pedagogical readiness. The teacher remains central in orchestrating the interaction between AI tools, assessment processes, and learning activities.

5.1 Pedagogical Tensions in AI-Supported Geometry Instruction

While the proposed framework emphasizes pedagogical coherence, its implementation inevitably involves several tensions that require careful consideration. These tensions do not represent shortcomings of technology itself, but rather structural dilemmas inherent in the integration of AI-supported tools into constructivist learning environments.

AI Scaffolding vs. Productive Struggle
One central tension concerns the balance between scaffolding and productive struggle. Geometry learning benefits from cognitive challenge, exploration, and the gradual construction of insight. If AI-generated hints or explanations are provided too readily, they may reduce opportunities for meaningful intellectual effort. Conversely, insufficient guidance may leave students disengaged or confused. Effective implementation

therefore requires carefully calibrated AI support that sustains cognitive demand while preventing frustration.

Automated Feedback vs. Epistemic Depth
Formative e-assessment systems typically priorities efficiency and immediacy. However, geometric reasoning involves justification, relational thinking, and structural insight that may not be fully captured through automated feedback. There is a risk that digital assessment reduces complex reasoning to correctness indicators or procedural checks. To maintain epistemic depth, feedback mechanisms must be aligned with reasoning processes rather than solely with final answers.

Adaptivity vs. Learner Autonomy
AI-supported differentiation can personalize learning trajectories by adjusting task difficulty and providing tailored explanations. Yet excessive adaptivity may unintentionally limit students' agency by narrowing the range of exploratory possibilities. In geometry, where multiple solution paths and representational strategies are valuable, maintaining learner autonomy remains essential. Adaptivity should therefore support, rather than constrain, exploratory engagement. Previous research on AI-supported self-regulated learning suggests that while adaptive systems can enhance monitoring and feedback processes, they may also influence learners' autonomy and decision-making patterns (Chiu, 2022).

Technological Mediation vs. Teacher Orchestration
Although AI tools offer adaptive functionalities, the pedagogical responsibility remains with the teacher. A further tension emerges between automated system responses and teacher-led orchestration. Overreliance on AI-mediated explanations may weaken the teacher's role in facilitating collective reflection and mathematical discourse. Sustained instructional coherence depends on the teacher's capacity to integrate AI feedback into broader pedagogical intentions.

5.2 Risks of AI-Supported Geometry Instruction

While AI-supported tools offer significant pedagogical potential, their integration into geometry instruction also raises several important concerns. One key risk is the overreliance on AI-generated explanations, which may reduce students' engagement in independent problem-solving and limit opportunities for productive struggle. Another challenge relates to the accuracy and quality of AI-generated responses. Inaccurate or oversimplified explanations may lead to misconceptions if they are not critically evaluated within the classroom context. This highlights the importance of positioning AI as a supportive rather than authoritative source of knowledge. Furthermore, excessive dependence on automated feedback may shift students' attention from reasoning processes to surface-level correctness. To preserve the epistemic depth of geometry learning, AI-supported feedback must be aligned with tasks that require justification, explanation, and conceptual understanding.

These considerations suggest that the effective integration of AI into geometry education depends not only on technological availability, but also on careful pedagogical design and critical teacher mediation.

6 Conclusion

This paper proposed a pedagogically grounded framework for integrating AI-supported tools and formative e-assessment into geometry instruction at the lower secondary level. The framework builds on constructivist principles and emphasises the complementary interaction between adaptive technological support, continuous formative feedback, and active knowledge construction. A key contribution of the paper lies in its integrative perspective, bringing together AI technologies, pedagogical principles, and the subject-specific characteristics of geometry. Rather than positioning technology as an autonomous instructional agent, the approach highlights the importance of coherent pedagogical design and teacher mediation. AI-supported tools may enhance visualisation, differentiation, and reflective problem solving, while formative e-assessment supports timely feedback and informed instructional decision-making. When embedded within inquiry-oriented tasks, these elements can contribute to more engaging and conceptually focused geometry learning environments. The framework offers a structured yet flexible model that can be adapted to diverse classroom contexts without fundamental curricular changes. At the same time, effective AI integration depends not only on technological availability but also on teachers' digital competence and pedagogical readiness (Chiu & Chai, 2020). Future research should further explore implementation in real classroom settings, long-term effects on student learning, and the evolving role of AI in mathematics education. Recognising the inherent pedagogical tensions in AI-supported instruction underscores that technological innovation must be accompanied by deliberate didactic structuring, as emphasised in research on technology-enhanced learning (Mishra & Koehler, 2006).

Acknowledgement. SGS Integrating E-Assessment into Constructivist Approaches to Geometry in Pre- Service Mathematics Teacher Education.

Erasmus+ 2023-1-CZ01-KA220-HED-000160664: "Accelerating STEAM-related Knowledge and Skills via 3D Modelling and 3D Printing".

References

Aljarrah, A., Ababneh, M., Karagozlu, D., Ozdamli, F.: Artificial intelligence techniques for distance education: a systematic literature review. TEM J. **10**(4), 1621–1629 (2021). https://doi.org/10.18421/TEM104-18

Bautista, G., López-Costa, M.: Smart learning spaces considering the integration of the pedagogical, environmental and digital dimensions: a systematic review. Learn. Environ. Res. **28**, 455–472 (2025). https://doi.org/10.1007/s10984-025-09551-2

Chiu, T.K.F.: Applying artificial intelligence in self-regulated learning: a systematic review. Educ. Technol. Soc. **25**(3), 1–14 (2022)

Chiu, T.K.F.: The impact of generative AI (e.g., ChatGPT) on teaching and learning: opportunities and challenges. Comput. Educ.: Artif. Intell. **4**, 100134 (2023)

Chiu, T.K.F., Chai, C.S.: Sustainable development of digital competence in teachers: technology acceptance and pedagogical readiness. Comput. Educ. **150**, 103840 (2020)

Holmes, W., Bialik, M., Fadel, C.: Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign, Boston, MA (2019)

Kasneci, E., Sessler, K., Küchemann, S., et al.: ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Individ. Differ. **103**, 102274 (2023)

Klimova, B., Pikhart, M.: Artificial intelligence and mathematics education: a review. Front. Psychol. (2025)

Lavicza, Z.: Integrating technology into mathematics teaching at the university level. ZDM Math. Educ. **42**, 105–119 (2010)

Lavicza, Z., Weinhandl, R., Prodromou, T., et al.: Developing and evaluating educational innovations for STEAM education in rapidly changing digital technology environments. Sustainability. **14**(12), 7237 (2022)

Mariyono, M.: Artificial intelligence in mathematics education: Challenges and opportunities. Educ. Sci. (2025)

Mishra, P., Koehler, M.J.: Technological pedagogical content knowledge: a framework for teacher knowledge. Teach. Coll. Rec. **108**(6), 1017–1054 (2006)

Mustafa, M.: AI-based learning environments in mathematics education. J. Educ. Technol. (2024)

Schmid, A., Nam, P.S., Lavicza, Z., Houghton, T.: Augmented reality as a tool to support the connection between reality and mathematical knowledge: a case for teaching parallel relations of planes in space. Int. J. Technol. Math. Educ. **31**(3), 135–142 (2024)

Šarmanová, J., Kostolányová, K.: Adaptive e-learning: from theory to practice. Int. J. Inf. Commun. Technol. Educ. **4**(4), 34–47 (2015)

VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ. Psychol. **46**(4), 197–221 (2011)

Žilková, K., Záhorec, J., Munk, M.: Analysis of the level of geometric thinking of pupils in Slovakia. Educ. Sci. **15**(8), 1020 (2025)

Žilková, K.: Method of virtual spatial manipulation in teaching of mathematics. J. Technol. Inf. Educ. **1**(2), 33–39 (2009)

Interaction Between Real and Virtual Environments in Chemistry Education Using the Example of Redox Reactions

Hana Hlavsova[1], Veronika Machkova[2], and Martin Bilek[1](✉)

[1] Faculty of Education, Charles University, M. Rettigové 4, Prague, Czechia
martin.bilek@pedf.cuni.cz

[2] Faculty of Science, University of Hradec Kralove, Rokitanskeho 62, Hradec Kralove, Czechia

Abstract. Electrochemistry dealing with redox reactions is considered one of the most challenging topics for students due to its abstract nature, interdisciplinary relations and complexity. Digital modelling and computer simulations provide students with opportunities to observe chemical phenomena as well as to explore processes at the microscopic level supporting deeper conceptual understanding. They can facilitate learning process across all levels of knowledge addressed in chemistry—at symbolic, sub-microscopic and macroscopic levels. This study investigates the effectiveness of selected computer simulations of redox processes—specifically electrolysis and galvanic cells—in secondary school chemistry education. A pedagogical experiment was conducted with secondary school students to compare the impact of the implemented simulation-based instructions and identical real-world experimental activities. Students' knowledge was assessed before and after the activities, and their knowledge retention was evaluated by using didactic tests. Post-tests and retention tests also measured students' confidence in their answers. The findings indicate that simulations are particularly effective in improving students' understanding of redox processes at the microscopic level, whereas real-world experiments contribute more significantly to understanding at the macroscopic level. A supplementary survey examining students' attitudes toward the teaching methods revealed that both approaches were perceived as interesting, engaging, and useful. However, simulations generated slightly higher levels of interest. Overall, the results suggest that computer simulations are a valuable tool in teaching electrochemistry, especially when combined with real laboratory experimentation.

Keywords: Teaching electrochemistry at a secondary school · computer simulations · electrolysis · and galvanic cell

1 Introduction

The dynamic development of digital technologies has significantly contributed to the growing integration of simulations and visualizations into science education. These digital tools markedly enhance conceptual clarity, support interactive forms of learning,

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 63–74, 2027.
https://doi.org/10.1007/978-3-032-32104-6_6

and increase the overall attractiveness of scientific curricula. In the context of the global rise of STEM-oriented education, simulations represent not only a motivational component but also an effective pedagogical medium for presenting, structuring, and assessing subject matter [1, 2]. A further advantage lies in the ability of computer simulations to depict phenomena at the micro level, thereby granting students access to processes that would otherwise remain inaccessible through traditional instructional methods. At the same time, they enable the simplification of highly complex scientific events into didactically manageable forms without compromising factual accuracy or disciplinary rigor [3]. Consequently, computer simulations constitute a highly effective approach for facilitating comprehension of chemistry content.

Current levels of technological equipment in schools can generally be regarded as satisfactory, and most learners today possess at least basic proficiency in operating mobile digital devices such as tablets and smartphones [4]. Although teacher competence in using modern technologies continues to improve, didactic software still does not represent a standardized or consistently applied component of everyday instruction [5, 6]. Existing research on the use of digital educational tools in chemistry—particularly computer simulations—typically focuses on their impact on the learning process and on how these tools are evaluated by both teachers and students. Findings from Czech educational studies indicate that teachers often employ computer-based tools mainly as motivational supplements rather than fully exploiting their didactic potential. One of the frequently identified barriers remains insufficient awareness of available educational software [7, 8].

Within chemistry education, oxidation–reduction reactions have long been perceived as a particularly challenging thematic area. Their difficulty is attributed to the high degree of abstraction and substantial interdisciplinary overlap inherent to redox processes. Furthermore, understanding redox reactions is foundational for subsequent chemical knowledge and essential due to their extensive application in industrial technologies, where the development and utilization of redox principles continue to advance [9].

In this study, we focus on the use of selected computer simulations of electrolysis and galvanic cells in teaching chemistry at the secondary school as a general education subject in the field of redox reactions. The subject of the research was to analyze the effectiveness of implementing appropriate simulations in the process of teaching redox reactions being considered as a difficult topic in comparison with the use of an identical real activity. The research was carried out as a pedagogical experiment applied to an available sample of secondary school students. The students' level of knowledge was assessed before and after the implementation of simulation-based instructions in one group and an identical real experiment in another group. The students' retention of knowledge was assessed through didactic tests.

2 Background

With the development and availability of digital technologies, the use of computer models in science education as an informative, motivational, or assessment tool is growing [1, 2]. Teaching supported by computer models or simulations is often referred to as model-based learning or simulation-based education. Computer simulations offer students and

teachers the opportunity to explore phenomena or objects that cannot be explored in their real form or are not effective to explore. They also offer the repeatability of demanding procedures and support the clarity, interactivity, and attractiveness of the subject matter. Another advantage is the ability to view ongoing processes at a micro level that would otherwise be inaccessible to students. This simplifies complex processes making them easier for students to understand while maintaining factual accuracy and expertise.

Simulations also make it possible to include experiments in the teaching process that would otherwise be impossible to carry out at school due to their dangerous nature or the unavailability of raw materials [3]. Simulations are thus gradually becoming an important and integral part of chemistry teaching. However, research in this area is still insufficient, especially given the rapid pace of development in digital technologies. Most studies on the effectiveness of computer simulations are conducted at universities, where simulations are created or used to support experimental activities and detailed investigation of processes. Teaching with the support of simulations at the secondary or primary level is not often reflected in studies and therefore represents a challenge for further research [10]. The main objective of our research is therefore the effectiveness of implementing computer simulations in the process of teaching chemistry at a secondary school (grammar school, for students aged 15–16), specifically in the field of electrochemistry. Students´ level of knowledge and retention of knowledge in electrochemistry was observed on the basis of various conditions for carrying out the relevant laboratory activities.

With the growing number of available computer simulations, the problem arises of how to evaluate these tools and how to select suitable simulations for teaching. Using heuristic methods, Nielson [11] determined the basic evaluation criteria for online tools based on the experience of experts. These are difficult for ordinary teachers to apply today and are supplemented by new criteria, such as the soundscape of the tool, sufficient instruction, and the balance between the amount of information in the tool and its clarity [12]. When searching for suitable simulations in the field of electrochemistry, we therefore tried to apply relevant criteria for evaluating online simulations from evaluation tools in similarly focused research [13].

Another reason for our research was the fact that teaching redox reactions is considered a critical part of the chemistry curriculum [2, 14], and examples of electrolysis and galvanic cells are often used in secondary school curricula to illustrate redox reactions. Due to the high degree of abstraction, complexity, and broad scope of the subject matter, students may encounter a number of difficulties when learning electrochemistry. Students have trouble identifying redox reactions, cannot explain the difference between electrolysis and a galvanic cell, cannot describe the passage of electric current through metals and electrolytes, incorrectly label the processes occurring at the anode and cathode in diagrams, misinterpret the main principles of electrochemistry, etc. [15].

Based on a detailed analysis of several dozen professional studies on teaching electrochemistry using simulations, we have identified the main directions of research in this area to date [16]. These include comparing the effectiveness of real experiments with simulations in order to determine students' understanding of the process at the micro level, comparing different simulations with each other in order to select the most didactically

suitable option, the selection of simulations and their combination when observing students in order to create an effective series of simulations to form a more comprehensive picture of a chemical reaction, and the creation of new simulations suitable for teaching and researching their effectiveness [16]. Studies focused on understanding oxidation-reduction processes at the micro level have concluded that computer simulations have a positive effect on improving cognitive processes in students and on deepening their understanding of chemical reactions when real experiments are combined with corresponding simulations. When observing differences in student motivation between real experiments and simulations, it was found that students were more motivated to carry out real experiments, while university students were more motivated to work with simulations, which presented the process at the micro level and enabled them to gain a deeper understanding [17, 18]. When comparing the effectiveness of simulations with each other, the authors generally did not find statistically significant differences [4, 17, 19, 20]. In a series of simulations that varied in the level of detail with which they depicted a given event, the authors concluded that it is more effective for students to first watch a simpler simulation and only then move on to more complex or sophisticated simulations to deepen their knowledge [21]. The results also show that it is more suitable for students to watch segmented simulations, which work similarly to a video that can be repeated or paused and commented on, while also setting different response conditions. It is less suitable for students to watch a series of static images or images that move but do not follow on from each other smoothly like a video [16].

3 Research Design

Based on the findings presented in the previous section, we decided to investigate the effectiveness of using computer simulations of electrolysis and galvanic cells in comparison with a real performance of identical laboratory experiments. We designed the research as a pedagogical experiment with a cross-over design in which two groups of students were monitored. One group of students was engaged in real experimental activities, while the other group was engaged in simulated experimental activities. In the following activity the groups of students were swapped.

The main objective of the research was to examine the effectiveness of implementing computer simulations of electrolysis and galvanic cells in the teaching process of redox reactions in secondary school chemistry lessons. The research sought answers to the following questions: Is the level and retention of knowledge related to the topic of redox reactions achieved by secondary school students when implementing computer simulations the same as when implementing corresponding real experiments? How do students evaluate the teaching aids used, i.e., computer simulations compared to real experiments?

To achieve the above-stated goal, we used a pedagogical experiment in a cross-over design with didactic tests (pre-test, post-test, and retention test) to compare the level of knowledge and students' retention of knowledge depending on using either selected simulations or corresponding real experiments. Then we used questionnaires to detect students' attitudes and opinions on the teaching tool used. The research methods and the selection of the research sample of students were inspired by previous studies [2,

19, 22, 23], in which the authors worked with available respondents at the institutions where they themselves worked. Our research sample included secondary school students, specifically first-year students (aged 15–16) since the topic of electrochemistry is part of the chemistry curriculum taught at this grade. These students from one class were divided into two groups. We therefore opted for a purposeful non-probability sample to reveal the influence of two selected teaching aids on the level of knowledge acquired by students and the retention of the knowledge. The students were divided into two groups of 15, group A (average chemistry grade 1.57) and group B (average chemistry grade 1.50).

Two computer simulations were selected for the educational experiment: one modelling electrolysis and one modelling a galvanic cell. They were selected on the basis of a detailed didactic analysis of a total of 18 simulations (9 electrolysis and 9 galvanic cells) found using the Google search engine and the English keywords "electrolysis simulation/animation/visualization" and "galvanic/voltaic cell simulation/visualization/animation." For the didactic analysis of the simulations found we used criteria that generally evaluate simulation as a didactic tool. Following the model of Asirvatham [24], who published general criteria for evaluating available online teaching materials, we reformulated the individual criteria into questions for easier evaluation. Other specific evaluation criteria focused on content, didactic and user requirements, and technical parameters. We discussed our didactic analysis in detail in the article [13].

4 Research Implementation

The pedagogical experiment was designed to be carried out during scheduled chemistry lessons at a selected secondary school. The planned interventions and data collection took place during three consecutive lessons in the following order: (1) Lecture on Redox Reactions – Introduction to Electrochemistry (2) Pre-test (3) Laboratory exercise on the topic of electrolysis (4) Laboratory exercise on the topic of Galvanic cells (5) Post-test (6) Retention test (7) Questionnaire.

The first lesson was conducted as an explanation of new subject matter on redox reactions in which the concepts of oxidation, reduction, anode, cathode, electrolysis, galvanic cell and Beketov's series of metals were explained. At the end of the lesson, the students were given a didactic pre-test to determine their level of knowledge of the subject matter. The second lesson consisted of a laboratory activity on electrolysis that was introduced in the previous lecture as an example of the use of redox reactions. Group A worked with a simulation of electrolysis and group B with a real experiment as in normal laboratory activities. Each group received a worksheet that guided the students through the simulated or real environment. The worksheet introduced each student to the set of tools they would be working with during the laboratory activity, the procedure, and additional questions that guided the students what to focus on and what to think about during the experiment. Students who worked with the simulation had a picture of the simulation they were to work with and a description of all its functions that they could set (see Fig. 1) in the section on the tools presented. Teaching with computer simulation was carried out in a computer classroom.

The students who worked in a real-world environment had a photograph of the electrolysis process using the equipment available to them and a description of this

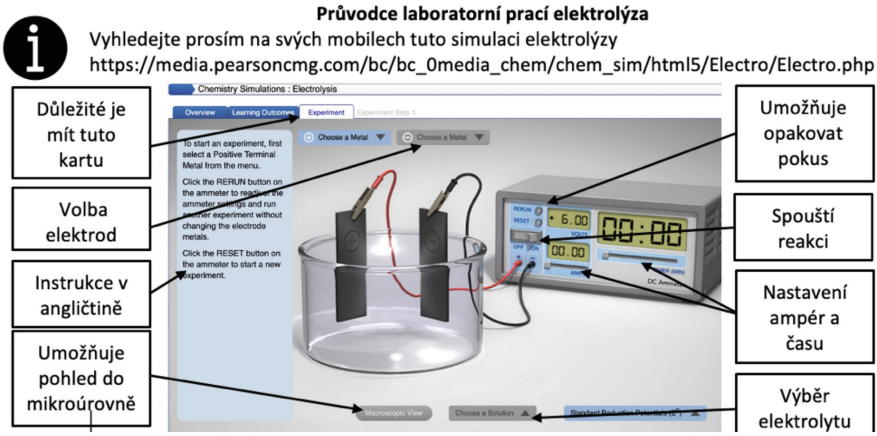


Fig. 1. Sample instructions (in Czech) for working with the electrolysis simulation (in English) in the worksheet

image (see Fig. 2) in the section on the equipment presented. This group of students worked in a specialized physics and chemistry classroom.

Fig. 2. Sample instructions (in Czech) for working with a real electrolysis experiment in the worksheet

The reason for compiling the worksheets in this way was to set the same conditions for both groups. Both groups had a picture of electrolysis and a description of everything they could do and what they would need. The description of the procedure in a real environment was designed to mirror the way students proceed when working with simulations. The worksheet also served as a record of the students' work results, i.e., as a laboratory report.

During the third lesson, laboratory work was carried out on the topic of galvanic cells. This time, the group that had worked with the simulation during the second lesson performed a real activity with the support of a similarly structured worksheet (Fig. 3).

The group that worked in a real environment during the second lesson now worked with a galvanic cell simulation supported by a worksheet (see Fig. 4).

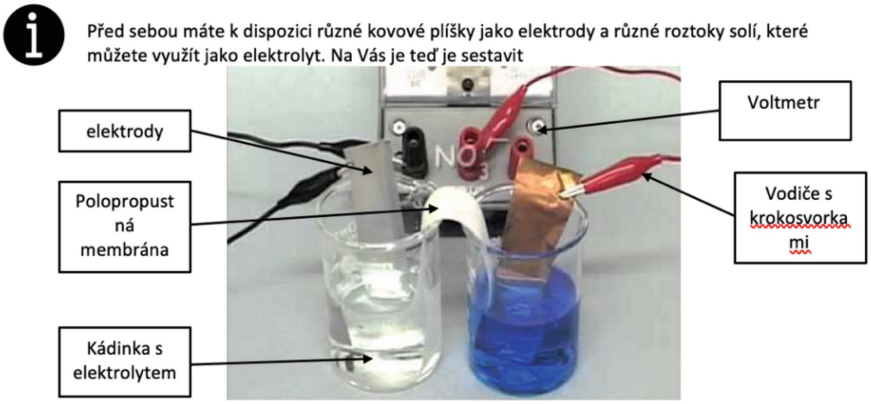


Fig. 3. Sample instructions (in Czech) for working with a real galvanic cell experiment in the accompanying worksheet.

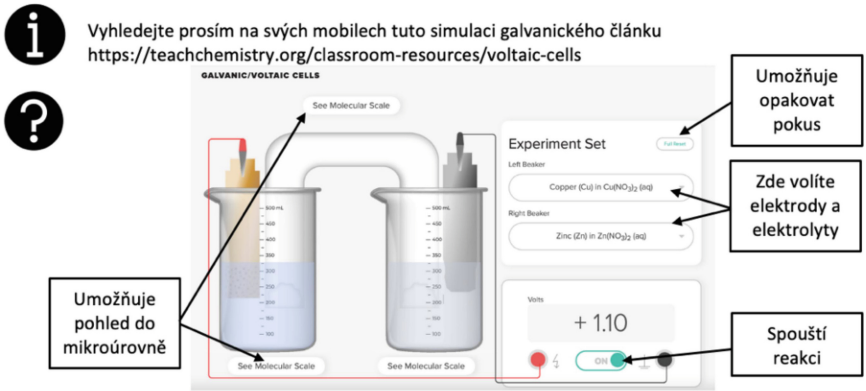


Fig. 4. Sample instructions (in Czech) for working with the galvanic cell simulation (in English) in the accompanying worksheet

At the end of the third lesson, both groups were given a didactic post-test and, after three months, a retention test with the same content as the pre-test. The didactic test was developed on the basis of teaching objectives that teachers usually formulate at this level of education in the thematic Electrochemistry unit. Semi-structured interviews with teachers were used to identify these objectives, and based on the results, specific thematic areas were identified. These areas served as the basis for the items of the didactic: Explanation of electrolysis, galvanic cells, and differences between processes; Course of reactions at electrodes; Modification of reaction conditions; and Condition of electrodes after reaction [25]. The didactic test compiled in this way was piloted, and based on the results, three items whose wording was not understandable to students were modified. In the retention test, the items were rephrased while retaining their original meaning. In addition, confidence levels were added to the post-test and retention test. The confidence levels for each question were analyzed separately, thus completing the overall

picture of the students' progress in knowledge and their confidence in understanding the topic. The didactic test had two parts, one focused on electrolysis and the other on galvanic cells. The Electrolysis part included nine items (two open-ended and seven close-ended, with two items requiring students to explain their answers). The Galvanic Cell part contained ten items (one open-ended and nine close-ended, with two items requiring students to explain their answers), one item concerned the comparison of electrolysis and galvanic cells.

5 Results

The cross-over design of the pedagogical experiment yielded the following results. In terms of student knowledge, it can be stated that almost all students in both groups achieved better results in the didactic post-test than in the pre-test, while indicating a medium level of confidence in their answers (around 5 on a ten-point scale). However, if we also take into account the results of the retention tests, it can be said that students working with simulations achieved better results in terms of knowledge progress than those working with real experimental activities. Students using simulations also indicated a higher level of confidence in their answers. In terms of knowledge retention, i.e., when comparing the average results of the post-test and the knowledge retention test after three months, there was only a slight deterioration with all students (a few percent). The progress of students in both groups working with different teaching aids was evident. Results show that when working with the electrolysis simulation, the students achieved better or the same results in the post-test in comparison with the students working in the real environment, i.e., there was a significant progress in their knowledge of the process itself compared to students' initial knowledge. As far as long-term retention of information gained is concerned, it cannot be said unequivocally that work with simulations is more suitable for all teaching objectives. For example, in Item 3 where students were asked to indicate the direction of electron flow in a diagram, there was a slight improvement in students' performance after using the simulation; however in the retention test, no one answered the item correctly, regardless of the teaching tool used. In Item 2 where students named the electrodes in the diagram, all students in both groups performed at the same level in the post-test, but in the retention test, no student from the group working with the simulation named the electrodes correctly, while students working in the real environment performed significantly better. However, in all other items, the students working with the simulation showed overall better results in terms of knowledge retention.

The success rate of students in didactic tests after working with a galvanic cell, are not as clear-cut as in the case of electrolysis. In the post-test on galvanic cells, students working in a real environment performed better or equally well as students working with a simulation. The only difference can be seen in identifying the purpose of the salt bridge in a galvanic cell. In terms of the retention of knowledge, the use of a real environment shows better results than a simulation, unlike in the case of electrolysis.

In searching for an answer to the research question of whether the level and retention of students' knowledge is achieved when implementing computer simulations in the topic of redox reactions compared to implementing the corresponding real environment,

we used a simple statistical analysis of data (Student's paired t-test). The research question was divided into two sub-questions, for which the following null hypotheses were formulated: (1) After implementing the simulation of electrolysis and galvanic cells, the students did not achieve better results and did not retain their knowledge in the long term. (2) After implementing the real experiment, the students did not achieve better results and did not retain their knowledge in the long term.

In both cases, it was possible to reject the hypothesis for comparing the pre-test and post-test, i.e., both the simulation and the real environment resulted in a statistical increase in students' knowledge at a significance level of 0.05 ($p = 0.000016$, $p = 000002$) after the implementation of the teaching. However, when comparing the pretest and retention test, it was possible to reject the hypothesis only for the use of simulation ($p = 0.039$), but not for the use of a real environment ($p = 0.4022$), i.e., after using simulation, there was a statistically significant increase in students' knowledge. To these conclusions it should be added that when comparing the post-test and retention test, there were no significant changes in the knowledge achieved in both cases.

We paid specific attention to the open-ended items in the didactic tests, even though they were also evaluated bipolarly (correct – incorrect). In the Electrolysis section, there were two open-ended questions and two closed-ended questions with explanations of the answers. The students answered mostly correctly according to the codebook in both the pre-test and post-test. When working with the simulation they often showed higher confidence values, and in the retention test, a larger proportion of them remained at the same level as in the pretest (47%), unlike students working in a real environment (27%). The definitions of electrolysis often referred to the use of the process rather than explaining the principle (e.g., it is the plating of metals, the opposite of a galvanic cell, it is a reaction in a single beaker, etc.). Most incorrect answers also remained without any explanation, or with the comment "I don't know." In the Galvanic Cell section, there was one open-ended question concerning the explanation of its principle. This item showed no significant improvement compared to the initial knowledge. Among the answers that were not incorrect but did not describe the phenomenon in question, there were often one-word answers such as "battery." Thus, the students again did not describe the principle of the phenomenon, but they knew its practical application.

After the retention test, the students were also given a questionnaire to assess their interest in, usefulness of, and attractiveness of the activities carried out. The questionnaire was an adapted version of the well-known IMI (Intrinsic Motivation Inventory) tool according to Teplá et al. [26], who used the questionnaire in her research on the use of 3D models in teaching. The IMI questionnaire can be used to measure seven areas of intrinsic motivation and self-regulation, namely: interest/enjoyment, perceived competence, effort, value/usefulness, perceived pressure, and tension, perceived choice in performing the activity, and experience of social relationships. From these areas, we selected only items focused on usefulness and interest, and we reformulated the items from the above-mentioned questionnaire to reflect students' motivation when using simulation or real experiments. All students were given two questionnaires, one focusing on their experiences with both simulations and the other on real activities. Students considered working with simulations and real-life activities to be interesting, fun, and important, but they are not entirely sure whether the simulations helped them understand

the subject matter. With the slight preference of simulations, students considered both options to be a way of making lessons more interesting and would generally be willing to participate in similar activities again, as they would have no problem with their use.

6 Limitations of the Research

The research conducted has its limitations that should be taken into account. These mainly concern the short-term nature of the interventions, i.e., the use of simulations and real-life activities, which took place only during three lessons in one school year and in one class at one school. To minimize this limitation, it would be advisable to include these activities over a longer period of time and in various subjects. We are also aware that the conclusions we draw after conducting the educational experiment with selected high school students cannot be clearly generalized to the secondary school population throughout the country, nor to all students at the school itself, nor to the specific activities used. Therefore, only simple statistics was used to evaluate the data obtained, and greater emphasis was placed on describing and interpreting the identified connections in specific cases. The scale used to measure the certainty of students' answers in didactic tests also has its limitations. Traditionally, its application is associated with the modification of points of individual items in the didactic test, e.g., in the case of an incorrect answer and a high level of confidence expressed, or vice versa. In our case, this was not used, but we tried to identify possible misconceptions in the confidence expressed in the students' answers.

7 Conclusions

The development of digital technologies has led to their increasingly frequent and easier use in teaching science subjects. Computer simulations aim to improve understanding of real systems and offer new possibilities for teaching. The aim of our study was to examine the effectiveness of implementing computer simulations in secondary school chemistry teaching in the field of electrochemistry, i.e., to design and optimize the implementation of computer simulations of electrolysis and galvanic cells in secondary school chemistry teaching. The aim was to identify the level of knowledge after using computer simulations in comparison with activities in a real environment and the attitudes of students towards the implementation of these activities in teaching.

The pedagogical experiment shows that the use of both simulations and real-life environments improved students' knowledge with computer simulations having a more significant effect (especially in the topic of electrolysis). Students working with simulations responded better to items related to the micro world, while students working in a real environment responded better to items related to the macro world (more significantly in the topic of galvanic cells). However, in neither variant there was a progress in students' ability to correctly identify the direction of electron flow in an ongoing chemical reaction, and this was also the item in all didactic tests in which students were least confident in their answers. On the contrary, the greatest progress was achieved in both cases in the ability to determine the state of the electrodes after the chemical reactions had taken place. In a supplementary survey focusing on students' opinions and attitudes

towards the use of simulation or real environments, students expressed slightly greater interest in the use of simulation in teaching process. Overall, however, they considered both simulations and real experiments to be interesting, fun, and useful activities that are worth including in science education.

Acknowledgement. The authors would like to thank for the support of the Erasmus+ project STEMkey and the Faculty of Education Charles University Cooperatio programme.

References

1. Fadzli, S. et al.: Environment based virtual interaction to enhance motivation of STEM education: the qualitative interview design and analysis. Educ. Inf. Technol. **25**, 775–790 (2020)
2. Huri, N.H., Karpudewan, M.: Evaluating the effectiveness of integrated STEM-lab activities in improving secondary school students' understanding of electrolysis. CERP. **20**, 495–508 (2019)
3. Tang, H., Abraham, M.R.: Effect of computer simulations at the particulate and macroscopic levels on students' understanding of the particulate nature of matter. J. Chem. Educ. **93**(1), 31–38 (2016)
4. Mikeš, J., Hlavatý, J.: Možnosti využití virtuálních laboratoří při výuce chemie na středních školách. Praha: Chemické listy. **114**(4), 291–294 (2020)
5. Oujezdský, A., Slíva, A., Brázda, R.: Using ICT in education: measuring systems interfaced to computers. In: Inted 2015 Proceedings, pp. 7509–7512 (2015)
6. Vojíř, K., Rusek, M.: Science education textbook research trends: a systematic literature review. Int. J. Sci. Educ. **41**(11), 1496–1516 (2019)
7. Chroustova, K., Bilek, M., Sorgo, A.: Validation of theoretical constructs toward suitability of educational software for chemistry education: differences between users and nonusers. J. Balt. Sci. Educ. **16**(6), 873–897 (2017)
8. Balogh, Z., Turčáni, M., Magdin, M.: Design and creation of a universal model of educational process with the support of petri nets. In: Li, S., Jin, Q., Jiang, X., Park, J.J. (eds.) Frontier and Future Development of Information Technology in Medicine and Education, pp. 1049–1060. Springer Netherlands, Dordrecht (2014)
9. De Jong, O., Acampo, J., Verdonk, A.: Problems in teaching the topic of redox reactions: actions and conceptions of chemistry teachers. J. Res. Sci. Teach. **32**(10), 1097–1110 (1995)
10. Campos, N., Nogal, M., Cáliz, C., Juan, A.A.: Simulation-based education involving online and on-campus models in different European universities. Int. J. Educ. Technol. High. Educ. **17**, 1–15 (2020)
11. Nielson, J.: Heuristic Evaluation of User Interfaces. Association of Computing Machinery Conference, Seattle, WA (1990)
12. Leacock, T., Nesbit, J.: A framework for evaluating the quality of multimedia learning resources. Educ. Technol. Soc. **10**, 44–59 (2007)
13. Henychová, H., Machková, V.: Počítačové animace a simulace elektrolýzy 2 a galvanického článku jako prostředek podpory názornosti ve výuce chemie na střední škole. Biologie-Chemie-Zeměpis. **32**(2), 2–9 (2023)
14. Rychtera, J. et al.: Která jsou klíčová, kritická a dynamická místa počáteční výuky chemie v České republice? Arnica. **8**(1), 35–44 (2018)
15. Tsaparlis, G.: Teaching and learning electrochemistry. Isr. J. Chem. **59**(10), 478–492 (2018)

16. Henychová, H., Machková, V.: The use of computer simulation to improve learning of redox reactions. In: Active Science Education, pp. 35–48. Pedagogiucal University of Cracow, Krakow (2022)
17. Ramírez, J. et al.: A virtual laboratory to support chemical reaction engineering courses using real-life problems and industrial software. Educ. Chem. Eng. **33**, 36–44 (2020)
18. Vries, L., May, M.: Virtual laboratory simulation in the education of laboratory technicians–motivation and study intensity. Biochem. Mol. Biol. Educ. **47**(3), 257–262 (2019)
19. Hensen, C., Glinowiecka-Cox, G., Barbera, J.J.: Assessing differences between three virtual general chemistry experiments and similar hands-on experiments. Chem. Educ. **97**(3), 616–625 (2020)
20. Hendel, S.J., Young, E.R.: Introduction to electrochemistry and the use of electrochemistry to synthesize and evaluate catalysts for water oxidation and reduction. J. Chem. Educ. **93**, 1951–1956 (2016)
21. Cole, M., Rosenthal, D., Sanger, M.: Two studies comparing students' explanations of an oxidation-reduction reaction after viewing a single computer animation: the effect of varying the complexity of visual images and depicting water molecules. CERP. **20**, 738–759 (2019)
22. Carlos, D., Ruben, M., Emilio, G., Mercedes, O., Alonso, M.: A virtual lab as a complement to traditional hands-on labs: characterization of an alkaline electrolyzer for hydrogen production. Educ. Chem. Eng. **23**, 7–17 (2018)
23. Nxumalo-Dlamini, N.L., Gaigher, E.: Teachers' use of computer-based simulations in teaching electrolysis: a case study in Eswatini. Afr. J. Res. Math., Sci. Technol. Educ. **23**(3), 320–331 (2018)
24. Asirvatham, D.: Quality Framework for Assessment of Multimedia Learning Materials Version 1.0. Elsevier Ltd.. https://www.researchgate.net/publication/257718377_Quality_Framework_for_Assessment_of_Multimedia_Learning_Materials_Version_10 (2012). Last accessed 2021/14/08
25. Henychová, H.: Dissertation Thesis. Faculty of Education Charles University, Prague (2023)
26. Teplá, M., Teplý, P., Distler, P., Šmejkal, P.: Vzdělávací 3D modely Corinth a jejich využití ve výuce chemie a ostatních přírodovědných předmětů. Chemické Listy. **115**(7), 383–386 (2021)

An Indicator-Based Framework for Adaptive Distance Learning: Design of a Learner-Needs Model and Expert Validation

Iryna Bilous[1](✉), Viktoriia Homotiuk[1], Oksana Gomotiyk[1], Azeta Tartaraj[2], Beata Akimjakova[3], and Oto Mucka[3]

[1] West Ukrainian National University, 11 Lvivska st., 47000 Ternopil, Ukraine
1982bilous@gmail.com, o.homotiuk@wunu.edu.ua
[2] New Campus, Rr. Miqësia, Spitallë, Durrës 2009, Albania
azetatartaraj@uamd.edu.al
[3] Faculty of Theology, Catholic University in Ružomberok, Spišská Kapitula 12, 053 04 Spišské Podhradie, Slovakia
beata.akimjakova@ku.sk

Abstract. Higher education distance learning increasingly requires transparent, implementable personalization to address heterogeneity in students' prior knowledge, pace, self-regulation, engagement, and accessibility needs. This paper proposes an indicator-based framework for adaptive distance learning in Moodle that links a Learner-Needs Model, a 20-item indicator matrix (needs → measurable LMS signals → decision rules), and Moodle-native adaptation mechanisms (conditional access, activity completion, assessment settings, and targeted support actions). To validate the framework at the design stage, an expert panel of nine university instructors (N = 9) evaluated each indicator on two 4-point criteria: relevance (construct representation) and feasibility in Moodle (measurability and implementability). Content validity was quantified using item-level and scale-level Content Validity Index (I-CVI; S-CVI/Ave). The results demonstrate high content validity of the indicator set: S-CVI/Ave = 0.95 for relevance and 0.91 for feasibility; all indicators met the minimum I-CVI threshold (≥0.78). Four indicators were borderline on feasibility (I-CVI = 0.78) and were earmarked for minor operational clarification related to reporting/analytics configuration and context-specific thresholding. The validated framework provides a reproducible, platform-aligned instrument for designing adaptive Moodle courses in higher education and establishes a basis for subsequent pilot studies of learning outcomes, persistence, and responsible analytics use.

Keywords: Adaptive Learning · Distance Education · Moodle · Learner Needs

1 Introduction

Higher education distance and blended learning increasingly requires implementable personalization to address learner heterogeneity in prior knowledge, pace, self-regulation, engagement, and accessibility needs. Adaptive learning provides a principled

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 75–84, 2027.
https://doi.org/10.1007/978-3-032-32104-6_7

basis for such personalization through learner modeling and rule-based instructional decisions [1, 2]. As a widely adopted LMS, Moodle is a practical platform for this purpose because it offers native mechanisms for transparent course design and controllable adaptation (e.g., conditional access, activity completion, and assessment settings) [3–5].

A persistent design gap, however, is the lack of reusable, operational mappings between learner needs and measurable LMS signals that can trigger pedagogically justified actions. Without explicit indicators and decision rules, "adaptive" LMS implementations often remain feature-driven and difficult to evaluate or govern. Moreover, the use of LMS traces for intervention must rely on interpretable criteria and purpose-bound data use to avoid overreach and misinterpretation [6].

To address this gap, this paper proposes an indicator-based framework for adaptive distance learning in Moodle that integrates a Learner-Needs Model, a 20-item indicator matrix linking needs to measurable Moodle signals, decision rules, and platform-native adaptation mechanisms, and design-stage expert validation using the Content Validity Index (CVI) [7–9]. An expert panel of nine university instructors (N = 9) evaluated each indicator on relevance and feasibility in Moodle (4-point scales). The instrument demonstrated high content validity (S-CVI/Ave = 0.95 for relevance; 0.91 for feasibility); all items met the minimum I-CVI threshold (≥0.78), with four indicators at the feasibility borderline requiring minor operational clarification.

2 Related Work

Adaptive learning has a long research tradition in adaptive hypermedia and adaptive educational systems, where personalization is structured around a learner model, a domain model, and an adaptation (decision) model [1, 2]. Typical mechanisms include adaptive navigation and presentation, personalized feedback, and adaptive assessment, all aimed at aligning instruction with the learner's current state and needs [1, 2]. A persistent practical challenge is translating these principles into implementable rules and measurable variables in authentic LMS-based learning environments.

In higher education distance learning, learner modeling must extend beyond prior knowledge to incorporate self-regulation and motivational factors that shape persistence, pacing, and task completion. Self-regulated learning highlights planning, monitoring, and control strategies as predictors of online learning success [10, 11], while motivation frameworks (e.g., self-determination theory) explain how autonomy, competence, and task value influence engagement and sustained effort [12]. These perspectives support the inclusion of indicators capturing not only cognitive progress but also pacing behaviors, deadline adherence, and help-seeking patterns. Recent higher-education research further specifies key learner- and context-level prerequisites for adaptive online learning, showing that student readiness is shaped by technology access, learner perceptions, and lecturer roles, while digitalisation intensifies demands for virtual pedagogical competence and structured instructional support [13, 14].

Inclusive and accessible design is another foundation for adaptive distance learning. Universal Design for Learning (UDL) emphasizes systematic variability in representation, engagement, and action/expression to provide meaningful alternatives for diverse

learners [15]. In LMS contexts, UDL principles are commonly enacted through multimodal content, alternative activity formats, and flexible assessment options, which can be combined with conditional rules to support structured yet inclusive learning pathways.

Accordingly, Moodle provides a practical substrate for rule-based adaptation through native features such as Restrict access and Activity completion [4, 5]. However, many Moodle implementations remain feature-centric and lack reusable mappings between learner needs, platform-measurable indicators, and pedagogically justified decision rules. In parallel, the growth of Moodle analytics expands opportunities for data-informed interventions [16], but also reinforces the need for transparent design logic and responsible governance given the ambiguity of behavioral traces and associated privacy and fairness risks [6]. Therefore, analytics signals should be embedded into interpretable pedagogical rules rather than used as opaque triggers [17, 18].

When adaptation is operationalized through indicator sets and decision rules, a methodological gap concerns how to validate indicator relevance and feasibility before costly pilots. Content validity approaches, particularly the Content Validity Index (CVI), offer a structured expert-based procedure to assess construct representation and practical implementability [7–9]. While prior work often reports adaptation principles or LMS feature use, comparatively fewer studies provide validated indicator instruments that explicitly connect learner needs → measurable LMS signals → adaptation actions.

Recent Moodle-focused research also differentiates between assessing whether a course is structurally prepared for adaptive learning and specifying how adaptive decisions should be triggered and enacted. For example, Morze et al. propose an implementation framework spanning technical, pedagogical, and administrative stages, including analytics- and AI-supported personalization [19]. Semerikov et al. developed ALRAF as a readiness assessment instrument for estimating whether a Moodle course contains key prerequisites for adaptive learning [20]. In contrast, the present study focuses on the next step – process-level adaptation logic – by defining learner-needs indicators as measurable LMS signals and linking them to explicit decision rules and Moodle-native actions, providing a transparent and reproducible operational layer.

UDL-oriented Moodle studies similarly emphasize that inclusivity requires explicit pedagogical strategy rather than technical features alone. Recent frameworks and case studies show how Moodle mechanisms (conditional activities, flexible assessment, multimodal content) can support inclusive and adaptive learning when guided by UDL principles [21, 22]. Related learning-analytics work connects UDL-guided design with interpretable patterns of learner interaction, underscoring the value of transparent indicators for supporting diverse learners [23]. At the same time, validated measurement work on social competencies and soft skills development underscores the importance of collaboration-related and transversal outcomes in LMS-based course design, whereas teacher acceptance of educational technology remains a practical adoption constraint – together reinforcing the need for transparent, indicator-driven adaptation that is both measurable and feasible in real Moodle implementations [24–26]. Building on these directions, our framework links learner needs (including accessibility and self-regulation) to platform-aligned indicators and concrete Moodle actions, and validates the indicator set through expert appraisal.

3 Materials and Methods

This study follows a design science approach: it develops a reusable design artifact – an indicator-based adaptive-learning framework for Moodle – and evaluates it through expert appraisal prior to classroom-scale piloting. The artifact comprises a Learner-Needs Model, an indicator matrix linking needs → measurable indicators → decision rules → Moodle mechanisms → expected effects, and a design-stage validation protocol.

3.1 Indicator Set Development and Structure

Scope and Operationalization. The indicator set targets higher education distance/blended courses delivered in Moodle. Each indicator is operational: it specifies what is measured, the Moodle data source, and a corresponding adaptive action implementable via standard Moodle features (e.g., conditional access, activity completion) [4, 5].

Need Clusters and Coverage. The Learner-Needs Model groups indicators into six clusters grounded in adaptive learning, self-regulation, motivation, inclusion, and analytics-informed support [1, 2, 6, 10–12, 15, 17, 18]. Indicators were produced through an iterative workflow (initial pool → screening → consolidation → final set), where screening ensured theoretical alignment and Moodle measurability, and consolidation removed redundancy and harmonized item wording. The final instrument includes $k = 20$ indicators covering cognitive baseline/progress, assessment behavior, self-regulation (pacing), engagement, accessibility/UDL, digital conditions, support needs, and collaboration.

Item Structure. Each item includes: indicator statement, operational definition, Moodle evidence source, and adaptive action. This format reduces ambiguity and supports implementation feasibility in Moodle [4, 5].

3.2 Expert Panel and Procedure

Experts. Content validation involved $N = 9$ university instructors with ≥ 3 years of experience designing and delivering Moodle-based distance or blended courses, supporting the credibility of feasibility judgments. The panel size aligns with common CVI recommendations for early-stage instrument development [7–9].

Ratings. Experts rated each indicator on two 4-point criteria:

1. Relevance (construct representation for higher education distance learning), and
2. Feasibility in Moodle (measurability and implementability using standard Moodle functionality).

A 4-point scale (1–4) was used to avoid a neutral midpoint and enable dichotomization into acceptable (3–4) vs. unacceptable (1–2) ratings [7–9].

Data Handling. Ratings were analyzed in aggregate; no personally identifying expert information is reported.

3.3 Content Validity Analysis (CVI)

Item-Level CVI (I-CVI) For each indicator i, the Item-level Content Validity Index is:

$$I - CVIi = \frac{n_{(3-4)}}{N} \tag{1}$$

where n(3–4) is the number of experts rating the item 3 or 4, and N is the panel size [7–9].

Scale-Level CVI (S-CVI/Ave). Overall content validity is summarized as the mean I-CVI across all k indicators:

$$S - \frac{CVI}{Ave} = \frac{1}{k} \sum_{i=1}^{k} I - CVI_i \tag{2}$$

CVI was computed separately for Relevance and Feasibility [7–9].

Decision Rule. Following CVI guidance, indicators with I-CVI < 0.78 (i.e., <7 of 9 experts rating 3–4) are flagged for revision (clarifying operational definitions, thresholds, or Moodle evidence/actions) rather than immediate deletion; S-CVI/Ave is used as an overall adequacy signal with refinement focused on borderline items [7–9].

4 Results

Expert ratings were obtained for k = 20 indicators, each evaluated on two criteria (relevance and feasibility in Moodle) using a 4-point scale. With N = 9 experts, this yielded 180 relevance ratings and 180 feasibility ratings across the instrument.

4.1 Overall Content Validity (S-CVI/Ave)

Table 1. Content validity summary (N = 9; k = 20)

Metric	Relevance	Feasibility in Moodle
Acceptable ratings (3–4), n/N_total	171/180	164/180
S-CVI/Ave	**0.95**	**0.91**
Min I-CVI	0.89 (8/9)	0.78 (7/9)
Max I-CVI	1.00 (9/9)	1.00 (9/9)

These results indicate high content validity at the scale level for both relevance and feasibility, consistent with CVI interpretation guidance [7–9] (Table 1). All indicators achieved I-CVI ≥ 0.78 on feasibility and I-CVI ≥ 0.89 on relevance, implying that the indicator set is acceptable for implementation with only targeted refinements.

Table 2. Item-level CVI for the indicator set (N = 9)

ID	Cluster	Short label	I-CVI (Relevance)	I-CVI (Feasibility)
I1	Cognitive baseline	Pre-test score baseline gaps	1.00 (9/9)	0.89 (8/9)
I2	Cognitive baseline	Checkpoint quiz mastery score	1.00 (9/9)	1.00 (9/9)
I3	Progress stability	Quiz score variability (SD)	1.00 (9/9)	**0.78 (7/9)**
I4	Assessment	Multiple attempts without mastery	0.89 (8/9)	0.89 (8/9)
I5	Assessment	Critical-item low performance	0.89 (8/9)	**0.78 (7/9)**
I6	Self-regulation	≥2 late submissions/module	1.00 (9/9)	1.00 (9/9)
I7	Self-regulation	Last-minute submission pattern	0.89 (8/9)	0.89 (8/9)
I8	Engagement	Low weekly activity	1.00 (9/9)	1.00 (9/9)
I9	Engagement	7-day inactivity	0.89 (8/9)	**0.78 (7/9)**
I10	Engagement	Low topic completion rate	1.00 (9/9)	0.89 (8/9)
I11	Motivation	Low task value (survey)	0.89 (8/9)	1.00 (9/9)
I12	Motivation	Low self-efficacy (survey)	1.00 (9/9)	0.89 (8/9)
I13	Accessibility/UDL	Need alternative format	0.89 (8/9)	1.00 (9/9)
I14	Accessibility/UDL	Modality preference	1.00 (9/9)	0.89 (8/9)
I15	Digital conditions	Unstable access (survey)	0.89 (8/9)	1.00 (9/9)
I16	Digital literacy	Low digital literacy/onboarding	1.00 (9/9)	0.89 (8/9)
I17	Support need	Frequent help requests	0.89 (8/9)	1.00 (9/9)
I18	Collaboration	Low group participation	1.00 (9/9)	**0.78 (7/9)**
I19	Feedback responsiveness	Delayed response to feedback	0.89 (8/9)	1.00 (9/9)
I20	Path preference	Path choice preference	1.00 (9/9)	0.89 (8/9)

4.2 Item-Level Results (I-CVI) and Refinement Targets

Feasibility-borderline indicators (I-CVI = 0.78; 7/9) (Table 2). Four indicators reached the minimum feasibility threshold and were therefore earmarked for minor operational clarifications (rather than removal), in line with the refinement rule for CVI-based instrument development [7–9]:

- I3 (score variability / SD): clarify that SD is computed via Gradebook export (CSV) or a simple analytics script; specify a default threshold (e.g., 15 percentage points) as a starting point, to be tuned per discipline.
- I5 (critical-item performance): specify prerequisites for feasibility (e.g., enabling Quiz statistics and tagging items in the question bank); offer a fallback option (manual review of tagged items) when statistics are not enabled.
- I9 (7-day inactivity): define inactivity operationally using Moodle Logs/Course participation report (no events within 7 days) and note that the threshold can be adjusted for module duration.
- I18 (group participation): specify the collaborative activity type (Forum/Workshop/Wiki/Group assignment) and define "participation" as ≥ 1 contribution by midpoint to avoid ambiguity across course designs.

Overall, the expert panel provided strong support for the indicator set's relevance (11 items with I-CVI = 1.00; 9 items with I-CVI = 0.89) and high feasibility (8 items at 1.00; 8 items at 0.89), with feasibility clarifications concentrated in indicators that depend on advanced reporting/analytics settings or course-specific definitions.

5 Discussion

The expert validation provides strong evidence that the proposed indicator set is content-valid for designing adaptive distance learning in higher education Moodle courses. Scale-level results (S-CVI/Ave = 0.95 for relevance; 0.91 for feasibility) indicate high expert agreement that the indicators represent meaningful learner-need constructs and are largely implementable with standard Moodle capabilities, consistent with CVI guidance for early-stage appraisal [7–9]. The relevance–feasibility gap is itself informative: it suggests that adaptation goals are well understood pedagogically, while implementation in routine LMS workflows may still require additional configuration, reporting literacy, or institutional support.

Design Implications. Most indicators were rated as fully feasible (I-CVI = 1.00) or feasible with minor edits (I-CVI = 0.89). This supports an implementation strategy centered on Moodle-native mechanisms (e.g., conditional access and completion tracking) rather than external adaptive platforms [4, 5], and reinforces the artifact's design-science character as a reusable mapping from learner needs to measurable LMS signals and explicit adaptive actions that can be encoded in course settings [27].

Borderline Feasibility and Implementation Pathways. Four indicators reached the minimum feasibility threshold (I3, I5, I9, I18; I-CVI = 0.78). Their common constraint is dependence on aggregated computations not available in core conditional rules, fine-grained assessment evidence beyond activity-level grades, or time-based inactivity and

collaboration metrics. Feasibility can be strengthened without changing pedagogical intent by specifying accessible evidence sources, default thresholds as starting points, and fallbacks when advanced reports are unavailable [7–9]. Concretely: score variability (I3) can be computed via Moodle analytics/custom aggregations or grade history exports, with notifications governed by transparent rules [13, 16]; critical-item performance (I5) can be approximated through competency-based design and pass conditions aligned with critical tasks rather than question-level gating [14]; inactivity (I9) requires time-based re-engagement mechanisms (e.g., reminders via a re-engagement plugin or workplace rules) because Moodle conditions typically trigger on actions; and collaboration (I18) is most feasible when defined through measurable participation proxies (e.g., minimum contributions by a checkpoint using completion tracking) rather than automated quality judgments [5]. Overall, the pattern is expected: analytics enlarges the intervention toolbox [6, 7] but increases requirements for configurability, interpretability safeguards, and context-sensitive thresholding [8].

Responsible, Interpretable Adaptation and AI-Enabled Contexts. The framework intentionally favors transparent, rule-based adaptation (signal → rule → action) to support interpretability and accountable, purpose-bound data use given the ambiguity and ethical sensitivity of LMS traces [6]. In AI-enabled Moodle ecosystems, the indicator matrix can additionally serve as a governance layer by constraining recommendations and interventions to explicitly defined indicators, thresholds, and policies aligned with learning outcomes and assessment logic, thereby improving traceability for instructors, students, and quality reviewers.

Limitations and Next Steps. CVI provides expert evidence of content quality but does not establish learning impact or predictive validity. The expert panel included only instructors; future validation may involve instructional designers and students. The next step is piloting in authentic Moodle courses and evaluating outcomes (achievement, persistence, learner experience) and potential differential effects across learner subgroups, consistent with responsible analytics principles [6, 17, 18]. Future iterations may also refine the Learner-Needs Model by separating a dedicated social interaction cluster beyond the current collaboration indicator (I18).

6 Conclusion

This paper proposed and validated an indicator-based framework for adaptive distance learning in higher education using Moodle, combining a Learner-Needs Model, an operational indicator matrix (needs → Moodle signals → decision rules → adaptive actions), and a CVI-based expert validation protocol. Expert evaluation by N = 9 university instructors demonstrated high overall content validity of the indicator set (S-CVI/Ave = 0.95 for relevance; 0.91 for feasibility), with only four indicators requiring minor feasibility clarifications (borderline at the 0.78 threshold). These results support the framework's readiness for implementation as a practical, transparent design instrument for adaptive Moodle course development, and motivate the next research step: pilot deployment and outcome evaluation in authentic higher education settings.

References

1. Brusilovsky, P.: Adaptive hypermedia. User Model. User-Adap. Inter. **11**, 87–110 (2001)
2. Shute, V.J., Zapata-Rivera, D.: Adaptive educational systems. In: Adaptive Technologies for Training and Education, pp. 7–27. Cambridge University Press (2012)
3. Dougiamas, M., Taylor, P.C.: Moodle: using learning communities to create an open source course management system. In: Proceedings of EdMedia, vol. 2003, pp. 171–178 (2003)
4. MoodleDocs: Restrict access settings. https://docs.moodle.org/en/Restrict_access_settings, last accessed 2026/01/24
5. MoodleDocs: Using Activity completion. https://docs.moodle.org/en/Using_Activity_completion, last accessed 2026/01/24
6. Slade, S., Prinsloo, P.: Learning analytics: ethical issues and dilemmas. Am. Behav. Sci. **57**(10), 1510–1529 (2013)
7. Polit, D.F., Beck, C.T.: The content validity index: are you sure you know what's being reported? Critique and recommendations. Res. Nurs. Health. **29**(5), 489–497 (2006)
8. Polit, D.F., Beck, C.T., Owen, S.V.: Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res. Nurs. Health. **30**(4), 459–467 (2007)
9. Lynn, M.R.: Determination and quantification of content validity. Nurs. Res. **35**(6), 382–385 (1986)
10. Zimmerman, B.J.: Becoming a self-regulated learner: an overview. Theory Pract. **41**(2), 64–70 (2002)
11. Pintrich, P.R.: The role of goal orientation in self-regulated learning. In: Boekaerts, M., Pintrich, P.R., Zeidner, M. (eds.) Handbook of Self-Regulation, pp. 451–502. Academic Press (2000)
12. Deci, E.L., Ryan, R.M.: The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol. Inq. **11**(4), 227–268 (2000)
13. Alam, S., Hameed, A., Balogh, Z., Haque, S.A.: Digitalisation of education: challenges and issues in virtual pedagogical competence. J. Educ. Cult. Soc. **16**(1), 165–177 (2025)
14. Diputra, Y., Husin, M., Fadhilah, Karimi, A., Yunusa, A.A., Muthma'innah, M.M.: Factors shaping student readiness for online learning: technology, perception, and lecturer roles. J. Educ. Cult. Soc. **16**(1), 365–387 (2025)
15. CAST: Universal Design for Learning Guidelines, version 2.2. CAST (2018)
16. MoodleDocs: Analytics. https://docs.moodle.org/en/Analytics, last accessed 2026/01/24
17. Siemens, G., Long, P.: Penetrating the fog: analytics in learning and education. Educ. Rev. **46**(5), 30–40 (2011)
18. Gašević, D., Dawson, S., Siemens, G.: Let's not forget: learning analytics are about learning. TechTrends. **59**, 64–71 (2015)
19. Morze, N.V., Varchenko-Trotsenko, L.O., Terletska, T.S.: Comprehensive framework for adaptive learning implementation in Moodle LMS: technical, pedagogical, and administrative perspectives. In: AREdu 2025: 8th International Workshop on Augmented Reality in Education (co-located with ICHTML 2025). CEUR Workshop Proceedings, vol. 4060. CEUR-WS.org (2025)
20. Semerikov, S.O., Nechypurenko, P.P., Vakaliuk, T., Mintii, I.S., Fadieieva, L.: Development and validation of an adaptive learning readiness assessment framework for Moodle courses. Research Square preprint (posted Oct 7, 2025). https://doi.org/10.21203/rs.3.rs-7596476/v1
21. Al-Qora'n, L.F., Nganji, J.T., Alsuhimat, F.M.: designing inclusive and adaptive content in Moodle: A framework and a case study from Jordanian higher education. Multimodal Technol. Inter. **9**(6), 58 (2025). https://doi.org/10.3390/mti9060058
22. Montes, R., Herrera, L., Crisol, E.: Moodle usability assessment methodology using the universal design for learning perspective. arXiv:2403.10484. (2024)

23. Roski, M., Sebastian, R., Ewerth, R., Hoppe, A., Nehring, A.: Learning analytics and the universal Design for Learning (UDL): A clustering approach. Comput. Educ. **214**, 105028 (2024). https://doi.org/10.1016/j.compedu.2024.105028
24. Khampirat, B., Pusaboon, K., Bamrungsin, P., Chonsalasin, D.: Development and validation of a questionnaire to measure social competencies for undergraduate students. J. Educ. Cult. Soc. **16**(1), 535–553 (2025)
25. Kildė, L.: ESL teachers' approaches towards the acceptance of educational technology integration in non-formal education: A case from Kenya. J. Educ. Cult. Soc. **14**(1), 634–649 (2023)
26. Savitska, V., Homotiuk, O., Stefanyshyn, O., Múčka, O., Spuchľák, J., Kenderesova, E.: Innovative educational technologies as the tool of soft skills development in the context of higher education transformation. Clin. Soc. Work Health Interv. **16**(5), 81–98 (2025)
27. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)

Experiences with AI, Attitudes Towards Its Use and Ethical Concerns of Teaching Students: Correlation Analysis

Jana Burgerová and Vladimír Piskura(✉)

Faculty of Education, University of Presov, 17. November 15, 080 01 Presov, Slovakia
{jana.burgerova,vladimir.piskura}@unipo.sk

Abstract. In higher education, we are seeing an avalanche-like increase in the existence and use of artificial intelligence (AI) tools. For this reason, we focused our attention on a study that aimed to examine the relationships between understanding artificial intelligence (AI), level of practical experience, attitudes towards the use of AI in education, and ethical issues, usually concerns of students. The students were teacher training students. The research had a quantitative, non-experimental correlational design and was based on data obtained from an author-designed questionnaire (10 items, Likert scale 1–6; $\alpha = 0.798$). The sample consisted of 1,432 students in teacher training programmes at Slovak universities in the academic years 2023–2025. The relationships between variables were analysed using Spearman's correlation coefficient. The results confirmed that understanding of AI is significantly related to students' experience ($\rho = 0.181$–0.311; $p < 0.001$). At the same time, it was found that a higher level of experience is associated with more positive attitudes towards the implementation of AI in education ($\rho = 0.217$–0.405; $p < 0.001$). Ethical concerns correlated with more negative attitudes towards AI and a greater need for and support of regulatory measures ($\rho = 0.266$–0.398; $p < 0.001$). The findings suggest that understanding, experience, and ethical attitudes form an interconnected system of factors influencing the readiness of future teachers to work with AI. At the same time, they emphasise the need for balanced training combining the practical use of AI with systematic ethical reflection.

Keywords: artificial intelligence · generative AI · teacher training · digital competence · student attitudes · AI ethics · higher education

1 Introduction

Artificial intelligence (AI) has, of course, been around for a long time, but it is only in recent years that it has become one of the significant phenomena influencing the process and transformation of education. Language models (LLMs), which have exploded in popularity since 2022, are changing the ways in which students and teachers work with content, communicate, solve problems and evaluate learning outcomes [1–3]. In the context of training future teachers, this technological change is particularly significant, as

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 85–97, 2027.
https://doi.org/10.1007/978-3-032-32104-6_8

their professional lives will be shaped by an environment where AI will be an integral part of school reality. The demands placed on teachers are therefore changing fundamentally [4] – from basic digital literacy to the ability to use intelligent technologies critically and creatively in planning and implementing teaching [5]. A key competence for future teachers is becoming AI literacy [6], which includes not only a technical understanding of how AI works, but also the ability to assess its pedagogical value, risks and ethical implications [7]. Research shows that students with a higher understanding (cognitive as well as manual) of technologies have significantly more positive attitudes towards their use in school [8] and are also more open to experimenting with innovative teaching methods supported by AI tools [9]. However, understanding is not an isolated factor – it is closely related to experience itself. Direct work with AI increases the feeling of competence, reduces uncertainty and supports the establishment of technology in pedagogical practice [10, 11]. These connections suggest that the relationship between knowledge, experience and acceptance of AI is reciprocal and evolving very dynamically.

Like everything in life, the positive potential of AI also brings threats, especially ethical challenges. These are particularly specific in the school environment, where we constantly deal with issues of fairness, transparency, privacy, and student autonomy. Current research indicates that ethical concerns can significantly influence the willingness of future teachers to use AI or support its integration into teaching processes [12, 13]. A higher level of concern is often associated with the demand for stricter regulatory mechanisms.

Despite growing interest in this issue, there is not much research in the European context that analyses the cognitive (understanding), behavioral (experience) and value (attitudes and concerns) components of teacher students' relationship to AI. Most of the available studies deal with these components partially and therefore do not provide a comprehensive view of the structure of mutual relationships. We believe that it is future teachers who will determine the limits, possibilities and pedagogical strategies for working with AI in schools in the coming decade. The aim of the study is to create a comprehensive empirical picture of how the key factors mentioned above influence each other and how they correlate, and to identify implications for the training of future teachers in this context based on the findings.

2 Research Methodology

2.1 Research Design

The research was conducted as a quantitative, non-experimental, correlational study focusing on identifying relationships between understanding AI as such, practical experience, attitudes towards the use of AI in education, and ethical concerns of teacher training students. The study used questionnaire data collection via an online tool, which made it possible to effectively reach a wide population of respondents in different years and fields of study at the university. Given the research objective – to verify the relationships between several psychological and attitudinal variables – a correlational design was chosen, followed by testing of statistical hypotheses using non-parametric methods.

2.2 Research Objectives and Hypotheses

The aim of the research was to analyse how the following are interrelated: understanding of AI, practical experience with its use, attitudes towards its pedagogical integration, and ethical concerns among a population of teacher training students.

Based on theoretical assumptions, three hypotheses were established:

H1: Greater understanding of AI (i1) is associated with a higher level of experience with it (i4, i5, i6).
H2: Greater experience with AI (i4, i5, i6) is associated with more positive attitudes towards its use in education (i2, i3, i7).
H3: Higher ethical concerns (i8) are associated with stronger rejection of AI and support for its regulation (i9, i10).

These hypotheses allow for a comprehensive examination of the relationships between the cognitive, behavioural, and value aspects of AI literacy.

2.3 Research Sample

Students of teacher training programmes at two Slovak universities (*University of Presov in Presov, University of Pavol Jozef Safarik*) participated in the research. Data from three academic years (2023, 2024 and 2025) were included in the analysis. The sample included full-time and part-time students from all years of bachelor's and master's degree programmes (Table 1).

Table 1. Participants

		Form			
Year	University	Full-time	Part-time	Not specified	SUM
2023	UNIPO	121	53	336	510
	UPJŠ	84	0	0	84
2024	UNIPO	277	220	0	497
2025	UNIPO	207	133	0	340

The research sample consisted of *1,432* students who participated in the questionnaire survey in three consecutive academic years: 2023 ($N = 595$), 2024 ($N = 497$) and 2025 ($N = 340$). The respondents were mainly from the University of Presov in Presov (94.1%), supplemented by a smaller representation of students from Pavol Jozef Safarik University in Kosice (5.9%). Participation was voluntary and anonymous. We are aware of the limitations posed by institutional imbalance, which may restrict the full generalization of the results at the national level. However, given the robust size of the total sample (N = 1,432), the three-year data collection period, and the fact that teacher preparation programs in Slovakia are subject to uniform national standards and regulation, we believe that the correlations found accurately reflect broader patterns of technological adaptation

among future teachers in the Slovak academic environment. This study thus serves as an important empirical baseline for future comparative analyses with institutions in a European or global context, which could further enhance the generalizability of these findings.

2.4 Research Tool

An original questionnaire containing ten items rated on a Likert scale ranging from 1 to 6 (*1 = very low rating, 6 = very high rating*) was used to collect data, divided into four thematic areas:

Understanding AI: i1
Attitudes towards the use of AI in education: i2, i3, i7
Experience with AI: i4, i5, i6
Ethical concerns and regulatory attitudes: i8, i9, i10

The reliability of the tool was verified using Cronbach's alpha ($\alpha = 0.798$), which demonstrated the consistency of the research tool (Fig. 1). The questionnaire and output data are publicly available in the OFS open repository at: http://bit.ly/3LSRtQF

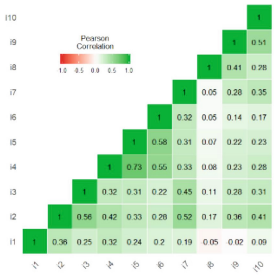


Fig. 1 Correlation Heatmap

2.5 Statistical Data Processing

At the beginning of each analysis, basic descriptive statistics (median, mean, standard deviation, frequency distributions) were calculated, which we then used to describe the research variables. Given the ordinality of the Likert scale and the non-parametric distribution of the research data, non-parametric statistical methods were chosen. For each of the hypotheses, Spearman's coefficient was calculated, which is a measure of monotonic (not necessarily linear) dependence between two variables and does not assume normal data distribution or equal intervals between scale values, which is why we consider it suitable for researching attitudinal, perceptual or subjective variables [14]. When interpreting the results of correlation analysis, it is important to distinguish not only the direction but also the strength of the relationship between variables. Spearman's coefficient (ρ) takes values from -1 to $+1$, with values around *0.10–0.19* indicating a

weak relationship, values *0.20–0.39* indicating a moderate relationship, and values *0.40–0.59* indicating a relatively strong correlation with practical significance. In the context of social and educational sciences, even moderately strong correlations are significant, as students' attitudes and behavior are influenced by many parallel factors and rarely reach extreme correlation values. Data processing was performed in Jamovi, which provides a robust framework for non-parametric correlation analysis. Significance was assessed at the $\alpha = 0.05$ level.

2.6 Results

The individual hypotheses are presented in three subsections, which deal separately with the relationships between understanding and experience (H1), experience and attitudes (H2), and ethical concerns and rejection of AI (H3). The interpretation of the results focuses on statistical significance and practical implications for teacher training.

The results of the correlation analysis H1 (*Higher understanding of AI (i1) is associated with a higher level of experience (i4, i5, i6)*) allowed us to determine whether students with a higher understanding of AI also show greater engagement in the use of AI in various areas.

Table 2. Correlation Matrix H1

		i1		i4		i5	
i1	Spearman's rho	–					
	p-value	–					
i4	Spearman's rho	**0.311**	***	–			
	p-value	< 0.001		–			
i5	Spearman's rho	**0.232**	***	0.720	***	–	
	p-value	< 0.001		<0.001		–	
i6	Spearman's rho	**0.181**	***	0.545	***	0.575	***
	p-value	< 0.001		< 0.001		< 0.001	

Note. * p < 0.05, ** p < 0.01, *** p < 0.001

Spearman's correlations (Table 2) show that there are mild to moderately strong and statistically highly significant positive relationships between AI understanding (*i1*) and individual types of experience (*i4, i5, i6*). Students who declare a higher level of understanding of AI also report that they have more practical experience with using these tools.

The strongest positive relationship was found between understanding of AI (*i1*) and general experience with AI (*i4*), with a correlation *of rho* $= 0.311$, $p < 0.001$, indicating a moderately strong positive association. A similar, albeit slightly weaker, relationship was found between understanding and experience in the school environment (*i5; rho* $= 0.232$, $p < 0.001$) and between understanding and the use of AI in preparing university assignments (*i6; rho* $= 0.181$, $p < 0.001$). These results suggest that students with

a higher declared understanding of AI tend to work with AI more actively and to a greater extent. At the same time, very strong relationships between the experience items themselves were also confirmed: *i4 and i5* (*rho = 0.720, p < 0.001*), i4 and i6 (*rho = 0.545, p < 0.001*), i5 and i6 (*rho = 0.575, p < 0.001*). If a student has experience in one area, they usually have it in others as well – this is a compact pattern of behaviour typical of AI users. The results are visualised in the form of a graphical matrix (Fig. 2) showing the relationships between AI understanding (*i1*) and individual types of experience (*i4, i5, i6*). It shows consistent positive linear trends, representing higher understanding associated with greater experience. The most significant connection is between *i1* and *i4*, while the relationships with *i5* and *i6* are milder but still clearly positive. The strong interconnections between *i4-i5-i6* suggest that students' experiences with AI accumulate across different contexts.

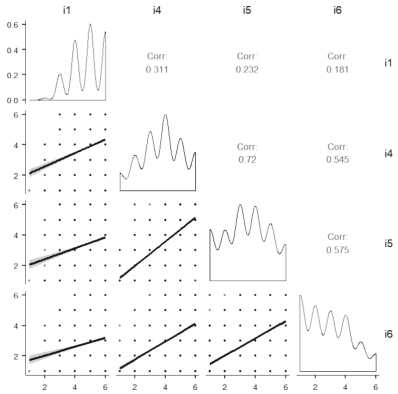


Fig. 2 Visualised Correlation Matrix H1

Overall, the results robustly support hypothesis H1, i.e., understanding of AI is closely related to active use of AI in various contexts, with experience also carrying overall consistency across multiple areas of AI use in education.

To test hypothesis H2 (*Greater experience with AI (i4, i5, i6) is associated with more positive attitudes towards its use in education (i2, i3, i7)),* we analysed the relationship between students' level of experience with AI and their attitudes towards AI in education. Since this is ordinal data, we again used Spearman's correlation, which allows us to assess the strength and direction of monotonic relationships between variables.

Table 3. Correlation Matrix H2

		i2		i3		i7		i4		i5	
i2	Spearman's rho	–									
	p-value	–									

(*continued*)

Table 3. (*continued*)

		i2		i3		i7		i4		i5	
i2	Spearman's rho	–									
i3	Spearman's rho	0.552	***	–							
	p-value	<0.001		–							
i7	Spearman's rho	0.516	***	0.443	***	–					
	p-value	<0.001		<0.001		–					
i4	Spearman's rho	**0.405**	***	**0.303**	***	**0.324**	***	–			
	p-value	<0.001		<0.001		<0.001		–			
i5	Spearman's rho	**0.324**	***	**0.301**	***	**0.299**	***	0.720	***	–	
	p-value	<0.001		<0.001		<0.001		<0.001		–	
i6	Spearman's rho	**0.263**	***	**0.217**	***	**0.310**	***	0.545	***	0.575	***
	p-value	<0.001		<0.001		<0.001		<0.001		<0.001	–

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Results of testing hypothesis H2 (Table 3) – there is a clear and statistically significant correlation between the level of students' experience in using AI (*i4, i5, i6*) and more positive attitudes towards its use in education (*i2, i3, i7*). Students who report higher scores in the area of experience rate the contribution of AI to education and its potential for primary education significantly more positively. The strongest relationship was found between general experience (*i4*) and the assessment of the contribution of AI to education (*i2*), where the correlation was $rho = 0.405, p < 0.001$, similarly to $i4 \times i3$ ($rho = 0.303, p < 0.001$) and $i4 \times i7$ ($rho = 0.324, p < 0.001$).

Equally strong relationships were also found for item *i5*, which captures the use of AI in school practice. Students with stronger practical experience evaluate AI more positively in the context of education, as confirmed by the correlations $i5 \times i2$ ($rho = 0.324, p < 0.001$), $i5 \times i3$ ($rho = 0.301, p < 0.001$) and $i5 \times i7$ ($rho = 0.299, p < 0.001$). The third area of experience, represented by item *i6* (use of AI in the creation of university theses), also shows consistent, moderately strong positive relationships with attitude variables ($i6 \times i2$: $rho = 0.263, p < 0.001$; $i6 \times i3$: $rho = 0.217, p < 0.001$; $i6 \times i7$: $rho = 0.310, p < 0.001$).

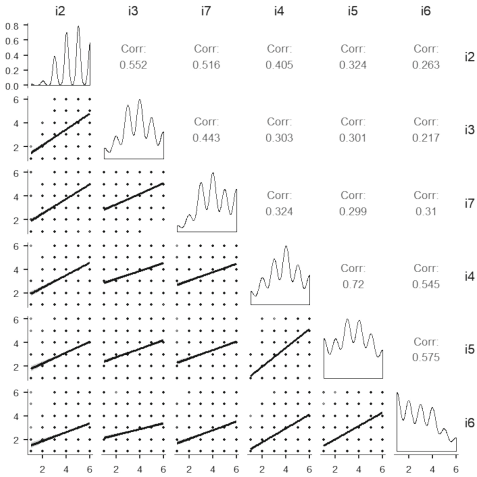


Fig. 3 Visualised Correlation Matrix H2

The correlation matrix (Fig. 3) also visually confirms the consistent positive relationships between experience with AI and attitudes towards its use in education. The positive monotonic trend in all pairs of variables is documented by upward-sloping lines, confirming that students with more experience also have more favourable attitudes. The most pronounced upward lines can be observed in the relationships *i4* × *i2* and *i4* × *i7*, including numerically stronger correlations.

We can conclude that the more experience students have, the more positively they evaluate the use of AI in education, as well as the need to prepare teachers to work with these tools. The correlations are consistent across all experience items and significant at the $p < 0.001$ level, confirming a stable relationship between the practical use of AI and the acceptance of its pedagogical potential.

To test hypothesis H3 (*Higher ethical concerns (i8) are associated with stronger rejection of AI and support for its regulation (i9, i10)*), we analysed the relationship between ethical concerns about the use of AI (*i8*) and the degree of rejection and demand for regulation (*i9, i10*). Since the variables are ordinal in nature, we used Spearman's correlation in this case as well, which allows us to assess the direction and strength of the relationships between these attitudes.

Table 4. Correlation Matrix H3

		i8		i9		i10	
i8	Spearman's rho	–					
	p-value	–					
i9	Spearman's rho	**0.398**	***	–			
	p-value	< 0.001		–			
i10	Spearman's rho	**0.266**	***	0.506	***	–	

(*continued*)

Table 4. (*continued*)

		i8		i9		i10	
i8	Spearman's rho	–					
	p-value	< 0.001		< 0.001		–	

Note. * p < 0.05, ** p < 0.01, *** p < 0.001

The results (Table 4) clearly confirm that higher ethical concerns about AI are associated with stronger rejection of its use and more pronounced agreement with statements of a regulatory nature. Ethical concerns (*i8*) correlated significantly with the belief that AI should not interfere with education (*i9*), with a correlation coefficient of *rho = 0.398, p < 0.001*, representing a moderately strong positive relationship. Similarly, although somewhat weaker, the results documenting the relationship between ethical concerns and the belief that AI development should be halted (*i10*) showed a correlation *of rho = 0.266, p < 0.001*.

A strong relationship also emerged between the attitudes of rejection themselves – *i9* and *i10* (*rho = 0.506, p < 0.001*), suggesting that students with regulatory sceptical attitudes are also more likely to support more radical restrictions on AI development.

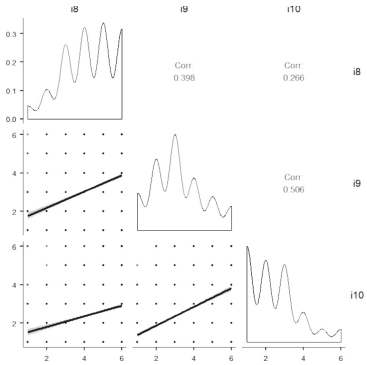


Fig. 4. Visualised correlation matrix H3

The correlation matrix (Fig. 4) demonstrates consistent positive relationships between variables *i8*, *i9* and *i10*, with all trends being upward. The most pronounced of these is between *i9* and *i10*, where a clear linear upward pattern can be seen (***rho = 0.506***). The relationships between *i8* and both negative attitudes (*i9, i10*) are milder but also clearly positive. The matrix thus supports the numerical results and confirms that students' concerns and negative attitudes are cumulative. The overall pattern of results clearly confirms hypothesis H3: *the greater the ethical concerns a student declares, the more likely they are to hold negative or restrictive attitudes towards AI and its use in education.*

3 Discussion

The results of the study provide a comprehensive view of the connection between understanding artificial intelligence, practical experience, attitudes towards its use, and ethical concerns of future teachers. With robust empirical support, all three hypotheses were confirmed, suggesting that the cognitive, behavioural, and value components of AI Literacy form a coherent system of interdependent factors. The findings thus expand on existing knowledge about teacher training in AI, which has often examined individual dimensions in isolation [2, 9].

The results of the study show that understanding of AI naturally develops with experience, with the highest correlation detected between general experience and understanding (*rho = 0.311*). Previous empirical findings emphasise that interaction with and knowledge of technology increases confidence, understanding of how it works, and the ability to assess its pedagogical value [15, 16]. Strong correlations between experience items (*i4, i5, i6*) mean that the use of AI in different contexts is cumulative – in other words, students who use AI to solve study problems are likely to use it in their everyday lives and subsequently in their teaching practice. Correlation analysis further supports practical experience as a key predictor of positive attitudes towards the use of AI in education. The strongest correlation was between general experience and general belief in the benefits of AI for education (*rho = 0.405*), which correlates with findings that acceptance of technology is conditioned by a sense of competence and personal experience [17, 18].These findings also correlate closely with current global studies, thereby placing our European perspective within a broader international context. Similar trends on a transnational scale are reported by [19], who emphasizes that the acceptance of AI technologies is universally conditioned by the level of the user's prior positive experience. The significance of this relationship from a global perspective is also underscored by a recent study by [20] published in *Humanities and Social Sciences Communications*, which demonstrates, using the TAM (Technology Acceptance Model) on an international sample, that practical experience and trust in the technology are key factors influencing students' attitudes toward its use. The alignment of our local findings with these global trends suggests that Slovak teacher education students exhibit very similar patterns of behavior in adapting to AI as their peers in other regions of the world. Significant correlations for items *i5* and *i6* also indicate that experience from real teaching situations and the academic environment are equally important. These results demonstrate the need for the systematic integration of AI into educational activities within teacher training. From the perspective of the practical training of future teachers, the correlations identified (rho = 0.217 to 0.405) are of fundamental importance for the design of study programs. The fact that positive attitudes toward AI are moderately influenced by prior experience suggests that theoretical reflection on the phenomenon of AI within lectures alone is insufficient for effective technological adaptation. For faculties training teachers, this means the necessity of transitioning from a descriptive approach to active "hands-on" learning, where students integrate AI tools directly into the creation of lesson plans and teaching materials. Emphasizing the practical component in the curriculum serves as a key catalyst for digital transformation, as it is precisely the sense of pedagogical competence when working with AI that demonstrably lowers barriers and increases the willingness of future teachers to utilize these innovations in their future school practice. Based on

the results, we further conclude that ethical concerns correlate significantly with negative attitudes and support for regulation, with the highest correlation between negative attitudes *i9* and *i10* (*rho = 0.506)*. This finding is consistent with studies [19, 21], which demonstrate that acceptance of AI is closely related to the level of experience and trust, with practical interaction with technologies significantly strengthening the willingness of AI users to accept and use it. A positive, albeit weaker, relationship between ethical concerns and support for a complete halt to AI development suggests that some students perceive technological risks as significant, which has important implications for the design of courses focused on ethics and the responsible use of AI.

In a broader context, our findings indicate that future teachers approach AI pragmatically, their attitudes determined by two opposing but concurrent factors: growing experience and deepening ethical concerns. Experience promotes acceptance and visibly shifts attitudes towards the use of AI, while ethical concerns increase caution and promote regulatory attitudes. This dual model confirms that preparing teachers to work with AI is not enough on a technical level alone – it is equally important to cultivate their ability to critically understand and reflect on the ethical implications. At the same time, it is important to emphasise that although the correlations between experience and attitudes are higher than the correlations between ethical concerns and rejection, both types of relationships are sufficiently strong for the purposes of behavioural science. Future teachers are able to appreciate the benefits of AI while being sensitive to its risks – which is in line with modern concepts of "AI literacy" and "responsible AI," which emphasise the balance between innovation and ethical responsibility [22, 23].

Based on our findings, we propose specific strategies for integrating AI literacy and ethical discussions into the teacher education curriculum: (1) Hands-on integration: Instead of isolated theoretical courses, we recommend implementing AI tools directly into the pedagogy of individual subjects. An example is the use of generative AI to personalize instructional texts for students with special educational needs or the automated creation of formative tests directly during seminars. (2) Ethics Labs: Since ethical concerns correlate significantly with regulatory attitudes, discussions on ethics should be based on addressing real-world dilemmas. We recommend the case study method, such as analyzing AI hallucinations in student work or critically assessing copyright issues in AI-generated materials. (3) The 'AI-as-Partner' Concept: Students should be guided to use AI as a collaborative partner in brainstorming and designing lessons. According to our data, this approach directly strengthens their sense of pedagogical competence and reduces the technological anxiety associated with an unfamiliar tool.

4 Conclusion

The research findings show that understanding of AI, practical experience, attitudes towards the use of AI, and ethical concerns form an interconnected system among teacher training students that significantly influences their readiness to work with modern technologies. It has been confirmed that understanding of AI increases with the level of experience, which emphasises the need to provide students with authentic opportunities to work with AI tools. Similarly, experience with AI has been shown to be a key factor in positive attitudes towards its use in the teaching process, confirming the importance

of practice-oriented teacher training. The findings also point out that ethical concerns are a stable and significant predictor of regulatory and negative attitudes, highlighting the need for the systematic integration of ethical and critical discussions about AI into the curriculum.

Overall, the results suggest that the development of AI literacy must be balanced: it is not enough to build only technical competencies, but also value-based and ethical thinking. For the university training of future teachers, this means the need to create educational environments in which students gain experience, understanding and the ability to critically reflect on the consequences of using AI in education and will be able to use AI in their future practice in the context of improving the educational environment in its broadest sense. We believe that the study provides an important framework for the creation of modern teacher training programmes and, at the same time, opens up space for further research focused on the long-term development of attitudes towards AI during study and practice. Specifically, we propose conducting longitudinal studies that would track the stability of future teachers' attitudes from the beginning of their studies to their entry into the profession, which would allow us to capture causal connections in the process of their technological adaptation. An equally important direction for future research is to examine the direct impact of teachers' AI literacy on measurable teaching outcomes and the cognitive progress of their students, thereby shifting the focus from the analysis of subjective attitudes to the objective measurement of real pedagogical impact in school practice.

Acknowledgement. We, the authors, declare that we have used the AI Tool ChatGPT in the development of this paper for the following purpose: proofreading of the selected parts of the text.

This paper was created as part of the VEGA 1/0163/25 project "Systemic solution for the level of digital competences in the university environment in accordance with EU standards", which is being carried out at the Faculty of Education of the University of Presov in Presov.

References

1. Zawacki-Richter, O., Marín, V.I., Bond, M., Gouverneur, F.: Systematic review of research on artificial intelligence applications in higher education – where are the educators? Int. J. Educ. Technol. High. Educ. **16**, 39 (2019). https://doi.org/10.1186/s41239-019-0171-0
2. Holmes, W., Bialik, M., Fadel, C.: Artificial Intelligence in Education. Promise and Implications for Teaching and Learning (2019)
3. Shah, P.: AI and the Future of Education: Teaching in the Age of Artificial Intelligence. John Wiley & Sons (2023)
4. Yu, H.: The application and challenges of ChatGPT in educational transformation: new demands for teachers' roles. Heliyon. **10** (2024)
5. Ng, D.T.K., Leung, J.K.L., Chu, S.K.W., Qiao, M.S.: Conceptualizing AI literacy: an exploratory review. Comput. Educ. Artif. Intell. **2**, 100041 (2021). https://doi.org/10.1016/j.caeai.2021.100041
6. Yi, Y.: Establishing the concept of AI literacy. Jahr–Europ. J. Bioeth. **12**, 353–368 (2021)
7. Long, D., Magerko, B.: What is AI literacy? Competencies and design considerations. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–16. Association for Computing Machinery, New York, NY, USA (2020)

8. Ardies, J., De Maeyer, S., Gijbels, D., van Keulen, H.: Students attitudes towards technology. Int. J. Technol. Des. Educ. **25**, 43–65 (2015)
9. Siau, K., Wang, W.: Building trust in artificial intelligence, machine learning, and robotics. Cut. Bus. Technol. J. **31**, 47 (2018)
10. Yilmaz, R., Karaoglan Yilmaz, F.G.: The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. Comput. Educ. Artif. Intell. **4**, 100147 (2023). https://doi.org/10.1016/j.caeai.2023.100147
11. García-Peñalvo, F.J.: Generative artificial intelligence. open challenges, opportunities, and risks in higher education. In: Proceedings 14th International Conference on eLearning, pp. 4–15 (2023)
12. Akgun, S., Greenhow, C.: Artificial intelligence in education: addressing ethical challenges in K-12 settings. AI Ethics. **2**, 431–440 (2022). https://doi.org/10.1007/s43681-021-00096-7
13. Delello, J.A., Sung, W., Mokhtari, K., Hebert, J., Bronson, A., De Giuseppe, T.: AI in the classroom: insights from educators on usage, challenges, and mental health. Educ. Sci. **15** (2025). https://doi.org/10.3390/educsci15020113
14. Hauke, J., Kossowski, T.: Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest. Geogr. **30**, 87–93 (2011). https://doi.org/10.2478/v10117-011-0021-1
15. Wu, W.V., Yen, L.L., Marek, M.: Using online EFL interaction to increase confidence, motivation, and ability. J. Educ. Technol. Soc. **14**, 118–129 (2011)
16. Estradas, A., Ribeiro, I., Raposo, A., Durão, A.: Technology as a tool to enhance motivation and learning. (2019)
17. Durodolu, O.: Technology acceptance model as a predictor of using information system'to acquire information literacy skills. Libr. Philos. Pract. (2016)
18. Roca, J.C., Chiu, C.-M., Martínez, F.J.: Understanding e-learning continuance intention: an extension of the technology acceptance model. Int. J. Hum. Comput. Stud. **64**, 683–696 (2006). https://doi.org/10.1016/j.ijhcs.2006.01.003
19. Gerlich, M.: Perceptions and acceptance of artificial intelligence: a multi-dimensional study. Soc. Sci. **12**, 502 (2023)
20. Khoso, A.K., Honggang, W., Darazi, M.A.: Trust and attitude towards AI as pathways to creativity: a TAM model study of EFL students' digital literacy and AI acceptance. Humanit. Soc. Sci. Commun. **13**, 69 (2025)
21. Oprea, S., Nica, I., Bâra, A., Georgescu, I.: Are skepticism and moderation dominating attitudes toward AI-based technologies? Am. J. Econ. Sociol. **83**, 567–607 (2024)
22. Abbas, T.: Ethical implications of AI in modern education: balancing innovation and responsibility. Soc. Sci. Spectr. **2**, 51–57 (2023)
23. Megdad, M.M.M., et al.: Ethics in AI: balancing innovation and responsibility. (2024)

STEM in Education: A Comparative Analysis of the Influence of Physical Hardware and Simulation on the Development of Algorithmic Thinking

Petr Coufal(✉) and Eva Milková

Faculty of Education, University of Hradec Kralove, Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic
{petr.coufal,eva.milkova}@uhk.cz

Abstract. This article examines the role of the STEM concept and programming within the context of secondary vocational education for students in economics-based programs. The study aims to compare the effectiveness of using physical hardware (the micro:bit microcontroller) versus a virtual development environment (the Microsoft MakeCode simulator). The research employs a mixed-methods design, combining a quantitative questionnaire survey of 56 respondents with a qualitative analysis of semi-structured interviews. The results reveal the existence of a so-called "haptic paradox." Physical hardware demonstrates significantly higher motivational potential and strengthens the perceived meaningfulness of the work by providing students with tangible evidence of their digital competence. Conversely, the virtual simulator proved to be a more didactically effective tool for developing procedural knowledge and understanding abstract algorithmic structures, primarily due to the elimination of technical barriers and a lower cognitive load during debugging. Based on the findings, the study proposes the implementation of a hybrid instructional model. This model consists of the initial use of a simulator as a cognitive "scaffolding" tool, followed by work with physical devices to anchor attitudes toward technology and bridge the gap between code and real-world practice. This paper provides insight into programming instructional methodologies for a specific target group of students for whom computer science is not a core major subject.

Keywords: STEM education · physical computing · micro:bit · algorithmic thinking · Microsoft MakeCode

1 Introduction

The STEM (Science, Technology, Engineering, Mathematics) concept represents an integrated approach to education that emphasizes applied science and real-world problem-solving [1]. Science, Technology, Engineering and Mathematics education fields encompass all the subjects which teach students to use logical reasoning in the pursuit of scientific truths and their application to our reality. From biology to astronomy, or robotics,

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 98–109, 2027.
https://doi.org/10.1007/978-3-032-32104-6_9

these disciplines make use of the scientific process of observing the natural world, setting up hypothesis, collecting data, and reporting them as objectively as possible. At the same time, especially in the last two centuries, they focus on using and further developing technologies and engineering processes to advance scientific research and apply its findings into economic flow [2].

1.1 Role of Programming in Vocational Training

Within the context of secondary vocational education, programming is perceived not merely as an isolated technical skill, but as a key tool for understanding the digital transformation of society. According to Osborne [3], students' attitudes toward technology are shaped by their personal experience with tools that allow them to see the direct impact of their work. For students in economic fields, where computer science is not a core subject, programming becomes a means of developing so-called "transversal skills," which are essential for their future adaptability in the labor market during the era of Industry 4.0 [4].

1.2 Development of Algorithmic and Computational Thinking

Algorithmic thinking forms the core of Computational Thinking, a cognitive process involving the formulation of problems and their solutions in the form of steps that can be executed by an information-processing agent [5]. In this regard, instruction at secondary schools draws on Piaget's theory of cognitive development, specifically the stage of formal operations, where students are capable of working with abstraction and logical hypotheses [6]. The development of these competencies is particularly effective when utilizing the "scaffolding" methodology, where students are guided from visual blocks to understanding the deeper structures of code [7]. The objective is not merely knowledge of syntax, but the ability to decompose problems, recognize patterns, and create efficient algorithms.

Teaching programming basics is a challenging task for both students and teachers. Every year, many students fail introductory programming courses at universities, despite their effort and an active aid from lecturers [8].

2 Materials and Methods

2.1 Micro:bit and the MakeCode Environment

The micro:bit is a comprehensive microcomputer whose hardware architecture is intentionally designed to be readable and understandable for students. The dominant feature of the front side is a 5×5 LED matrix, which serves as a basic display unit for text, icons, and simple animations, thereby providing immediate visual feedback. The hardware layer further includes programmable buttons, a touch sensor in the logo, and a wide range of integrated sensors (accelerometer, compass, light sensor, microphone). As stated by Sentance [9], this concentration of sensors on a single board allows students to transition from abstract code to physical interaction with the environment—a key element of so-called physical computing. The micro:bit is shown in Fig. 1.

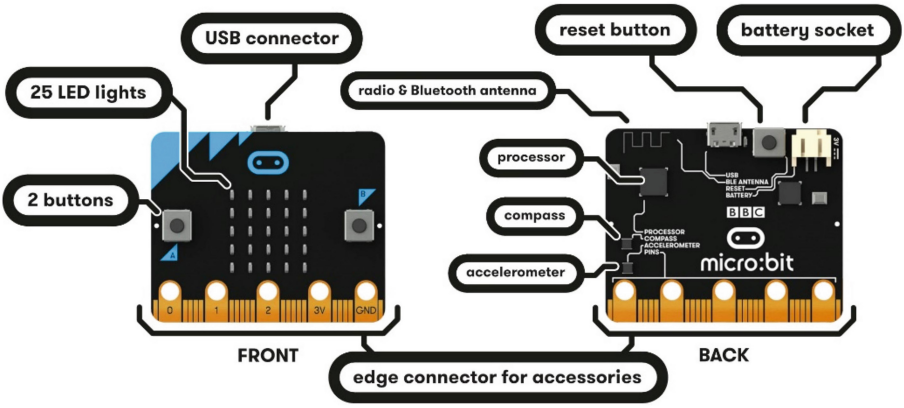


Fig. 1. Micro:bit: description of the microcomputer and its elements [10].

The Microsoft MakeCode development environment complements the hardware with an intuitive software layer based on visual programming. The interface is divided into three functional blocks that form a logical unit for the learning process:

- Interactive Simulator: Located on the left side of the screen, it provides a faithful digital copy of the micro:bit that interprets the created code in real-time. Students can simulate inputs (e.g., button presses or temperature changes) and immediately see the reaction, which minimizes frustration from malfunctioning hardware and accelerates the debugging cycle [11].
- Command Palette (Toolbox): The middle column categorizes color-coded blocks (Logic, Loops, Variables), facilitating students' orientation within the syntax through visual metaphors.
- Workspace: In the right section, students assemble the program like a puzzle. This "drag-and-drop" method eliminates syntactic errors (missing semicolons, brackets), allowing students in economic fields to focus primarily on the algorithm rather than the grammar of the programming language (Fig. 2).

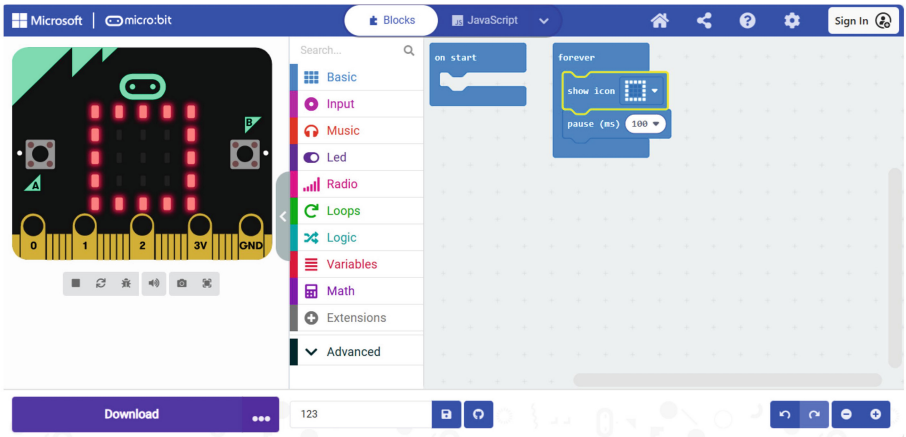


Fig. 2. Makecode: sample development environment [12].

3 Research Methodology

The present research employs a mixed-methods research design, combining quantitative and qualitative approaches. This synergy allows not only for measuring the frequency of specific attitudes but also for gaining a deeper understanding of students' motivations and subjective experiences within the context of STEM education [13]. The primary objective of the investigation is the triangulation of data to obtain a comprehensive picture of the effectiveness of simulation versus physical computing.

Quantitative Phase: Questionnaire Survey.
The primary data collection instrument is a non-standardized questionnaire focused on the affective and cognitive components of students' attitudes. A five-point Likert scale (1 = strongly agree, 5 = strongly disagree) is utilized to measure the intensity of these attitudes. Monitored Dimensions: Motivation, perceived difficulty, technical reliability, and the degree of connection to real-world practice. Analysis: Data are processed using descriptive statistical methods to compare attitudes toward physical hardware versus the online simulator.

Qualitative Phase: Semi-structured Interviews.
To gain deeper insight into the phenomenon, semi-structured interviews were conducted. This method allows for flexible responses to participants' answers and the identification of specific barriers or motivational factors that a quantitative scale might not capture [14]. Respondent Selection: Purposive sampling based on the results of the questionnaire. Analysis: The interviews are transcribed and subjected to thematic analysis [15], aimed at identifying key narratives in students' perceptions of the instruction.

3.1 Research Ethics

The research was conducted anonymously and in accordance with the ethical standards of pedagogical research.

3.2 Research Sample

The research investigation was carried out at a secondary vocational school with an economic focus. The selected sample consisted of students in a field where computer science and programming do not represent the primary core discipline. This context is crucial for an objective assessment of the intuitiveness and motivational potential of the examined tools (simulator vs. hardware). Size and Composition: The group consisted of two classes with a total of 56 students. Age and Level of Study: Respondents were 1st and 2nd-year students aged 15–16 years. This age group is in the stage of developing abstract thinking, which is essential for algorithmization [6]. Initial Competencies: A high degree of heterogeneity was identified regarding students' prior experience with block-based programming (e.g., Scratch). Although the majority of students declared a basic awareness from primary school, their practical skills were assessed as elementary and inconsistent.

3.3 Research Procedure

The investigation was conducted in several logically sequential phases during the first semester of the school year. The entire process was designed to minimize external variables and ensure maximum comparability of data between the group utilizing physical hardware and the group working with the simulator.

Preparatory Phase.
Prior to the commencement of programming instruction, a set of 10 graded tasks was developed, and identical methodological materials and worksheets were prepared for both groups of students. For group A, a basic Micro:bit kit—including a USB cable and a battery power source—was prepared.

Implementation Phase (Intervention).
The implementation phase, representing the core intervention of the study, took place in both groups concurrently within the regular informatics curriculum. To ensure maximum control over the process, both classes were taught by the same instructor using identical methodology based on the ten graded tasks. Students worked individually at computer stations. In group A, interaction involved physical hardware (connecting the board, flashing .hex files), while group B's work was limited to the virtual environment of the web simulator. Progress through the tasks was linear, ranging from simple visualizations on the LED matrix to logical structures and the complex utilization of integrated sensors.

Quantitative Research Phase (Surveying).
Upon completion of the instructional block, students were presented with an online digital questionnaire created via the Google Forms platform. This platform was selected for its high user accessibility, intuitive interface for respondents, and efficient data export capabilities for subsequent statistical analysis. The questionnaire was administered immediately after the conclusion of the teaching cycle to ensure the freshness of students' reflections. A key element of the investigation was total anonymity, which was declared to the students at the beginning of the form. Respondents did not provide names

or other identifying information, aiming to eliminate social desirability bias and ensure the authenticity of their attitudes toward the instruction. The questionnaire consisted of items using a five-point Likert scale. The digital collection format also allowed for automated validity checks of responses and minimized errors arising from manual data transcription.

Qualitative Research Phase (Interviews).
One week after the questionnaire survey and following the evaluation of its primary trends, the qualitative data collection phase commenced in the form of semi-structured interviews. For this purpose, purposive sampling was applied: three students were selected from each class to represent the distinct spectrum of attitudes identified in the questionnaire (positive, neutral, and critical evaluations). The interviews were conducted individually in a quiet school environment to ensure an atmosphere of trust, with a time allocation of 10 to 15 min per respondent. With the participants' consent, all statements were audio-recorded, subsequently underwent verbatim transcription, and were subjected to thematic analysis to gain a deeper understanding of students' subjective experiences with both types of programming environments.

4 Research Results

4.1 Analysis of Individual Questionnaire Items

This chapter presents a comparative analysis of students' perceptions of programming instruction conducted in two different learning environments: a physical micro:bit device and a virtual simulator. The evaluation focuses on motivational aspects, perceived difficulty, confidence, debugging processes, and the perceived meaningfulness of learning outcomes. By examining differences in students' responses across twelve questionnaire items (Q1–Q12), the analysis aims to identify the respective strengths and limitations of hardware-based and simulator-based instruction in STEM education.

The results indicate that both groups evaluated the instruction positively; however, the group working with the physical micro:bit demonstrated a higher level of enthusiasm (Q1). Direct interaction with the device increased the attractiveness of the lessons and confirms the role of hardware as a significant motivational factor.

In terms of understanding the program (Q2), the simulator group achieved more favorable mean values. The virtual environment eliminates technical variables and enables students to focus more effectively on algorithmic logic and the transparency of code processes. A similar pattern emerged regarding confidence while solving tasks (Q6), where the simulator appeared to provide a less stressful and more stable learning environment.

Conversely, one of the strongest positive responses within the hardware group concerned observing the results of coding (Q3). The opportunity to see the output on a physical display held in one's hand was perceived as substantially more attractive than viewing the same output on a computer monitor. The assessment of programming difficulty (Q4) was relatively balanced across both groups, suggesting that the progression of tasks was appropriately aligned with students' abilities regardless of the development environment.

The willingness to continue working in this format (Q5) was higher among students using the real device, supporting the assumption that hardware possesses stronger long-term motivational potential. The most pronounced divergence emerged in relation to debugging (Q7). While error correction in the simulator is intuitive due to immediate feedback, hardware-based programming introduces additional uncertainty regarding whether issues stem from the code, connection, or flashing process.

A substantial difference was also observed in the perception of meaningfulness (Q8). Students working with the physical device perceived their creations as "real technology," whereas the simulator was more frequently viewed as a school-based tool. The preference for working with a tangible device was further confirmed in general attitude items (Q9, Q11), where students across the sample expressed a stronger interest in physical interaction.

An interesting paradox emerged in the evaluation of learning effectiveness (Q10, Q12). Students in the simulator group expressed greater confidence that programming skills can be acquired equally well in a virtual environment, whereas the hardware group was more skeptical. These findings suggest that while the simulator represents an efficient tool for conceptual understanding, the physical device contributes more substantially to motivational, emotional, and value-oriented dimensions of learning (Table 1).

Table 1.

Dimension	Group A		Group B		Difference in Means
	Mean	Std. Dev.	Mean	Std. Dev.	
I. Motivation and Engagement	1.43	0.59	1.99	0.62	−0.56
II. Difficulty and Self-confidence	2.46	0.88	1.95	0.62	0.51
III. Cognitive Understanding	2.45	0.75	1.81	0.62	0.64
IV. Haptic and Functional Preference	1.25	0.46	1.96	0.57	−0.71

1. *Dimension of Motivation and Engagement (Q1, Q3, Q5).*
   This dimension confirms that working with physical hardware generates a stronger affective response. The mean value of 1.43 for Group A indicates a very high level of consensus regarding the attractiveness of the instruction. The interviews complement this result: students perceive the hardware as a "reward" and find working with it less monotonous than standard monitor-based instruction.
2. *Dimension of Perceived Difficulty and Self-confidence (Q4, Q6, Q7).*
   In this case, the statistics speak clearly in favor of the simulator (1.95). The higher standard deviation in the hardware group (0.88) suggests that physical computing creates inequalities among students—while technically proficient students remain

confident, those less skilled experience significantly higher stress from hardware connectivity errors, which reduces their overall sense of competence.

3. *Dimension of Cognitive Understanding (Q2, Q10, Q12).*

   The simulator is perceived by students as a more effective tool for understanding the "internal workings" of a program (1.81). In the interviews, students frequently mentioned that the simulator is "clearer" and allows them to focus on pure algorithmization without the need to manage code uploading. Thus, for teaching abstract concepts (variables, loops), the simulator appears to be a didactically stronger medium.

4. *Dimension of Haptic and Functional Preference (Q8, Q9, Q11).*

   This dimension exhibited the greatest polarization of attitudes. Group A shows a mean of 1.25, representing nearly absolute consensus on the importance of a tangible result. Students from both groups consistently stated in interviews that "controlling a thing" is a qualitatively different experience than "controlling an image on a screen." This dimension confirms that STEM education at vocational economic schools should not abandon hardware, as it is the primary carrier of the sense of "technological reality."

Triangulation of Quantitative Data and Qualitative Interview Statements.
The triangulation of quantitative data with qualitative interview responses allows for a deeper interpretation of the identified statistical differences. Below are the analyses of student experiences for each dimension, accompanied by citations that illustrate their subjective perceptions of the instruction.

Dimension of Motivation and Engagement.
The statistically high level of engagement in the hardware group is explained in the interviews by the element of surprise and the tangible results. Students appreciated that the instruction "left the monitor," which represented a welcome change from standard informatics lessons. "The simulator is fine, but when those LEDs actually lit up in my hand, it was a completely different feeling. I was genuinely happy that it wasn't just a picture, but a thing I could actually show someone," described one respondent from Group A. Conversely, students from Group B admitted that their interest waned after the initial tasks: "I enjoyed it, but after a while, it felt like any other clicking on the web."

Dimension of Perceived Difficulty and Self-confidence.
The interviews confirmed that the higher level of uncertainty in the hardware group (0.88) stems from the technical "unpredictability" of the physical world. While the simulator is perceived as flawless, the hardware introduces stress factors into the process. "The most frustrating part was when my code was correct, but the micro:bit did nothing. I constantly had to check if the cable was plugged in right or if it had flashed correctly. In the simulator, I know immediately that if it doesn't work, I made a mistake in the blocks," explained one student, illustrating why they felt more confident using the simulator. This statement demonstrates that the simulator functions as a "protective layer" that relieves students of technical stress.

Dimension of Cognitive Understanding.
In terms of understanding principles, the simulator proved to be a didactically more transparent tool. In the interviews, students highlighted the visualization capabilities of the MakeCode environment, which helped them track the logical flow of the program. "On the computer, I can see how the yellow frame jumps from one block to another, so I know exactly where the program is at that moment. With the board, it's a black box—either it works or it doesn't, but I can't see inside to what's happening," stated a respondent from Group B. This reflection confirms that the simulator is a more effective tool for the deconstruction and understanding of abstract algorithmic structures.

Dimension of Haptic and Functional Preference.
The strongest consensus in the interviews concerned the significance of the project's "reality." For students of non-technical majors, a tangible output is proof of their competence in the digital world. "When we made the thermometer, in the simulator you just move a bar with the mouse; that's boring. But with the micro:bit, I could breathe on it or warm it up and watch the temperature actually rise. That felt like real science, not just a computer exercise," concluded one of the students. This haptic experience is identified in the interviews as the pivotal moment when the student "connects the code with reality."

The interviews reveal that the simulator is perceived by students as a tool for learning (clarity, debugging, confidence), while the physical micro:bit is perceived as a tool for creation (motivation, pride in results, connection to practice). This conclusion is fundamental for the discussion of the results.

5 Discussion

The research results point to the existence of a so-called "haptic paradox" in programming instruction for secondary vocational students. While the physical hardware (BBC micro:bit) elicited a significantly higher level of affective engagement and a sense of "task reality" (1.43 in the motivation dimension), it simultaneously introduced technical barriers into the learning process that led to uncertainty among some students. This finding corresponds with Seymour Papert's constructionist theory, which emphasizes that learning is most effective when students create a tangible artifact [16]. The interviews confirmed that for students in economic fields who do not consider themselves "technical types," the physical existence of a product is a key validator of their success, supporting Osborne's [3] findings regarding the formation of positive attitudes toward science and technology.

On the other hand, data in the dimensions of cognitive understanding and procedural certainty clearly favor the web-based simulator (1.81). This difference, supported by student citations regarding the "clarity of the yellow frame" and a "safe environment," suggests that simulation effectively reduces cognitive load (Cognitive Load Theory). As Sweller [17] notes, the elimination of extraneous elements—in our case, problematic connectivity or flashing .hex files—allows students to fully allocate their mental capacity to perceiving algorithmic structures. Thus, the simulator functions as an ideal "scaffolding" tool in the early stages of learning, providing immediate feedback without the frustration of hardware technical failures.

A comparison of both approaches shows that for the target group of vocational students, it is not a question of which tool is superior, but how to optimally combine them. The absence of haptic experience in a purely virtual environment leads to the devaluation of the task into a "school game," while exclusive work with hardware can provoke resistance in less proficient students due to the technical difficulty of debugging. In accordance with Wing's [5] research on the development of computational thinking, a hybrid instructional model is recommended: utilizing the simulator as a cognitive trainer for understanding logic and the physical micro:bit as a motivational climax that connects abstract code with the physical reality of sensorimotor perceptions [18, 19].

5.1 Recommendations for Pedagogical Practice

Based on the conducted analysis, we recommend that secondary vocational schools implement a hybrid instructional model based on a gradual transition from simulation to reality:

Initial Phase: Focus on cognitive anchoring and understanding abstract algorithmic structures (e.g., variables, loops) exclusively using the online simulator. This provides students with safe "scaffolding," eliminates technical hardware stress, and allows for efficient debugging.
Application Phase: Once students have mastered the program logic, instruction should seamlessly transition to working with the physical micro:bit (approx. 70% of the time allocation). This phase is crucial for transforming theoretical knowledge into real competencies and for strengthening the perceived utility of technology through haptic experience.

6 Conclusion and Future Work

The presented investigation confirmed that the choice of instructional environment for secondary vocational students in economic fields fundamentally influences not only their motivation but also the process of building algorithmic competencies. It was demonstrated that physical hardware is an indispensable element for building positive attitudes toward technology and for experiencing the meaningfulness of work (the sense of a "real" product).

Conversely, the virtual simulator proved to be a more effective tool for developing procedural knowledge and debugging, as it provides a safe environment free from technical noise. The primary contribution of this work is the finding that for the target group of "non-techies," a purely software-based path is not sustainable in the long term regarding interest, while a purely hardware-based path can be demotivating due to its technical complexity.

6.1 Future Research

The results open space for further exploration in several directions:

Investigating whether students who worked with hardware maintain interest in technology in later years of study more than those who used only simulation.

Focusing on how hardware specifically assists (or hinders) students with weaker initial competencies compared to gifted individuals.
Determining whether the ability to fix an error in a simulator leads to a better capacity for solving technical problems in the real world.

References

1. Honey, M.A., Pearson, G., Schweingruber, H.: STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research. National Academies Press, Washington, DC (2014). https://doi.org/10.17226/18612
2. Moldoveanu, M., Milkova, E.: The influence of students' personality traits and their parents' parenting style on Romanian students' performance in science subjects from secondary to doctoral level. J. Contemp. Cent. East. Eur. **32**(3), 711–732 (2024). https://doi.org/10.1080/25739638.2024.2426331
3. Osborne, J., Simon, S., Collins, S.: Attitudes towards science: a review of the literature and its implications. Int. J. Sci. Educ. **25**(9), 1049–1079 (2003)
4. Schwab, K.: The fourth industrial revolution: what it means, how to respond. In: Handbook of Research on Strategic Leadership in the Fourth Industrial Revolution, pp. 29–34. Edward Elgar Publishing, Cheltenham (2024)
5. Wing, J.: Computational thinking. Commun. ACM. **49**(3), 33–35 (2006). https://doi.org/10.1145/1118178.1118215
6. Piaget, J.: The Principles of Genetic Epistemology. Routledge & Kegan Paul, London (1972)
7. Brennan, K., Resnick, M.: New frameworks for studying and assessing the development of computational thinking. In: Proceedings of the 2012 Annual Meeting of the American Educational Research Association. Vancouver, 13–17 April 2012 (2012)
8. Petránek, K., Janečka, P., Milková, E.: Testing programming aptitude: an in-depth analysis. In: Efficiency and Responsibility in Education 2013, pp. 497–502. Czech University of Life Sciences, Prague (2013)
9. Sentance, S., Waite, J., Hodges, S., Macleod, E., Yeomans, L.: "Creating cool stuff": Pupils' experience of the BBC micro:bit. In: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, pp. 531–536 (2017). https://doi.org/10.1145/3017680.3017749
10. Microbit-Challenges: Micro:bit hardware access diagram [online image]. In: UCL's BBC Micro:bit Tutorials. [cit. 11. 2. 2026]. Dostupné z: https://microbit-challenges.readthedocs.io/en/latest/_images/microbit-hardware-access.jpg (n.d.)
11. Ball, T., Chatra, A., de Halleux, P., Hodges, S., Moskal, M., Russell, J.: Microsoft MakeCode: embedded programming for education in blocks and TypeScript. In: Proceedings of the 2019 ACM SIGPLAN SPLASH-E Symposium. Association for Computing Machinery, New York (2019)
12. Microsoft: Microsoft MakeCode for micro:bit [online]. [cit. 11. 2. 2026]. Dostupné z: https://makecode.microbit.org/#editor (n.d.)
13. Creswell, J.W., Plano Clark, V.L.: Designing and Conducting Mixed Methods Research, 3rd edn. SAGE Publications, Thousand Oaks (2017)
14. Kvale, S., Brinkmann, S.: InterViews: Learning the Craft of Qualitative Research Interviewing, 2nd edn. SAGE Publications, Thousand Oaks (2009)
15. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006)
16. Papert, S.: Mindstorms: Children, Computers, and Powerful Ideas. Basic Books, New York (1980) ISBN 0-465-04674-6

17. Sweller, J., Ayres, P., Kalyuga, S.: Cognitive Load Theory. Springer, New York (2011) ISBN 978-1-4419-8126-4
18. Voborník, P.: The motivation and learning by making. In: 12th International Conference of Education, Research and Innovation. IATED, Valencia (2019) ISBN 978-84-09-14755-7.
19. Němec, R., Hubálovská, M.: Computer measuring systems and their impact on the popularity of individual parts of the physics subject in primary schools in the Czech Republic. In: DIVAI 2020: 13th International Scientific Conference on Distance Learning in Applied Informatics. Wolters Kluwer, Praha (2020) ISBN 978-80-7598-841-6.

Critical Thinking and Media Use in Technical Education

Elod Gogh[1](✉) and Attila Kovari[1,2,3,4]

[1] Faculty of Informatics, Institute of Digital Technology, Eszterházy Károly Catholic University, Eger 3300, Hungary
gogh.elod@uni-eszterhazy.hu
[2] Institute of Computer Engineering, University of Dunaujvaros, Dunaujvaros 2400, Hungary
[3] Kandó Kálmán Faculty of Electrical Engineering, Institute of Electronics and Communication Systems, Óbuda University, Budapest 1034, Hungary
[4] GAMF Faculty of Engineering and Computer Science, John von Neumann University, Kecskemét 6000, Hungary

Abstract. This study explores the role of the digital information environment in increasing the amount of misleading and manipulated information to which students are exposed and, as a consequence, the importance of critical thinking and conscious media use as key aspects of the educational process. Specifically, it examines the extent of support that technical college students perceive regarding the development of critical thinking in school education and the role that critical thinking plays in problem-solving, personal development, and resistance to manipulation. A questionnaire-based quantitative research approach was used among 184 students of grades 9 to 11 of a technical college. For the analysis of the data, descriptive statistics, non-parametric correlation tests, and linear regression analysis were used. Overall, the findings show that critical thinking, as a whole, and critical thinking in relation to personal development are appreciated by the students, whereas the development of critical thinking in school education is perceived less clearly. A weak positive relationship between age and the total critical thinking index was found. From the point of view of the pedagogical process, the results suggest that critical thinking can be developed through the learning organization and not through technology alone.

Keywords: critical thinking · media literacy · secondary education

1 Introduction

In recent years, it has been observed that digital media have redefined the way in which information is accessed, filtered, and interpreted. In the context of online media, it has been shown that misinformation has the potential for wide-scale dissemination, thereby posing direct educational risks for adolescents, as they increasingly rely on digital media for day-to-day guidance [1]. It has been noted that the phenomenon of misinformation is further characterized by the presence of "information disorder," which includes misinformation, disinformation, and malinformation, thereby posing challenges for students in making credible judgments in authentic media contexts [2].

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 110–118, 2027.
https://doi.org/10.1007/978-3-032-32104-6_10

In this context, critical thinking has gained significant importance as a core cognitive and metacognitive ability. In this regard, it has been defined that critical thinking encompasses purposeful judgment in the form of analysis, evaluation, inference, as well as the justification of judgments based on evidence and reasoning [3]. In this regard, it has been emphasized that the transfer of critical thinking from one context to another requires not just cognitive skills but also dispositions that are helpful in providing control over one's own thinking [4]. In this context, secondary education has been identified as an important context in which students' autonomy-related learning strategies are still in the process of development, while the cultivation of self-regulated learning provides the basis for long-term reflective competencies [5].

Evidence from instructional studies shows that critical thinking does not develop automatically from the modernization of the curriculum or the availability of technology. Instead, it develops in a learning environment that makes critical thinking, argumentation, and reflection explicit learning goals [6]. Similarly, media and information literacy approaches highlight that, in the information age, it is not enough for an individual simply to have access to technology, but that it is necessary to evaluate information, verify information, and reflect on information, among other things [7]. Research into civic online reasoning also shows that the ability to evaluate online information has implications for civic competence and, therefore, autonomy in the online world [8].

Domestic perspectives are consistent with these findings. Challenges related to digital literacy have been understood as mainly pedagogical issues of relevance, reliability, and interpretation, rather than issues of access [9]. Research on the so-called net generations has shown that rapid consumption, divided attention, and superficial processing may interfere with deeper interpretation and may even facilitate manipulation. Therefore, teacher competence is crucial, as digital competence involves didactic awareness that can place technology at the service of thinking development, and current technological opportunities, including artificial intelligence, can contribute only when aligned with clearly formulated educational goals and reflective learning scenarios.

The present study focuses on perceptions, as perceived support and competence provide valuable feedback for teaching practice. The main objective is to examine how technical college students perceive support for developing critical thinking and how they interpret critical thinking in relation to problem-solving, personal development, and resistance to manipulation in the digital information space.

2 Theoretical Background

Critical thinking in educational research has traditionally been conceived as a system of cognitive and metacognitive processes, as opposed to a discrete and isolated skill. The fundamental elements of critical thinking include the analysis and evaluation of information, the evaluation of the quality of judgment, and reflective decision making [3], as well as the dispositions and metacognitive control that promote transfer to other subject domains [4]. From this perspective, the cultivation of critical thinking is closely related to reflective and self-regulated learning, as learners need to be aware of and regulate their own understanding [5].

The digital information environment creates a situation where critical thinking is increasingly needed, as credibility signals are often ambiguous, content is curated by

algorithms, and visual and rhetorical manipulation is common. The impact of these factors is also significantly felt in large Big Data-based systems and online learning environments [10, 11]. These features increase the relevance of media and information literacy, such as source evaluation and verification, and civic online reasoning, where evaluation practices are implemented in authentic online settings [7, 8]. Research findings about the spread of misinformation online [1] and descriptions of information disorder more broadly [2] support the claim that schools face a structural issue: students must be equipped to deal with high-speed, low-trust information conditions [12, 13].

From a pedagogical perspective, the literature suggests that effective development is based on instructional design and opportunities for reflective learning [14, 15], rather than the use of technology itself [6]. In domestic discourse, digital literacy and critical evaluation are seen as methodological problems in the processes of meaning-making and reliability assessment [9]. In addition, students' media practices and learning styles also play an important role in outcomes, as the characteristics of the net generation may hinder sustained reflective learning [16]. In this context, teachers' methodological awareness and didactic digital competence are seen as enabling factors, while new technologies, including artificial intelligence, can play an active role only when embedded in reflective instructional designs [17]. In addition, it is worth developing appropriate regulations and guidelines for the ethical use of mobile communication devices and social behaviors [18, 19]. Alongside the spread of AI, students are increasingly affected by human-machine interfaces, which may somewhat limit the development of STEAM skills and soft skills [20, 21].

3 Research Methodology

3.1 Research Objectives and Design

The research explored technical college students' perception of school support for the development of critical thinking and their evaluation of the role of critical thinking in problem-solving, personal development, and resisting manipulation. A quantitative research approach using a questionnaire-based research design was utilized to examine students' subjective perceptions and self-evaluations in a form that can be statistically analyzed. Consistent with the exploratory nature of the study, the research questions that formed the basis of the study were: RQ1: How do students understand the importance of critical thinking in the context of solving problems, individual growth, and defending oneself against manipulations? RQ2: To what extent do students think that education at school promotes the development of critical thinking? RQ3: What is the connection between CTI_total and age, sleep length, and attitude towards media usage?

3.2 Research Sample

Participants were students in grades 9 through 11 at the BGéSzC Kossuth Lajos Bilingual Technical College. Data were collected through an anonymous paper-based questionnaire administered in a classroom environment and then converted into electronic form. In total, 184 questionnaires were returned and included in the study. The sample included

145 boys and 39 girls, consistent with the technical and IT-focused profile of the institution. Participants were aged between 14 and 19 (M = 16.34, median = 16, SD = 1.08).

3.3 Instrument

Critical thinking perception was measured by five Likert-type items scored on a five-point scale ranging from "not at all" to "completely." The issues included the following aspects: (1) analysis of situations while solving problems; (2) contribution of education to critical thinking skills; (3) analytical abilities of the individual; (4) resistance to manipulation and misinformation; and (5) deep thinking as well as personal growth and quality of life. As a result, the CTI_total score was calculated as the sum of answers to the five questions, meaning that a higher value indicates a more positive attitude towards critical thinking and its importance in education. Therefore, the theoretical range for the Critical Thinking Index (CTI_total) varies from 5 to 25 points. Cronbach's alpha was 0.597.

3.4 Data Analysis

Data processing involved descriptive statistics, cross-tabulation, and non-parametric tests. Spearman's rank correlation was used to examine relationships between CTI_total and background variables. Chi-square tests were applied to examine relationships between demographic variables and individual items. A linear regression model was applied to examine predictors of CTI_total, with age, sleep duration, and attitudes toward media use entered as independent variables. The significance level was set at $p < 0.05$. The factor of age was considered to be among the important developmental variables, whereas the duration of sleep was viewed as a general measure of the state of health and functionality that can impact reflective learning and attention. The attitudes towards media usage became a part of this study since media habits can affect information evaluation.

4 Results

Across all five questions, students showed positive attitudes toward critical thinking and reflective thinking. Perceived educational contribution was moderate to high (Fig. 1), and students also rated reflective thinking positively in relation to resistance to manipulation and misinformation (Fig. 2).

CTI_total values observed in the sample ranged from 11 to 25, with most students scoring in the middle to upper range, most frequently between 17 and 21 (see Fig. 3).

A weak but statistically significant positive correlation was found between age and CTI_total ($\rho = 0.148$; $p = 0.045$), indicating that higher age was associated with slightly higher CTI_total values (Fig. 4).

No significant relationships were found in cross-tabulation analyses between gender, grade level, and the critical thinking items. Linear regression analysis revealed low explanatory power ($R^2 = 0.035$). Age was a significant predictor ($\beta = 0.45$, $p = 0.033$), and sleep duration and attitudes toward media use did not reach significance as predictors

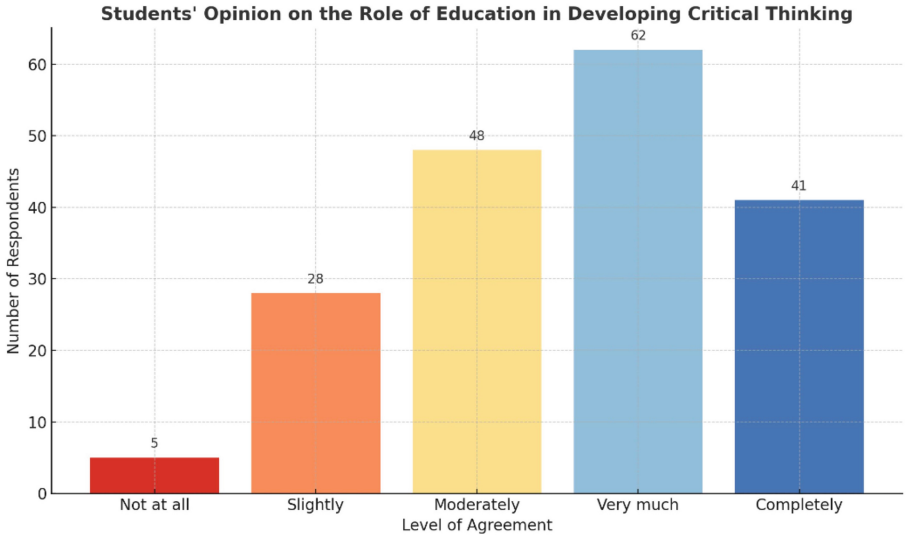


Fig. 1. Distribution of students' responses regarding the extent to which they believe school education supports the development of critical thinking.

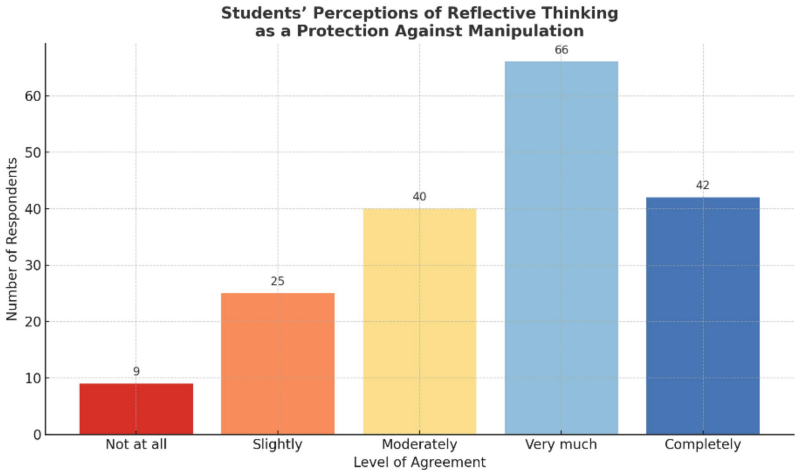


Fig. 2. Students' ratings of the role of reflective thinking in resisting manipulation and misinformation.

in the linear regression analysis. Sleep duration was negatively, although not significantly, correlated with CTI_total (Fig. 5)

The item-level association between age and the statement addressing resistance to manipulation and misinformation was positive but did not reach statistical significance ($\rho = 0.137$; $p = 0.064$).

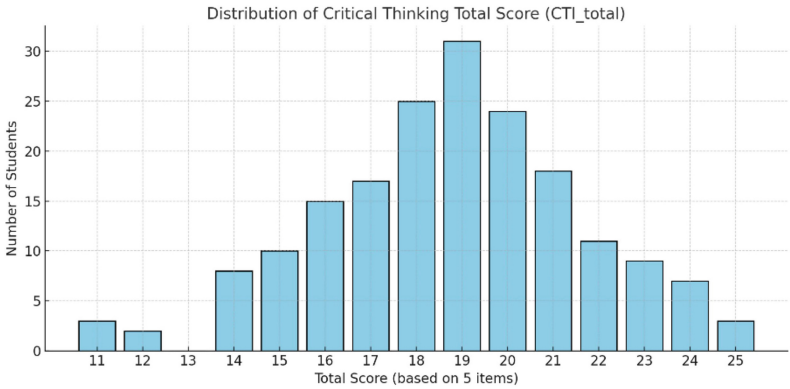


Fig. 3. Distribution of CTI_total scores based on the summed five-item Likert-type index.

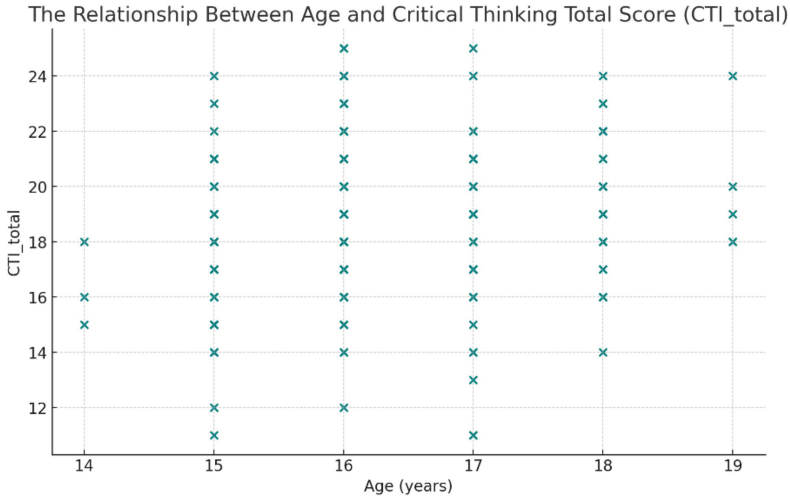


Fig. 4. Relationship between age and CTI_total, showing a weak positive association.

5 Interpretation and Pedagogical Implications

The results suggest that students generally perceive school support for the development of critical thinking positively, although the strength of this perception varies. This observation aligns with instructional evidence that the development of critical thinking depends on deliberate pedagogical design, as opposed to digital modernization. The weak but significant relationship between age and CTI_total, together with the positive but non-significant item-level association for resistance to manipulation, could imply that increased exposure to the information environment helps to foster an appreciation for the protective role of critical thinking.

Moreover, it is suggested that technology itself must be viewed as an environment, within which critical thinking is developed. As applied to the classroom setting,

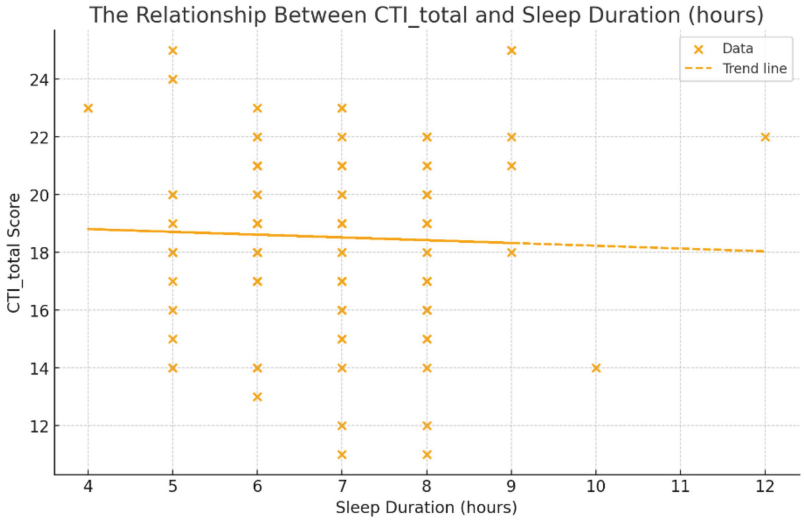


Fig. 5. Relationship between sleep duration and CTI_total, showing a negative but non-significant trend.

this means that it will be necessary to conduct short exercises involving comparing two sources, examining contradictory information available on the Internet, as well as discussing material produced by artificial intelligence.

The low explanatory power of the regression model indicates that self-assessed critical thinking is not well explained by basic demographic and lifestyle factors. This emphasizes the likely importance of classroom-level learning opportunities, teachers' methodological awareness, and reflective processes. Looking at teaching methodology, some approaches that can be taken to foster critical thinking in students include problem-focused assignments, source-based criticism, argument evaluation, and short checking activities with web-generated information or even artificial-intelligence-generated information.

6 Limitations and Further Research Directions

The generalization of the findings is limited by the fact that participants were drawn from only one technical college. Moreover, gender distribution reflects the institution-specific student profile of that college. Critical thinking was measured using self-assessment, which limits the interpretation of the findings in terms of actual competency levels. In addition, the present paper reports the structure of the instrument but not the full questionnaire wording, which limits exact replicability. Since the internal consistency of the five-item index was modest (Cronbach's a = 0.597), the index may cover several dimensions that have different development patterns; thus, further improvements can be made by developing subscales or combining the instrument with validated critical thinking tests.

Further studies could extend sampling to several institutions and diverse training profiles and use mixed-method approaches that combine questionnaires with interviews,

classroom observations, or portfolio analysis. Longitudinal studies could also help to clarify how perceptions of critical thinking and online media practices evolve together.

7 Summary

This research was focused on the analysis of technical college students' opinions concerning school assistance in promoting critical thinking and conscious media use among teenagers. Students were positively inclined toward critical thinking as well as its practical application in real-life situations but had less clarity about how it could be acquired systematically at school. A weak statistically significant positive association was found between respondents' age and the overall critical thinking index, whereas the item-level association between age and resistance to manipulation did not reach statistical significance. The regression model had low explanatory power, and neither sleep duration nor attitude toward media use was a statistically significant predictor. Consequently, one can make pedagogical conclusions about the necessity of organized learning process and reflection to develop critical thinking skills; otherwise, the presence of modern technologies does not necessarily lead to desired results.

References

1. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science. **359**(6380), 1146–1151 (2018)
2. Wardle, C., Derakhshan, H.: Information disorder: toward an interdisciplinary framework for research and policymaking. Council of Europe Report DGI(2017)09, Council of Europe, Strasbourg (2017)
3. Facione, P.A.: Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction (The Delphi Report). California Academic Press, Millbrae (1990)
4. Halpern, D.F.: Teaching critical thinking for transfer across domains: dispositions, skills, structure training, and metacognitive monitoring. Am. Psychol. **53**(4), 449–455 (1998)
5. Molnár, É.: Az önszabályozó tanulás. Iskolakultúra. **12**(9), 3–16 (2002)
6. Abrami, P.C., et al.: Strategies for teaching students to think critically: a meta-analysis. Rev. Educ. Res. **85**(2), 275–314 (2015)
7. Wilson, C., Grizzle, A., Tuazon, R., Akyempong, K., Cheung, C.-K.: Media and information literacy curriculum for teachers. UNESCO, Paris (2011)
8. Wineburg, S., McGrew, S., Breakstone, J., Ortega, T.: Evaluating Information: the Cornerstone of Civic Online Reasoning. Stanford Digital Repository, Stanford, CA (2016)
9. Molnár, G.: New challenges in the teacher career model with special regard to digital literacy. In: Diverse Pedagogical Culture: II. Educational Science and Professional Methodology Conference, pp. 365–373 (2014)
10. Molnár, G.: The role of electronic and virtual learning support systems in the learning process. In: IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI 2013), pp. 51–54 (2013)
11. Benedek, A., et al.: Practices of crowdsourcing in relation to big data analysis and education methods. In: IEEE 13th International Symposium on Intelligent Systems and Informatics, pp. 167–172 (2015)

12. Nagy, B., Cserné Pekkel, M., Takács, A.M., Váraljai, M.: Introducing AI into mathematics education: a course-based perspective. In: Szakál, A. (ed.) IEEE 25th International Symposium on Computational Intelligence and Informatics (CINTI 2025), pp. 121–126. IEEE Hungary Section, Piscataway (2025)
13. Kőkuti, T., Balázs, L., András, I., Rajcsányi-Molnár, M.: Collaborating with artificial intelligence - AI in business communication education. In: 6th International Conference and Workshop in Óbuda on Electrical and Power Engineering, pp. 287–294 (2023)
14. Takács, A.M., Takács, A., Augusztinyi, D.G., Nagy, B.: On some methods of digital assessment for mathematical courses. In: IEEE (ed.) 2024 IEEE 7th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE), pp. 313–316. Institute of Electrical and Electronics Engineers (IEEE), Piscataway (2024)
15. Kocsó, E.: A possible framework for examining student performance. J. Appl. Tech. Educ. Sci. **13**(3), 1–20 (2023)
16. András, I. et al.: Learning theories and the examination of the characteristics of new generations based on learning outcomes. In: Empirical Research in Education Across Borders, pp. 355–375 (2016)
17. Bartal, O., Rajcsányi-Molnár, M.: Teachers of the 21st century and the Mobile-tools. J. Appl. Tech. Educ. Sci. **10**(4), 53–66 (2020)
18. Molnár, G., Sik, D., Szűts, Z.: Methodological Possibilities of ICT-Based Mobile Communication Devices and Applications in Higher Education, New Ways of Learning, pp. 285–297. HERA (2017)
19. Molnár, G.: Technology and network-based learning forms and attitudes in the information society, with special regard to the basis of higher education. Információs Társadalom. **12**(3), 61–77 (2012)
20. Nagy, E., et al.: Exploring the role of human-robot interactions, within the context of the effectiveness of a NAO robot. Acta Polytech. Hungarica. **21**(3), 177–190 (2024)
21. Holik, I., et al.: The necessity of developing soft skills in STEM areas in higher education, with special focus on engineering training. Athens J. Technol. Eng. **10**(4), 199–214 (2023)

Student Perceptions of School Support for Creativity and Innovation in Technical Education

Elod Gogh[1] and Attila Kovari[1,2,3,4](✉)

[1] Faculty of Informatics, Institute of Digital Technology, Eszterházy Károly Catholic University, Eger 3300, Hungary
gogh.elod@uni-eszterhazy.hu, kovariati@outlook.com
[2] Institute of Computer Engineering, University of Dunaujvaros, Dunaujvaros 2400, Hungary
[3] Kandó Kálmán Faculty of Electrical Engineering, Institute of Electronics and Communication Systems, Óbuda University, Budapest 1034, Hungary
[4] GAMF Faculty of Engineering and Computer Science, John von Neumann University, Kecskemét 6000, Hungary

Abstract. Creativity and innovation skills have increasingly been recognized as key skills for the development of employability and lifelong learning. In the field of vocational technical education, these skills have specific relevance regarding the relationship with future professional performance. In the present research, the question is approached regarding the level to which students in a technical vocational school in Hungary perceive the support for the development of creativity and innovation skills in the school environment and the relationship with demographic and lifestyle-related variables. A quantitative methodological approach was chosen for the research. In the research, a total of 184 students completed the questionnaire. In the analysis, the relationships between the variables were examined with the help of Spearman rank-order correlation and linear regression analyses. In the research, it was found that the level of perceived support for the development of creativity and innovation skills is moderate or low. A weak negative relationship was found between the five-item index that indicates the level of support towards building creative and innovative abilities and the demographic factor of age. Moreover, a weak negative correlation was found between the index and perceived control over smartphone use, whereas a minor positive association was found between the index and sleep duration.

Keywords: creativity in education · innovation skills · student perceptions · vocational education

1 Introduction

Creativity and innovation are increasingly seen as key competencies for educational systems that have to deliver for a future characterized by uncertainty, technological change, and complex societal challenges. International policy documents highlight the importance of creative thinking in learning, work, and citizenship, and see it as a competence

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 119–128, 2027.
https://doi.org/10.1007/978-3-032-32104-6_11

that can be developed through schooling. This policy shift is seen in the prospective competence frameworks and in the more recent broad-based assessment frameworks that focus on creative thinking [1, 2].

This increased focus on creativity also challenges traditional approaches to learning that emphasize knowledge reproduction rather than the ability to frame problems, experiment, and be initiative-takers. Here, the role of the school is to create learning environments that promote initiative, collaboration, and risk-taking in a positive way, rather than focusing on content reproduction alone. Hungarian analyses of the state of education also highlight the importance of competence-oriented design and the promotion of learner agency in the learning environment for the future [3, 4].

Creativity-supportive learning environments tend to be marked by aspects of autonomy, open-ended tasks, and formative feedback that legitimizes exploration and multiple solution paths. While these can be instantiated in different ways, they demand a pedagogical design rather than sporadic classroom events [5, 6]. The challenge is especially salient in vocational and technical education because instructional conditions often face severe curricular and performance constraints, while labor market demands increasingly include independent problem-solving and creative thinking.

On one hand, the experience-based literature on creativity in an empirical or practice-oriented approach shows that students often experience constricted opportunities for creativity in formal education settings, including those that are digitally supported and in which creativity is assumed rather than designed. Discussions on the past in Hungary and abroad have pointed to structural barriers, teacher-centricity, or assessments as common barriers [7–10]. Creativity is heavily affected by the experience of the learning environment, and students' perceptions are thus a relevant source of evidence in determining whether the elements that support creativity are visible in the life of school. The impact of these factors is also significantly felt in large Big Data-based systems and online learning environments [11, 12].

The subject of the study is the ways in which students of the 9th to 11th grade of a technical vocational school perceive the school's support for creativity and innovation, and the relationship of these perceptions to demographic and lifestyle factors. The focus of the study is descriptive and explanatory, but not normative. Its aim is not to judge the phenomena under consideration, but to reveal patterns that could help in the development of creativity-supportive pedagogy in the context of vocational schools.

2 Theoretical Background

The definition of creativity generally involves the creation of new ideas or solutions that are appropriate within a particular context. This definition of creativity focuses on the fact that creativity not only involves newness, but it also involves appropriateness, and this definition can be used for all forms of learning, whether it is on a daily basis or on a high level [13].

Educational research is also moving towards a view that creativity is a developable capacity, not a fixed trait. Creativity is seen as emerging as part of normal classroom problem-solving when individuals create personally meaningful interpretations and establish new connections. From this perspective, creativity can be developed in

classrooms by legitimizing exploration and reducing the fear of making errors [14]. The concept of "mini-c" creativity emphasizes the personal dimension of creativity, which is particularly appropriate for adolescents whose creativity is still developing as a form of expression [15]. The Four-C Model also emphasizes different levels of creativity, from everyday creativity to eminent creativity, which supports the view that education needs to focus on developing everyday creativity, not just eminent creativity [16].

The componential model places the role of context in a central position across the main theories of creativity. According to the componential model, the key to creative performance lies in the interaction of domain-relevant skills, creative thinking processes, intrinsic motivational conditions, and a facilitating context. In terms of education, this suggests that the conditions of instruction and feedback may support or hinder creative engagement [17]. Further developments of the model highlight the role of organizational conditions and the dynamic interaction within an institutional context as a factor influencing creativity and innovation, a consideration which is particularly salient regarding schools as an institutional context [18]. In this regard, an exploration of students' perceptions of conditions for creative engagement provides a window into the institutional practice and the ways in which creative engagement is distributed among students. In addition, it is worth developing appropriate regulations and guidelines for the ethical use of mobile communication devices and social behaviors [19]. Alongside the spread of AI, students are increasingly affected by human-machine interfaces, which may somewhat limit the development of STEAM skills and soft skills [20, 21].

3 Research Methodology

3.1 Research Objective and Approach

The goal of the research was to examine students' perceived support for creativity and innovation in schools and to examine the relationships with some demographic and lifestyle-related variables. A quantitative design was used to allow for the statistical description and testing of relationships based on the students' self-reports.

3.2 Sample

The participants of this study consisted of students in grades 9, 10, and 11, all of whom were enrolled in a technical vocational program. The data was collected anonymously through a paper-based questionnaire that was later digitized for analysis. A total of 184 students completed the questionnaire. The gender composition of the participants was not even, with more male students participating, as expected of a technical vocational training program. The range of participants was between 14 and 19 years old ($M = 16.34$, median $= 16$, $SD = 1.08$).

3.3 Measurement Instrument

Perceived support for creativity/innovation in schools was measured using five Likert-type items with a five-point Likert scale. The items focused on perceived support for

encouraging independent thought, experimenting with new ideas, opportunities to do creative things, and supporting innovative ideas. The Creative Opportunities Index (COI) was computed by summing the five items, yielding a score ranging from 5 to 25. Cronbach's alpha was computed to examine internal consistency ($\alpha = 0.597$), which was deemed adequate since these items measure similar, though not the same, aspects of supporting creativity.

3.4 Data Analysis

Descriptive statistics were used to examine the distribution and measures of central tendency. Spearman rank correlations were also used to examine the relationship between COI and background variables. This is because the measurements were ordinal. Cross-tabulations and chi-square tests were also carried out for some items. A linear regression model was also developed to examine the combined effects of demographic and lifestyle variables on COI. Significance was established at $p < 0.05$.

4 Results

4.1 Overview of Perceived Creativity Support

Student feedback also shows that there is a moderate level of creativity support (Fig. 1). It seems that questions about the importance of teacher encouragement and thinking tend to be rated positively. This implies that most students believe that schools should foster thinking skills. However, questions about opportunities to develop and implement creative ideas are rated less positively.

4.2 Item Level Patterns

In terms of opportunities for developing and implementing creative solutions, most responses fall in the mid-range of the scale, with high ratings of support being rare. A similar pattern emerges in relation to the motivational and creativity-stimulating effect of the curriculum: students generally report that learning activities facilitate brainstorming and creative thinking to a limited or moderate degree. Judgments about the general school environment also fit this pattern, suggesting that the general climate of the school is rather regulated than experimental.

In summary, the results suggest that there is a disconnect between expectations and everyday experience. Creativity and reflective thinking are generally considered to be of great importance, yet the school environment is not always reported as conducive to the development of these abilities.

4.3 Distribution of the Creative Opportunities Index (COI)

Creative Opportunities Index scores show that the perception of opportunities for creativity is concentrated in the lower to middle part of the 5 to 25 scale. Most scores fall between 8 and 14, indicating that strong opportunities for creativity are not common and are found in a small percentage of students (Fig. 2).

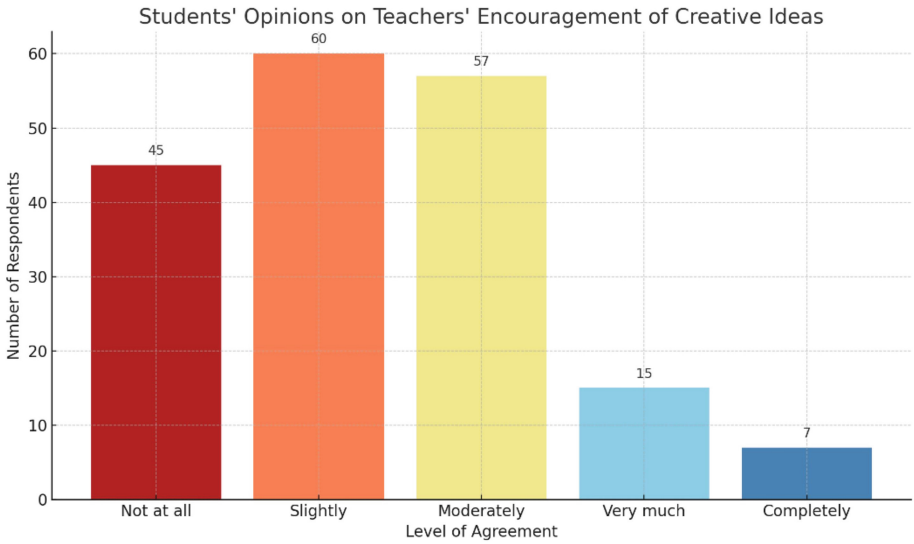


Fig. 1. Distribution of students' responses regarding perceived teacher encouragement for creative thinking.

4.4 Relationship Between COI and Age

A weak but statistically significant negative association was found between age and COI (Spearman rho = −0.166, p = 0.025) (Fig. 3). Older students tend to perceive slightly fewer creative opportunities than younger students.

4.5 Relationship Between COI and Digital Habits

For the most part, digital backgrounds did not show a statistically significant relationship with COI, suggesting that perceived school-level creativity support is not strongly influenced by patterns of digital media use in general. A single weak but statistically significant relationship was found: perceived control over smartphone use was weakly negatively related to COI (Spearman's rho = -0.18) (Fig. 4). Students who perceive lower levels of control tend to perceive slightly higher levels of creative opportunity.

4.6 Relationship Between COI and Sleep Duration

A slight positive trend is visible between the amount of sleep and COI on the linear regression graph. The estimates show that each additional hour of sleep is associated with a 0.13-point increase in COI (Fig. 5). This is minor and has much variation, but it is in the expected direction as it is assumed that well-being and restfulness are linked to the learning environment.

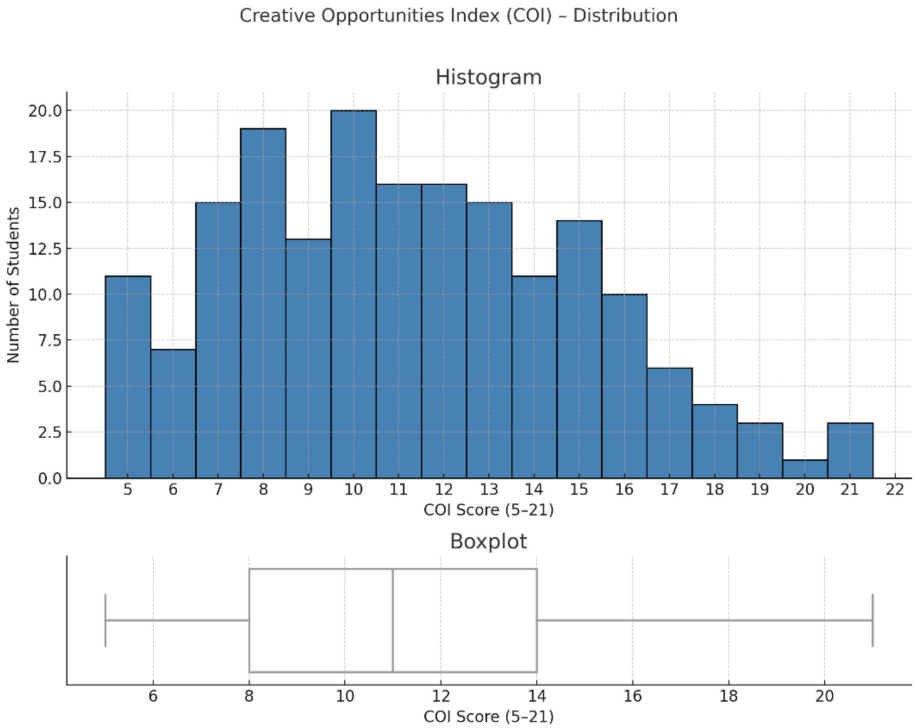


Fig. 2. Distribution of COI scores.

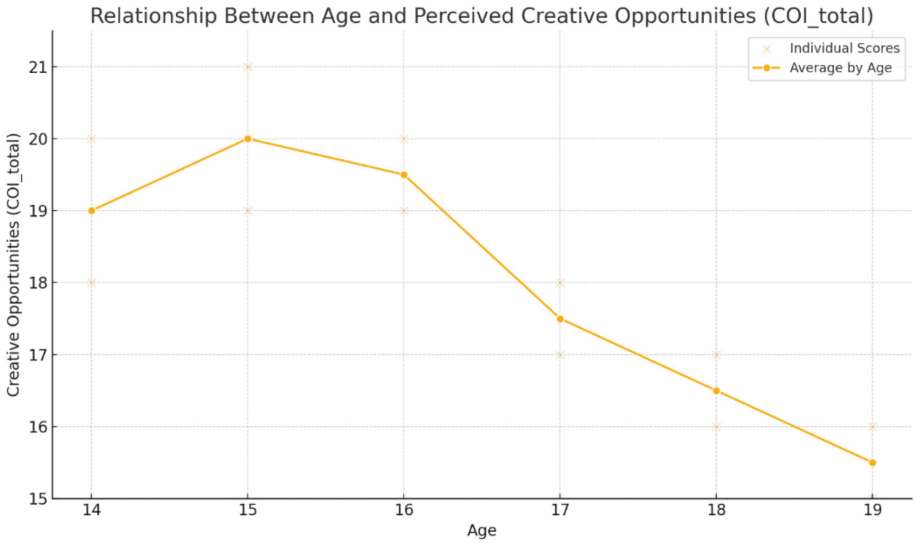


Fig. 3. Relationship between COI and age.

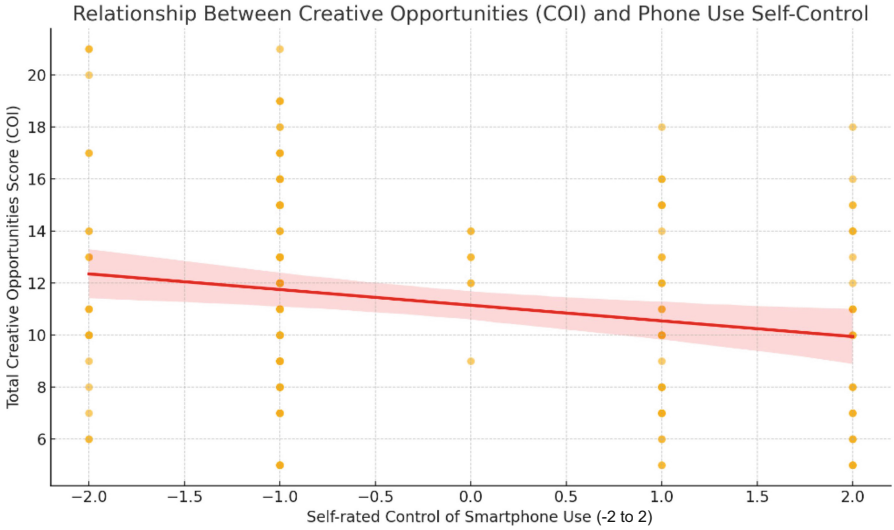


Fig. 4. Relationship between COI and self-control of smartphone use.

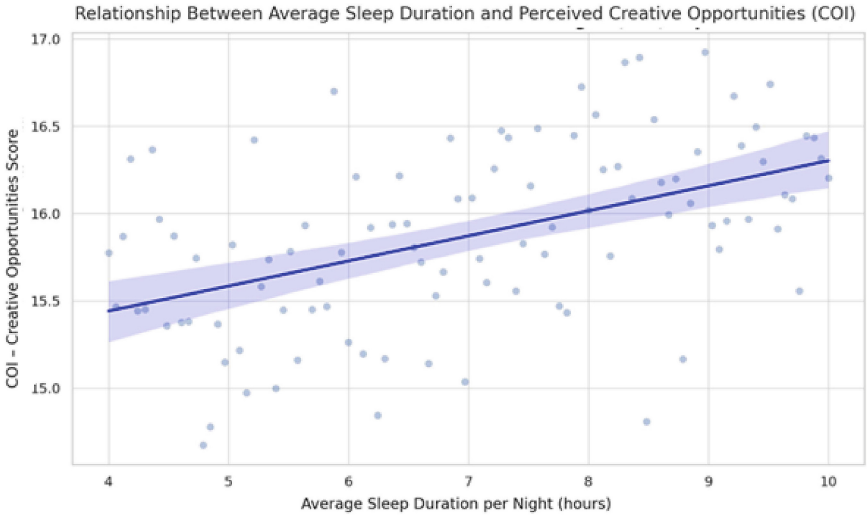


Fig. 5. Relationship between average sleep duration and COI.

5 Interpretation and Pedagogical Implications

The results suggest that generally, students perceive that there are only moderate or low levels of support for creativity and innovation in schools. This is supported by all results at the item level as well as the overall COI results. It implies that creativity support is not generally experienced as an integrated part of normal school life. It also implies that there are inequalities in the opportunities that exist for creativity support,

as high COI scores are rare and are confined to a narrow subgroup. At the same time, the statistically significant associations observed in the study are weak, therefore their practical significance should be interpreted with caution.

The slight negative correlation between age and perceived opportunities for creativity suggests that perceived opportunities decrease somewhat with increasing age. Due to the very low effect size, this trend requires careful analysis. One possible explanation is that the organization of the instruction appears more and more curriculum-focused and performance-oriented as the grade levels progress. This has important implications because the teenage years represent an important period of development in which autonomy and self-directed problem-solving are highly important.

The findings on digital habits do not support a simplistic view that creativity-related experiences at school are mainly due to more general digital lifestyles. This relationship is relatively tenuous, and thus it cannot be viewed as a primary cause. Instead, it needs to be seen as a relatively small correlation that may point to general trends in self-regulation and involvement in class.

As shown by the results of the regression model, there is a weak positive association between sleep duration and COI, which means that the level of well-being among students might have little relation to the perceived availability of creative opportunities. Sleep is not likely to be an important factor in the facilitation of creativity.

From a pedagogical developmental perspective, the results suggest that there is a need to think through environments that promote creativity in vocational education. Creativity support will be more visible when there are regular features of open-ended tasks, multiple solution paths, and formative feedback, as well as when there are institutional demands that promote these objectives. Well-being and self-regulation can be seen as potential facilitators of creative engagement. Given the weak associations, these implications concern general school design rather than strong predictors of individual student perceptions.

6 Limitations and Further Research Directions

Several limitations should be considered. Firstly, the findings are based on students' self-reported perceptions that may differ from actual teaching practice. Secondly, this study used a cross-sectional design. Therefore, the observed relationships should not be viewed causally. Finally, COI taps multiple but related aspects of creativity support. Future research should explore using factor analytic methods to test multi-dimensional structures of COI.

The generalizability of this study's findings is limited because this research was conducted at a single institution where the gender ratio was typical for technical and vocational training. Future research should include a broader sample. In addition, future research should include qualitative research methods such as interviewing students to determine what students mean by "creative opportunity" and what teaching situations students perceive as truly supportive. Longitudinal research would also be helpful to determine whether the age effect observed generalizes or whether it changes across developmental periods.

7 Conclusions

The present study aimed to examine the level of perceived school support for creativity and innovation among students in a technical vocational school and the relationships between these perceptions and certain background factors. The results show that most of the students perceived moderate or low levels of school support for creativity and innovation. The distribution of COI values shows that most of the values are distributed in the lower to mid-range of the scale, and perceived opportunities for creativity show a weak decline with increasing age.

Most of the digital lifestyle variables show no significant relationships with the COI, but weak associations were found with both control over smartphone use and sleep duration. These results suggest that well-being and self-regulation might play a role in the experience of opportunities for creativity and innovation. At the same time, these effects are small, so they should be interpreted as modest tendencies rather than strong explanatory factors.

In total, the results suggest that creativity support in the context of vocational education might not always be experienced as a stable and systemically integrated construct of the school context. Therefore, the design of supportive learning environments for creativity might require a holistic approach that goes beyond the design of the learning environment itself.

References

1. OECD: The Future of Education and Skills: Education 2030: the Future we Want. OECD (2018)
2. OECD: PISA 2022 Results (Volume III): Creative Minds, Creative Schools. OECD (2024). https://doi.org/10.1787/765ee8c2-en
3. Pellegrino, J.W., Hilton, M.L.: Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. The National Academies Press, Washington, DC (2012)
4. Váraljai, M., Nagy, B.: A survey in issues of disruptive technologies to broaden learning for the future students. In: 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2019), pp. 391–396 (2019)
5. Cropley, A.: Creativity in Education and Learning: A Guide for Teachers and Educators. Routledge, London (2015). https://doi.org/10.4324/9780203826270
6. Lucas, B., Claxton, G., Spencer, E.: Progression in student creativity in school: first steps towards new forms of formative assessments. In: OECD Education Working Papers No. 86. OECD Publishing, Paris (2013). https://doi.org/10.1787/5k4dp59msdwk-en
7. Bereczki, E.O.: Developing creativity in a learning environment supported by digital tools: what does research tell us for the classroom and when do we listen? Iskolakultúra. **29**(4–5), 50–70 (2019). https://doi.org/10.14232/ISKKULT.2019.4-5.50
8. Nagy, B., Váraljai, M., Mihalovicsné Kollár, A.: E-learning spaces to empower students' collaborative work serving individual goals. Acta Polytech. Hungarica. **17**(2), 97–114 (2020)
9. Hunya, M.: Kreativitás és innováció az oktatásban. Modern Iskola. https://modernisk ola.hu/2014/11/hunya-marta-kreativitas-es-innovacio-az-oktatasban/ (2014). Last accessed 2026/02/10
10. Vass, V.: A kreatív iskola. Anyanyelv-pedagógia. **5**(1), 1–8 (2012)

11. Molnár, G.: The role of electronic and virtual learning support systems in the learning process. In: 8th IEEE International Symposium on Applied Computational Intelligence and Informatics, pp. 51–54 (2013)
12. Benedek, A., Molnár, G., Szűts, Z.: Practices of Crowdsourcing in relation to big data analysis and education methods. In: 13th International Symposium on Intelligent Systems and Informatics, pp. 167–172 (2015)
13. Runco, M.A., Jaeger, G.J.: The standard definition of creativity. Creat. Res. J. **24**(1), 92–96 (2012). https://doi.org/10.1080/10400419.2012.650092
14. Sternberg, R.J., Lubart, T.I.: Defying the Crowd: Cultivating Creativity in a Culture of Conformity. Free Press, New York (1995)
15. Beghetto, R.A., Kaufman, J.C.: Toward a broader conception of creativity: a case for "mini-c" creativity. Psychol. Aesthet. Creat. Arts. **1**(2), 73–79 (2007). https://doi.org/10.1037/1931-3896.1.2.73
16. Kaufman, J.C., Beghetto, R.A.: Beyond big and little: the four C model of creativity. Rev. Gen. Psychol. **13**(1), 1–12 (2009). https://doi.org/10.1037/a0013688
17. Amabile, T.M.: Creativity in Context. Update to The Social Psychology of Creativity. Routledge (1996). https://doi.org/10.4324/9780429501234
18. Amabile, T.M., Pratt, M.G.: The dynamic componential model of creativity and innovation in organizations. Res. Organ. Behav. **36**, 157–183 (2016). https://doi.org/10.1016/j.riob.2016.10.001
19. Molnár, G., Sik, D., Szűts, Z.: Methodological Possibilities of ICT-Based Mobile Communication Devices and Applications in Higher Education, New Ways Of Learning, pp. 285–297. HERA (2017)
20. Nagy, E., et al.: Exploring the role of human-robot interactions, within the context of the effectiveness of a NAO robot. Acta Polytech. Hungarica. **21**(3), 177–190 (2024)
21. Holik, I., et al.: The necessity of developing soft skills in STEM areas in higher education, with special focus on engineering training. Athens J. Technol. Eng. **10**(4), 199–214 (2023)

Inquiry-Based Learning in Secondary Education Through Scavenger Applications

Imrich Jakab(✉), Zuzana Pucherová, Martina Agarski, and Dana Kollárová

Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia
{ijakab,zpucherova}@ukf.sk, {martina.agarski, dana.kollarova2}@student.ukf.sk

Abstract. The integration of digital technologies into education creates new opportunities for implementing student-oriented and inquiry-based approaches to learning. This article focuses on the application of inquiry-based learning in lower secondary education through the use of Scavenger Hunt Applications to create Interactive Educational Trails. Inquiry-Based Learning promotes active students' engagement by positioning them as researchers and emphasizing observation, data collection, and reflection. The study presents a conceptual and practical framework for implementing all four levels of Inquiry-Based Learning, by using Scavenger Hunt Applications. Special attention is paid to the highest level of Inquiry-Based Learning, where students independently design and create interactive educational trails. The methodological section describes a case study with 7th grade students who developed an interactive educational trail using the AVAtrails application as part of the Erasmus+ AVATAR project. The results indicate that this approach increases students' motivation, collaboration, digital skills, as well as their understanding of the surrounding environment. Despite the challenges students encountered during the creation of the Interactive Educational Trails, the results suggest that the combination of Inquiry-Based Learning, Scavenger Hunt Applications, and outdoor education represents an educational strategy with great potential for formal education.

Keywords: Inquiry-Based Learning · Scavenger Hunt Applications · Interactive educational trails

1 Introduction

The integration of digital technologies into education has become an integral part of contemporary schooling. In an increasingly digital society, digital literacy is essential for meaningful participation in education, work, and social life. Those who lack these competencies are therefore at growing risk of exclusion, which may be understood as a form of modern illiteracy. According to [1], technological development has created new opportunities in education, and mobile and remote learning have become indispensable components of many educational institutions.

Despite these advancements, the use of digital technologies in formal education still faces resistance, particularly among older generations. These concerns are often

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 129–140, 2027.
https://doi.org/10.1007/978-3-032-32104-6_12

associated with reduced contact with the real environment and potential negative effects on social and cognitive development. However, they typically arise from uncritical or poorly guided use rather than from pedagogically grounded integration.

Inquiry-Based Learning (IBL) includes various teaching approaches that support student learning through structured and increasingly independent exploration of complex questions and problems, which often have no single definitive solution [2]. According to [3], IBL is an educational approach in which students take on the role of researchers independently formulating hypotheses, designing and conducting experiments, collecting relevant information, and drawing conclusions. IBL can be described as a type of education that encourages students to actively participate in the learning process through inquiry, with teaching approaches that are strongly student-centered and based on research-oriented activities [4].

IBL consists of four levels of instructional approaches [5]:

- Confirmation Inquiry characterized by planned activities in which students verify and confirm a certain principle, while the expected results are predetermined.
- Structured Inquiry involves students exploring a question set by the teacher while applying a predetermined and clearly structured procedure.
- Guided Inquiry involves students exploring a question set by the teacher while independently designing or selecting appropriate research methods.
- Open Inquiry is characterized by students independently formulating research questions and conducting research using methods they design or select themselves.

One example of how IBL can be implemented is through the use of scavenger hunt applications (SHA). SHA transforms the exploration process into an interactive experience in which participants use mobile devices to obtain clues, solve tasks, and navigate between locations instead of relying on traditional printed materials [6]. According to [7], scavenger hunt activities are an effective tool for promotion and educational outreach, as they are based on student-centered active learning that employs innovative and dynamic methods to increase student engagement in achieving educational goals. As noted by [8], scavenger hunt activities offer a flexible and easily implementable instructional format that can be adapted to different age groups and educational environments, while allowing instructors to adjust question difficulty to their specific needs.

SHA can be considered as an information and communication technology (ICT) tool, as they employ digital technologies to support access to information and interactive learning processes. According to [9], ICT refers to a range of technologies that provide access to information through telecommunication systems. ICT, as described in [10], integrates various traditional educational technologies and forms of working with information, thereby linking different types of knowledge and literacy while blurring the boundaries between learning environments such as home, school, work, and the community.

As [11] emphasizes, students increasingly take responsibility for their own learning and its assessment, while developing expertise in the learning process. Additionally, positive motivational outcomes were most frequently observed when ICT supported student engagement, research activities, writing and editing, and the presentation of work. Even greater motivation was observed when ICT was embedded in both teaching and learning, rather than used solely as a teaching support tool [12].

Based on the theoretical basis of SHA as ICT-supported educational tools, we present a conceptual proposal for their use in the creation of interactive educational trails (IET) at all four levels of IBL.

At the first level (**Confirmation Inquiry**), SHA can serve as an effective tool for acquiring, consolidating, and verifying knowledge. The teacher defines the research topic, prepares a clear procedure for completing the tasks, and provides the correct answers, which are made available to students through the application. At the second level (**Structured Inquiry**), the teacher designs the educational trail, including its content, target group, and individual stations. Students actively participate by creating selected parts of the trail, for example by designing interactive elements or focusing on a specific section (e.g., the final station). At the third level (**Guided Inquiry**), the teacher determines the theme, basic content structure, and target group. Students then independently design and create the entire educational trail, adapting its content and teaching methods to the selected group, such as their peers. At the fourth level (**Open Inquiry**), the teacher provides only basic input for creating the trail or defines its location. Students independently set goals, select the target group, identify specific places of interest, propose content, and create the entire trail. In this context, SHA serves as a creative platform through which students take full responsibility for the entire inquiry process.

The aim of this paper is to highlight the possibilities of implementing the highest level of IBL (Open Inquiry) through the creation of IET by students of a model primary school.

2 Methodology

The basic idea of the paper is the creation of educational trails by students, representing the fourth level of IBL. Through inquiry, students learn about the cultural, social, natural, and environmental aspects of their surroundings through direct contact, applying acquired knowledge in practice and creating new, practical values.

The creation of IET is implemented by using the AVAtrails SHA, developed as part of the Erasmus+ AVATAR (Caring for All Vulnerable Animals Through an Active Relationship with Nature) project, in which Department of Ecology and Environmental Science, Faculty of Natural Sciences and Informatics, participates in collaboration with Department of Informatics. The model group consisted of 7th-grade students (class 7.C) from Devínska Elementary School in Nové Zámky, under the guidance of geography teacher PaedDr. Iveta Homolová.

The IET was created using the AVAtrails SHA (www.teachinggreen.eu), which enables the creation of IET, with each station consisting of the following components:

A – location
B – information about the destination
C – interactive element
D – answer
E – feedback
F – additional information

During the creation of IET, stages of students' independent work (marked with numbers 1–8 in the methodological procedure) alternated with methodological guidance and feedback from lecturers and teachers (marked with letters A–H, supplemented by the implementation period).

At the beginning of the work with primary school students, we defined the following hypotheses:

H1: The creation of IET content is expected to be less time-consuming for students than its technical implementation within the AVAtrails web application.
H2: The practical nature of the students' projects (i.e., the applicability of the project outcomes in real-life contexts) will serve as a motivation for their engagement in the project.
H3: 7th-grade primary school students are capable of designing an IET for their peers.
H4: The content of the educational trail created by the students will be engaging for their peers.

Methodological Procedure:
The methodological procedure began in November 2024 with the Presentation of the IET concept, project assignment, and division of tasks (A). This initial stage transitioned into the Preparation phase (B), which lasted through December 2024. During this time, students focused on the Conceptual phase (1), determining the theme and content of the trail, conducting target group analysis, and selecting a suitable location. This was followed by the Analytical phase, containing field research and mapping (2), where students identified points of interest with narrative potential and collected necessary data, such as photographs and GPS coordinates. In December 2024, students presented their initial ideas during the Concept presentation (C). This led into the Creative phase (D), which took place during January and February 2025. This stage was dedicated to Content creation and storytelling (3), where information was transformed into an engaging format, and Gamification (4), involving the design of interactive elements and multimedia support. By the end of February 2025, students reached the Presentation of the proposed IET (E), where they shared their designs for individual stations to receive feedback from teachers and lecturers. The project then proceeded to the Technical implementation phase (F) in March 2025. During the Technical phase of creating an IET (5), students programmed the trail by setting GPS locations and defining game logic and rewards. This stage also included rigorous Testing and debugging (Alfa & Beta test) (6) to verify the trail's functionality, including a peer review where groups tested each other's work and provided constructive feedback. The final stages involved the Presentation of IET evaluation results (G) in March 2025, followed by the Presentation phase containing Launch and Media Coverage (7). To ensure public awareness, a Marketing strategy (H) using social media and posters with QR codes was implemented from March 2025 through April 2026. Finally, the project culminated in the Dissemination of results (8) at the ŽiVEKO 2025 student science conference in April 2025.

Due to the fact that the AVAtrails SHA was still under development as part of the AVATAR project, it did not become fully operational until May 2025.

3 Results

The results include student outputs presented by students of Devínska Elementary School in Nové Zámky during the implementation of IBL.

1. **Preparation phase – conceptualization**
   During the preparation phase, students developed the IET named "The Beauty of Devínska Elementary School", specifically for a target group of peers and students aged 12 and older. The trail is located within the grounds of Devínska Elementary School in Nové Zámky and its surrounding area. The team identified twelve distinct stations maintained at a minimum distance of 20 m from each other, including Entrance to the grounds, Circular flower bed, Pebble playground, Classroom in nature, Immortal pine tree, and Birdhouses, Traffic playground, Our trees, Millennium I and Millennium II, Alley of Peace, and a stop dedicated to Separated waste. To ensure the quality of the trail, students established a set of criteria including educational value, functionality, entertainment, and attractiveness. They also focused on the appropriateness for the target group, clarity and comprehensibility, representativeness, and overall aesthetics.
2. **Analytical and creative phase**
   The students came up with introductory texts, interactive elements with feedback, and additional information for each station. The students had seven options to choose from for the interactive elements, from which they selected three: one correct answer, multiple correct answers, and matching pairs.
3. **Technical implementation (Prototyping), testing and debugging**

Table 1 Evaluation of the success rate of correct answers to IET questions in the questionnaire

S.n.	Station name:	Correct answers	Incorrect answers	% of student success
1.	Entrance to the grounds	28	2	93,33
2.	Circular flower bed	27	3	90,00
3.	Pebble playground	21	9	70,00
4.	Classroom in nature	21	9	70,00
5.	Immortal pine tree	25	5	83,33
6.	Birdhouses	29	1	96,67
7.	Traffic playground	7	23	23,33
8.	Our trees	29	1	96,67
9.	Millennium I	28	2	93,33
10.	Millennium II	20	10	66,67
11.	Alley of Peace	28	2	93,33
12.	Separated waste	25	5	83,33

Given that between 2024 and 2025, work was underway on developing the AVA-trails application and resolving technical issues with the application, testing of IET content (difficulty and comprehensibility of questions) was carried out by students in the form of a questionnaire.

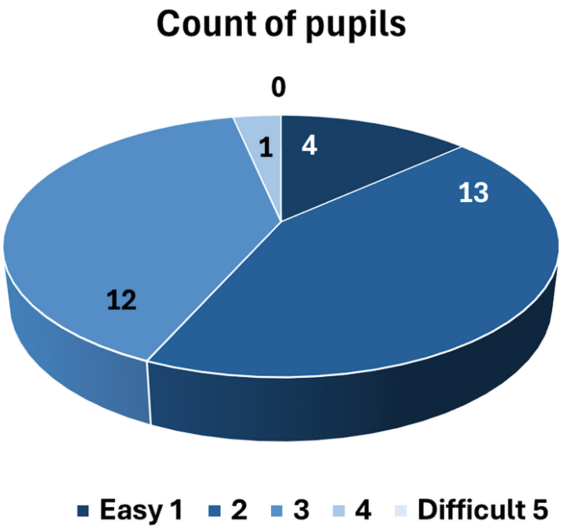
The respondents were 30 seventh-grade elementary school students. The students answered 12 questions, and the creators of IET were interested not only in the correct answers to individual questions (Table 1), but also in the difficulty and attractiveness of the IET questions. The students rated the overall difficulty on a scale of Easy 1 2 3 4 5 Difficult (Graph 1), and the overall interest of the questions on a scale of Interesting 1 2 3 4 5 Boring (Graph 2).

The assumption that the content created by students would be engaging for their peers (H4), can be confirmed based on data from the questionnaire survey. Respondents rated the IET as more interesting than boring (Graph 2), while also perceiving the questions as rather easy than difficult (Graph 1), which indicates that they were clear and comprehensible. The high rate of correct answers (Table 1) further supports this finding and shows that the creators were able to design an IET that is both attractive and appropriately tailored to their peers. Based on these results, students subsequently revised the final questions for each IET station.

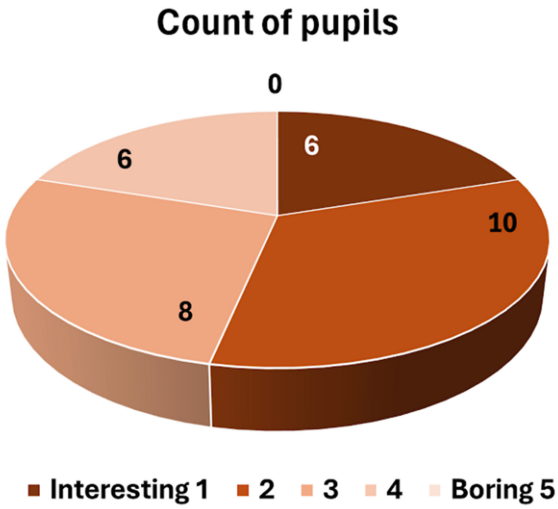
The successful completion of the project and the creation of a functional IET allow us to confirm the validity of the third hypothesis. The study results demonstrate that 7th-grade primary school students are capable of designing an IET for their peers. The students showed the ability to independently carry out all stages of IET development, from the initial conceptualization and selection of 12 stations to the analytical and creative phases, thereby confirming their competence to design a functional educational product.

4. **Marketing strategy**

The group of students who created IET registered for the ŽiVEKO 2025 Student Scientific Ecology Conference, which took place on April 16, 2025, at the Faculty of Natural Sciences and Informatics of the UKF in Nitra. Their contribution documenting the creation of IET was published in the Conference Proceedings, which is available at [13].



Graph 1 Assessment of the overall difficulty of questions by students



Graph 2 Evaluation of the overall interest of questions by students

4 Discussion

The creation of IET by students themselves represents a complex pedagogical model that merges **IBL**, **SHA** and **Outdoor Education**.

IBL is consistent with constructivist theories, which suggest that knowledge is more enduring when it results from active discovery rather than passive reception of information [14, 15]. An important benefit of IBL is the development of higher cognitive processes, such as critical thinking, scientific reasoning, and the ability to argue based

on evidence. According to [16], learning through inquiry can be considered to be the basis for developing reflective thinking, which is crucial for solving complex problems. Later research confirms that discovery-oriented teaching supports cognitive thinking and the ability to evaluate the quality of evidence [17, 18]. Moreover, systematic models of IBL show that students learn not only the content of the curriculum, but also the processes of scientific cognition themselves [19]. Another significant advantage of IBL is increased student motivation and engagement. Autonomy in posing questions and seeking solutions supports intrinsic motivation, which is consistent with self-determination theory [20, 21]. Research shows that students involved in IBL and project-based learning show higher interest in learning and more positive attitudes towards the subject than those in traditionally taught classes [22, 23].

In our case, the biggest motivation for the students was the fact that the result of IBL was the creation of an IET intended for practical use. Students expressed pride and enthusiasm in being the first creators of this type of ET in the city, as well as the youngest creators of IET in Slovakia. These findings confirm the validity of the second hypothesis, as the research indicates that the key motivational factor was not the use of digital technologies itself, but rather the real-world applicability of the students' work. Their enthusiasm stemmed from the opportunity to contribute to a project that will serve the needs of their school and the city in the future.

Similarly, the use of SHA in teaching increases student motivation and engagement through game elements such as points, tasks, and challenges [24]. Gamified activities promote intrinsic motivation and positive attitudes toward learning [25]. The use of both SHA and IBL supports a constructivist approach to learning, in which students construct knowledge through their own actions [26]. These location-based and mobile technologies enable contextual learning in real-world settings, such as cities or natural environments [27]. Research shows that location-based scavenger hunts improve understanding of subject matter and its connection to practice [28]. SHA promote collaboration and teamwork, as tasks are often solved in groups [29]. Collaborative learning through games develops students' communication and social skills [30]. SHA are suitable for interdisciplinary teaching, as they can combine language, science, history, and digital skills [31].

The creation of IET through SHA is largely carried out in the outdoor environment as a form of outdoor education. Outdoor education is a pedagogical approach that involves learning in natural or outdoor environments to support the cognitive, social, emotional, and physical development of students [32]. This approach is closely related to experiential and environmental education [33, 34]. Direct learning in an authentic environment increases the depth of understanding of the subject matter and its long-term memory [35]. Spending time and learning in nature has a positive effect on psychological well-being, reducing stress and anxiety. It also improves emotional regulation and students' self-confidence [36, 37]. Learning in nature supports physical activity, thereby contributing to a healthy lifestyle among children and youth [38]. Outdoor education can increase motivation to learn, especially in students with a weaker relationship with the traditional school environment [35].

IBL, the use of SHA, and outdoor education are more organizationally and logistically demanding to prepare than traditional classroom teaching [29, 38, 39]. Despite

its many benefits, IBL also has several limitations that are repeatedly discussed in the literature. One of the most frequently cited problems is the risk of excessive cognitive load, especially for students with low levels of prior knowledge. According to [40], minimal guidance during learning can lead to inefficient knowledge acquisition, as students' working memory is limited. This argument is supported by the theory of cognitive load, which suggests that unguided or poorly guided forms of inquiry can reduce the effectiveness of learning [41, 42]. Another limitation of IBL is its uneven effectiveness for different groups of students. Research indicates that students with lower metacognitive abilities or weaker basic knowledge may have difficulties in open-ended inquiry [43]. The effectiveness of IBL increases significantly when learning is accompanied by systematic pedagogical scaffolding and explicit teacher guidance [44].

Other difficulties may arise when using SHA in IBL. These problems are most often associated with technological requirements. Technological requirements, such as a smartphone, GPS, or a stable internet connection, can increase the digital divide among students [45]. Not all students possess the same level of technical skills, which can negatively affect their engagement in activities [46]. Other problems may arise during the stay in the outdoor environment itself. Outdoor activities involve safety risks, such as injuries or unpredictable situations in nature [32]. Teacher's responsibility for the safety of students may lead to concerns and restrictions on outdoor activities [47]. Not all students respond equally positively to outdoor education, particularly those with health limitations or low physical fitness [37].

In our case, the biggest negative was that the application for creating IET did not function properly. At the time, the AVAtrails application was under development and did not work perfectly on all mobile devices, which was demotivating for some students. The current version functions perfectly, and this limitation is not expected to recur in the future.

Despite this limitation, in relation to the evaluation of the project's time requirements, the first hypothesis (H1) was rejected. This hypothesis assumed that content creation would be less time-consuming for students than the technical implementation within the AVAtrails web application. However, an analysis of the methodological procedure timeline revealed that the creative phase focused on storytelling and gamification (Phase D) was carried out over a period of two months (January–February 2025), whereas the technical implementation of the trail (Phase F) was completed within one month (March 2025). This situation was primarily caused by the fact that students needed a considerable amount of time to find suitable and reliable information, which they then had to transform into an engaging and comprehensible form. The process of editing the texts in order to reduce their excessive length and ensure appropriateness for their peers required repeated consultations with teachers. From a temporal perspective, content development therefore proved to be a more demanding component of the project than its subsequent technical integration.

5 Conclusion

The findings of this study suggest that IBL supported by SHA can be considered a useful and effective approach to modern education. The creation of IET enables students to actively engage in the learning process, develop cognitive, social, and digital competencies, as well as connect theoretical knowledge with the real world. The presented study confirms that even at the level of lower secondary education, students are capable of successfully conducting research at the highest level if they receive appropriate guidance and support. Although this approach places greater demands on teachers and schools in regard to organization and technology, its educational benefits outweigh the identified limitations.

These findings are further supported by ongoing verification of the proposed model in practice. Currently, we are testing the proposed model of environmental education on smaller groups of students. In addition to primary school students, we have also had the opportunity to implement this model in cooperation with secondary school students, thereby expanding its applicability across different levels of education. The experience gained so far indicates its positive reception and effectiveness in the educational process.

Furthermore, there has been a significant improvement in the technical implementation. The technical phase is now proceeding more efficiently, mainly due to the elimination of the initial technical deficiencies of the application. This has contributed to increased overall efficiency and the removal of factors that could have negatively affected the IET development process.

Based on the experience gained from practical testing, we therefore recommend the implementation of this educational model into pedagogical practice, as it appears to be an effective tool for supporting environmental education and fostering students' active engagement in the learning process.

Acknowledgements. This article was written within the KEGA (037UKF-4/2025) and Erasmus+ (2023-1-SK01-KA220-SCH-000155430) projects.

References

1. Kang, J.: Benefits and challenges of mobile-learning brought to student learning outcomes in higher education: a systematic review from 2014-2023. Open Access J. **13**(1) (2024) E-ISSN: 2226-6348
2. Lee, V.S. (ed.): Teaching and Learning Through Inquiry: A Guidebook for Institutions and Instructors. Taylor & Francis (2023)
3. Keselman, A.: Supporting inquiry learning by promoting normative understanding of multivariable causality. J. Res. Sci. Teach. **40**(9), 898–921 (2003)
4. Levy, P., Little, S., Mckinney, P., Nibbs, A., Wood, J.: The Sheffield companion to inquiry-based learning. Centre for Inquiry-Based Learning in the Arts and Social Sciences (CILASS), The University of Sheffield, Brook Hill, UK (2010)
5. Banchi, H., Bell, R.: The many levels of inquiry. Sci. Child. **46**(2) (2008)
6. https://locatify.com/the-complete-guide-to-scavenger-hunt-apps/, last accessed 2026/01/17
7. Stark, R.K., Opuda, E., McElfresh, J., Kauffroath, K.: Scavenging for evidence: a systematic review of scavenger hunts in academic libraries. J. Acad. Librariansh. **47**(3), 102345 (2021)

8. Lijek, R.S., Fankhauser, S.C.: Using scavenger hunts to familiarize students with scientific journal articles. J. Microbiol. Biol. Educ. **17**(1), 125–128 (2016)
9. Ratheeswari, K.: Information communication technology in education. J. Appl. Adv. Res. **3**(1), 45–47 (2018)
10. Livingstone, S.: Critical reflections on the benefits of ICT in education. In: Digital Technologies in the Lives of Young People, pp. 9–24. Routledge (2015)
11. Carnoy, M.: ICT in education: possibilities and challenges. In: Inaugural Lecture of the UOC, 2005 (2004)
12. Passey, D., Rogers, C., Machell, J., McHugh, G., Allaway, D.: The Motivational Effect of ICT on Students, pp. 1–77. Department of Educational Research (2004)
13. https://www.fpvai.ukf.sk/images/KEE/kniznica/Zbornik_ZiVEKo_2025.pdf
14. Bruner, J.S.: The act of discovery. Harv. Educ. Rev. **31**(1), 21–32 (1961)
15. Ausubel, D.P.: Educational Psychology: a Cognitive View. Holt, Rinehart and Winston, New York (1968)
16. Dewey, J.: Democracy and Education: an Introduction to the Philosophy of Education, vol. 8. Macmillan, New York (1930)
17. Kuhn, D.: Science as argument: implications for teaching and learning scientific thinking. Sci. Educ. **77**(3), 319–337 (1993)
18. Osborne, J.: Teaching scientific practices: meeting the challenge of change. J. Sci. Teach. Educ. **25**(2), 177–196 (2014)
19. Pedaste, M., et al.: Phases of inquiry-based learning: definitions and the inquiry cycle. Educ. Res. Rev. **14**, 47–61 (2015)
20. Deci, E.L., Ryan, R.M.: Intrinsic Motivation and Self-Determination in Human Behavior. Springer Science & Business Media, Berlin (1985)
21. Ryan, R.M., Deci, E.L.: Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. The Guilford Press (2017)
22. Blumenfeld, P.C., Soloway, E., Marx, R.W., Krajcik, J.S., Guzdial, M., Palincsar, A.: Motivating project-based learning: sustaining the doing, supporting the learning. Educ. Psychol. **26**(3–4), 369–398 (1991)
23. Barron, B., Darling-Hammond, L.: Teaching for meaningful learning: a review of research on inquiry-based and cooperative learning. In: Sawyer, R.K. (ed.) The Cambridge Handbook of the Learning Sciences, pp. 275–297. Cambridge University Press (2008)
24. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness. Tampere, Finland (2011.09.28–2011.09.30). In: Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11, pp. 9–15 (2011)
25. Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work? Comput. Hum. Behav. **54**, 419–431. In proceedings of the 47th Hawaii International Conference on System Sciences, Hawaii, USA, January 6–9, (2014)
26. FitzGerald, E., Ferguson, R., Adams, A., Gaved, M., Mor, Y., Thomas, R.: Augmented reality and mobile learning: the state of the art. Int. J. Mobile Blended Learn. **5**(4), 43–58 (2013)
27. Sharples, M., Pea, R.: Mobile learning. In: Sawyer, K. (ed.) The Cambridge Handbook of the Learning Sciences: Second Edition, pp. 501–521. Cambridge University Press, New York, NY (2014)
28. Squire, K.: Mobile media learning: multiplicities of place. Horizon. **17**(1), 70–80 (2009)
29. Kapp, K.: The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education. Pfeiffer, San Francisco, CA (2012)
30. Johnson, D.W., Johnson, R.T.: An educational psychology success story: social interdependence theory and cooperative learning. Educ. Res. **38**(5), 365–379 (2009)

31. Looi, C.K., Seow, P., Zhang, B., So, H.J., Chen, W., Wong, L.H.: Leveraging Mobile Technology for Sustainable Seamless Learning: a research agenda. Br. J. Educ. Technol. **41**(2), 154–169 (2010)
32. Priest, S., Gass, M.A.: Effective leadership in adventure programming, 3rd edn, p. 448. ISBN: 9781492547860. Human Kinetics (2018)
33. Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development, p. 256. ISBN 0132952610. Prentice Hall, Englewood Cliffs, NJ (1984)
34. Beames, S., Higgins, P., Nicol, R.: Learning Outside the Classroom. Theory and Guidelines for Practice, p. 144. ISBN 978-0-415-89362-6. Routledge, New York, NY (2012)
35. Rickinson, M., et al.: A Review of Research on Outdoor Learning, p. 6. National Foundation for Educational Research and King's College London (2004)
36. Kaplan, R., Kaplan, S.: The Experience of Nature, p. 340. ISBN 0-521-34139-6. Cambridge University Press, New York, NY (1989)
37. Mygind, L., Kjeldsted, E., Hartmeyer, R., Mygind, E., Bølling, M., Bentsen, P.: Mental, physical and social health benefits of immersive nature-experience for children and adolescents: a systematic review and quality assessment of the evidence. Health Place. **58**, 102136 (2019)
38. Dyment, J.E., Bell, A.C.: Grounds for movement: green school grounds as sites for promoting physical activity. Health Educ. Res. **23**(6), 952–962 (2008)
39. Hargreaves, A., Fullan, M.: Professional Capital: Transforming Teaching in every School, p. 220. ISBN: 0807753327. Teachers College Press, New York, NY (2012)
40. Kirschner, P.A., Sweller, J., Clark, R.E.: Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educ. Psychol. **41**(2), 75–86 (2006)
41. Sweller, J., Aina, S.A.O., Aina, M.A.O.: Cognitive load theory. In: Psychology of Learning and Motivation, vol. 55, pp. 37–76. Academic Press (2011)
42. Mayer, R.E.: Should there be a three-strikes rule against pure discovery learning? Am. Psychol. **59**(1), 14–19 (2004)
43. Alfassi, M.: Reading to learn: effects of combined strategy instruction on high school students. J. Educ. Res. **97**(4), 171–185 (2004)
44. Lazonder, A.W., Harmsen, R.: Meta-analysis of inquiry-based learning: effects of guidance: effects of guidance. Rev. Educ. Res. **86**(3), 681–718 (2016)
45. Ally, M., Prieto-Blázquez, J.: What is the future of mobile learning in education? Mobile learning applications in higher education [special section]. Revista de Universidad y Sociedad del Conocimiento (RUSC). **11**(1), 142–151 (2014)
46. Traxler, J.: Learning with mobiles in the digital age. Pedagogika. **68**(3) (2018)
47. Waite, S.: Teaching and learning outside the classroom: personal values, alternative pedagogies and standards. Education 3–13. **39**(1), 65–82 (2011)

Didactic Potential of 3D Pen Activities Designed by Pre-service Teachers in the Context of the New National Curriculum: A Qualitative Content Analysis

Maria Anna Jedina(✉), Lilla Korenova, Angelika Schmid, and Katarina Zilkova

Faculty of Education, Comenius University Bratislava, Račianska 59, 813 34 Bratislava, Slovakia

jedina7@uniba.sk, {korenova,angelika.schmid, zilkova}@fedu.uniba.sk

Abstract. The rapid integration of digital technologies into primary education has intensified the need to examine not only technological tools themselves, but also how future teachers reason about their didactic use. In the field of technical education, 3D pens represent an accessible bridge between creative construction, embodied learning, and engineering-oriented STEAM approaches. However, limited research has focused on how pre-service teachers align such technologies with national curricular requirements. This study investigates the didactic potential of 3D pen activities designed by pre-service primary education teachers within the framework of the newly implemented National Curriculum (2023). The research explores how practical experiences with 3D pen integration manifest in the didactic reasoning of future teachers and how these activities align with selected curricular goals and performance standards. A qualitative content analysis was conducted on ten collaborative student projects (N = 50 participants). Each activity was evaluated using a goal–standard evaluation matrix (G1–G8 × S1–S7) on a 0–4 scale. Aggregated results were visualized through Sankey mapping to capture the intensity and structure of curricular alignment. The findings reveal strong alignment in areas related to project implementation, creativity, functionality, and problem-solving. Certain standards function as integrative nodes connecting multiple goals simultaneously. Less intensive representation was observed in communicative and regulatory dimensions. The study identifies a structured pattern of STEAM-oriented didactic reasoning, characterized by strong emphasis on project-based enactment and creativity, while communicative and regulatory dimensions remain comparatively underrepresented.

Keywords: 3D pen · Didactic Reasoning · Curriculum Alignment · STEAM education · Pre-service Teachers · Technical Primary Education

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 141–152, 2027.
https://doi.org/10.1007/978-3-032-32104-6_13

1 Introduction

The current digital era is characterized by the rapid emergence and progress of technologies, which brings an inherent and natural need to apply and integrate existing utilizable tools into the educational process. Within the environment of primary education, the 3D Doodler pen has gained significant attention as an accessible and intuitive bridge between digital modelling and the physical realization of objects [1]. Unlike stationary 3D printers, the 3D pen allows students and pupils to work in real-time, directly developing fine motor skills in conjunction with spatial awareness [2]. We decided to investigate the possibilities of using the 3D pen since immediate feedback on their own creation is critically important for primary school students. This tool enables real-time results and supports embodied learning through practical, hands–on engagement, which is essential at this developmental stage for understanding abstract concepts [3]. Using 3D pens in school settings is accessible to children's thinking and serves as a vital cognitive scaffolding for successfully transitioning from concrete manipulation to abstract spatial reasoning [4]. Technical education is currently shifting from traditional craft-based instruction toward engineering thinking and integrated STEAM perspectives [8]. This modern approach prioritizes divergent problem-solving and technical literacy over mere manual dexterity, aligning with the complex demands of the 21st-century curriculum [6]. While previous studies have explored the use of 3D pens in primary education and teacher education [1, 9], limited attention has been paid to how future teachers reason didactically about this technology. In particular, there is a lack of qualitative studies that systematically analyse pre-service teachers' reflections and project outputs in order to reconstruct their didactic reasoning regarding the educational use of 3D pens at the primary level. The aim of this study is to investigate the didactic potential of 3D pen activities designed by pre-service primary education teachers and to analyze how these activities align with the goals and performance standards of the newly implemented National Curriculum (2023).

2 Methods and Research Design

The research sample consisted of 50 pre-service primary education teachers (N = 50) currently enrolled in a master's degree program who also attended workshops about 3D printing. The study was conducted within the framework of the specialized university course *Technical skills in primary education* which focuses on modern didactic approaches in technical training. The participants' primary objective was to design original educational activities incorporating 3D pen technology as a tool for developing pupils' technical and creative skills. A key requirement of the instructional design process was the rigorous alignment of these activities with the content and performance standards defined by the newly implemented National Curriculum (2023). The students were tasked with selecting specific standards and demonstrating how their proposed activities fulfil the corresponding educational goals. These activity designs represent the primary data source for our analysis, allowing us to evaluate the didactic potential of the proposals and the students' ability to transform curricular requirements into innovative, technology-enhanced classroom practice.

Although the final dataset consists of 10 collaborative projects, it is important to emphasize the depth and complexity of these materials. In alignment with the principles of the new National Curriculum (2023), which emphasizes collaborative learning and group work as key pedagogical strategies, the participants also worked in groups. This allowed them to experience and implement the very cooperative methods they are expected to facilitate in their future classrooms. Each project was the result of intensive group work involving 50 pre-service teachers, ensuring a multi-perspective approach to instructional design and a high level of didactic reflection. This collaborative process allowed for a more robust integration of curricular standards than individual assignments might have produced. Consequently, these 10 units provide a sufficiently dense data source for a complex analysis via Sankey mapping, as they represent 50 unique viewpoints distilled into 10 structured educational frameworks. The thickness of the flows in our Sankey diagrams reflects this collective effort and provides a valid representation of how technical standards can be saturated through 3D pen technology.

The analysed projects are viewed as externalized manifestations of the students' internal pedagogical reasoning regarding the integration of 3D pen technology. To explore these underlying constructs, a qualitative content analysis was performed, targeting the explicit and implicit indicators of competency development within the instructional designs. To explore these underlying constructs, a qualitative content analysis [10] was performed, targeting the explicit and implicit indicators of competency development within the instructional designs. The analysis prioritized the depth of alignment between task architecture and the targeted competencies, such as technical literacy and divergent problem-solving. This qualitative framework facilitates an assessment of the didactic proficiency of pre-service teachers, specifically their capacity to operationalize the abstract requirements of the new National Curriculum into cohesive, competency-oriented pedagogical strategies.

As a practical example of the implementation of the previously defined standards and goals, we present the STEAM activity *Build a Bridge* which illustrates the process of transforming theoretical curricular requirements into a specific didactic design. The STEAM activity represents a comprehensive didactic model based on the principles of constructionism and project-based learning, where the 3D pen serves not merely as a technical tool, but as a medium for the material externalization of students' engineering reasoning. The primary objective is to develop technical and structural thinking by addressing authentic problems of stability and load-bearing capacity. Throughout the process, students explore the technical properties of materials-such as the strength, solidification, and adhesion of the filament-and seek their optimal application within an aesthetic design. The implementation follows a cyclical workflow, beginning with theoretical problematization and collaborative sketching, proceeding to the 3D modeling phase through material layering based on a template, and culminating in practical functionality testing under load. This methodological approach directly saturates the performance standards of the new National Curriculum in the area of *Man and the World of Work* [5] while fostering the students' ability to critically assess their own technical solutions and transform abstract structural rules into functional, stable products through formative assessment and reflection (Figs. 1 and 2).

Fig. 1. Bridge built by Doodler 3D Pen

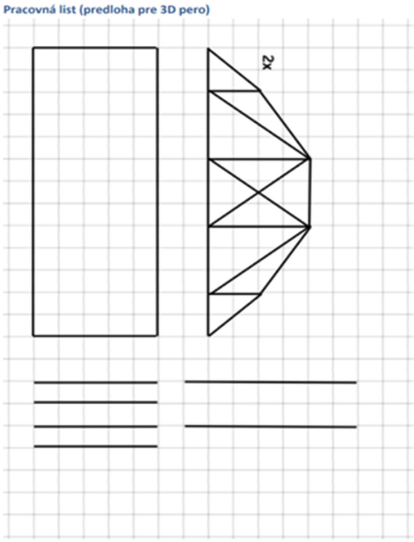


Fig. 2. STEAM Template

Sankey diagram is a visual analytical method used to represent flows between categories, where the magnitude of relationships is displayed through the width of the connecting bands. In educational research, this technique is increasingly employed to visualize distributions of coded qualitative data and transitions between competence domains. In this study, Sankey diagrams were utilized to map identified educational standards (S1–S7) against their corresponding national educational goals (G1–G8). This approach allows for a systematic tracking of how specific curricular standards, manifested in the students' activity designs, flow toward the fulfilment of broader objectives. By quantifying and visualizing these connections, the study provides a transparent overview of standard-to-goal alignment, highlighting which areas of the new curriculum are most effectively saturated by 3D pen integration.

2.1 Theoretical Framing of Didactic Reasoning

In this study, the concept of didactic reasoning is understood as a complex cognitive process through which pre-service teachers transform disciplinary knowledge into pedagogically meaningful and curriculum-aligned instructional situations. This construct is theoretically grounded in the framework of Pedagogical Content Knowledge (PCK) originally conceptualized by Shulman (1986) [14], and further extended through the Technological Pedagogical Content Knowledge (TPACK) model developed by Mishra and Koehler (2006).

Within the PCK framework, didactic reasoning refers to the capacity to transform subject-matter knowledge into representations and instructional strategies that make content comprehensible to learners (Shulman, 1986). The TPACK model expands this perspective by emphasizing the intersection of technological, pedagogical, and content

knowledge, highlighting that meaningful technology integration requires more than technical proficiency, it requires reflective and curriculum-sensitive instructional design [14]. In the context of 3D technologies in teacher education, previous research has shown that technology-supported instructional design fosters integrated STEAM-oriented pedagogical thinking, particularly when project-based and constructionist approaches are applied. Building on these foundations, we operationalize didactic reasoning into three analytically distinguishable, yet interrelated dimensions: **Instructional Planning Dimension** – how pre-service teachers conceptualize objectives, task structure, and methodological procedures in alignment with curricular goals and performance standards. **Design and Enactment Dimension** – the ability to operationalize technical content into structured project-based learning scenarios that foster problem-solving, creativity, and functional thinking. **Reflective and Regulatory Dimension** – the extent to which safety protocols, assessment strategies, presentation of outcomes, and metacognitive reflection are explicitly incorporated into instructional designs. Didactic reasoning is therefore conceptualized not as isolated technological decision-making, but as an integrative process linking content, pedagogy, technology, and curricular alignment within a STEAM-oriented educational framework [12].

RQ: How do practical experiences with 3D pen integration in technology education courses and specialized workshops manifest in the didactic reasoning of pre-service teachers regarding STEAM education and its alignment with the new National Curriculum?

2.2 Data Processing

The qualitative analysis of the 10 collaborative student projects was conducted through a structured coding process involving two independent coders. To ensure objectivity and analytical rigor, both coders systematically evaluated each project, identifying the manifestation of specific competencies and their alignment with the National Curriculum standards. Each alignment between a standard and an educational goal was evaluated on a 5-point scale (0–4), where 0 represents no potential and 4 represents a strong, systematically developed potential. Any discrepancies in the assessment were resolved through discussion to reach a consensus, and the final values were established by summing the scores from both coders across all analysed projects.

To assess the reliability of the coding procedure, inter-rater agreement was calculated prior to the consensus discussion. Across all 560 coded goal–standard units (G1–G8 × S1–S7), the initial percentage agreement reached 34.11%. Given the ordinal nature of the 0–4 evaluation scale, both unweighted and weighted Cohen's kappa coefficients were computed.

The unweighted Cohen's kappa was $\kappa = 0.157$, reflecting low exact agreement. However, when applying quadratic weighting to account for the degree of discrepancy between ratings, the weighted Cohen's kappa reached $\kappa w = 0.374$, indicating fair agreement beyond chance. This suggests that although exact agreement was limited, most discrepancies involved minor differences (e.g., adjacent scale points). Given the analytical complexity of evaluating multidimensional goal–standard relationships on a five-point ordinal scale, moderate levels of inter-rater agreement are not uncommon in qualitative educational research.

All disagreements were subsequently resolved through analytical discussion until full consensus was achieved. The final dataset used for aggregation and Sankey visualization therefore reflects consensually validated coding decisions.

This approach facilitated a seamless transition from detailed qualitative evaluations to the Sankey mapping, where the resulting visual flows represent the aggregate depth and intensity of the pedagogical links between the proposed activities, standards, and educational goals (Table 1).

Table 1. Assessment Scale for Curricular Alignment.

Score	Definition
0	No potential to fulfil the goal trough the given standard
1	Minimal potential to fulfil the goal through the given standard
2	Partial potential
3	Clear and feasible potential
4	Strong, systematically developed potential

The degree of alignment was assessed on a scale from 0 to 4: 0 (No alignment) for zero potential, 1 (Implicit) for minimal or indirect links, 2 (Partial) for activities requiring further refinement, 3 (Clear) for feasible, well-defined potential, and 4 (Systemic) for a strong, systematic integration of the standard into the core pedagogical strategy.

Overview of Educational Goals (G1–G8). For the purpose of this analysis, we extracted eight core educational goals. To ensure a systematic coding process and clarity within the subsequent visualization, these goals were operationalized and labeled with codes G1 through G8. This categorization allows for a precise mapping of how each student project aligns with the state-defined requirements for technical education. The following educational goals represent the strategic pillars against which the didactic potential was measured:

G1: Knowledge of materials and working procedures
G2: Application of knowledge in solving a technical task
G3: Implementation of a technical project
G4: Creativity and personal imagination in construction
G5: Safety and work regulations
G6: Presentation and communication of outcomes
G7: Work values and collaboration
G8: Practical and functional idea (usefulness)

List of Curricular Standards (S1–S7). The standards derived from the National Curriculum, define how pre-service teachers transform technical requirements into classroom practice. These standards serve as the operational nodes in our Sankey mapping to visualize the alignment between student projects and educational goals. For the purposes of this study, the following performance standards were analysed:

S1: Design and implement a technical project
S2: Create a sketch and construct a model accordingly
S3: Identify and solve a technical problem
S4: Use tools and materials safely
S5: Consider functionality, design and aesthetics
S6: Present and justify the work outcome
S7: Apply properties of technical materials

3 Results

Table 2. Matrix of Didactic Potential: Co-occurrence of Curricular Goals and Performance Standards

Goals/Standards	S1_Design and implement a technical project	S2_Create a sketch and construct a model accordingly	S3_Identify and solve a technical problem	S4_Use tools and materials safely	S5_Consider functionality, design and aesthetics	S6_Present and justify the work outcome	S7_Apply properties of technical materials
G1_Knowledge of materials and working procedures	58	23	41	14	26	6	75
G2_Application of knowledge in solving a technical task	31	62	63	9	53	12	19
G3_Implementation of a technical project	72	64	53	51	28	15	23
G4_Creativity and personal imagination in construction	52	66	27	10	72	14	13
G5_Safety and work regulations	26	15	13	68	20	8	20
G6_Presentation and communication of outcomes	21	5	36	7	14	66	5
G7_Work values and collaboration	54	9	34	30	10	54	5
G8_Practical and functional idea (usefulness)	57	31	63	11	66	24	25

As part of the research, a qualitative content analysis was conducted on ten 3D pen activities designed by pre-service primary education teachers within the course *Technical skills in primary education.* The objective of the analysis was to examine the didactic potential of these activities in the context of the new National Curriculum (ŠVP). Each

activity was evaluated through a matrix of relationships between eight curricular goals (G1–G8) and seven performance standards (S1–S7). The resulting Sankey visualization allowed for the capturing of the intensity and structure of the mutual interconnections between goals and standards based on aggregated evaluations.

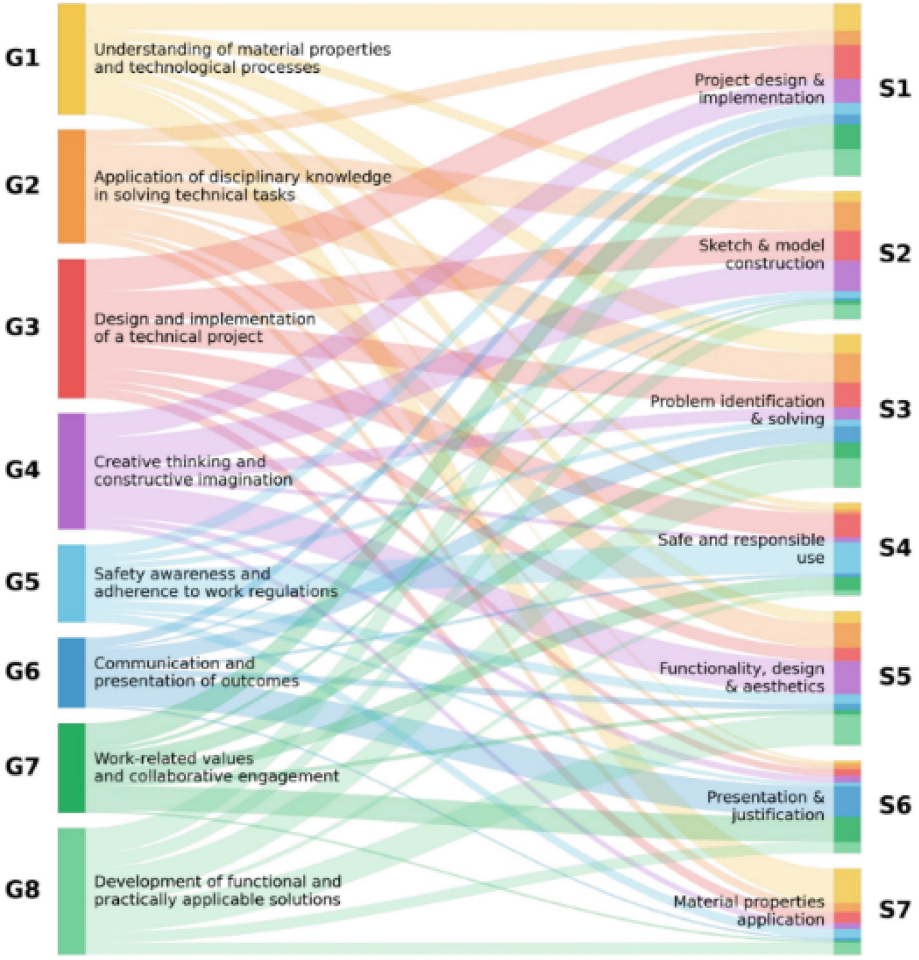


Fig. 3. Sankey Diagram of Goal–Standard Co-occurrence

The analysis demonstrates that the most significant didactic potential is concentrated primarily in goals related to the implementation of a technical project (G3), creativity and personal imagination in construction (G4), and the practical and functional utility of solutions (G8). The application of knowledge in solving technical tasks (G2) is also strongly represented. This pattern suggests that students, in their activity proposals, prefer a constructionist, project-oriented, and creative conception of technical education, where emphasis is placed on the active realization of a product and its functional value.

At the level of performance standards, there is a dominance of standards focused on the design and implementation of a technical project (S1), creating sketches and constructing models (S2), and identifying and solving technical problems (S3). Significant connections are also observed in the standard considering functionality, design, and aesthetics (S5). This result indicates that 3D pen activities are perceived as a means of developing practical and procedural skills that support technical thinking and the application of knowledge in specific construction situations.

The Sankey visualization further reveals a high degree of overlap between individual goals and performance standards. Several standard, particularly S1 and S3, perform an integrative function, contributing to the fulfilment of multiple curricular goals simultaneously. This suggests that the analysed activities are not isolated in nature but support complex and interconnected learning, where creativity, knowledge application, and project implementation are mutually dependent.

Relatively lower intensity is shown in areas related to safety and work regulations (G5), as well as the presentation and communication of outcomes (G6). At the level of performance standards, this is reflected in the weaker representation of standards regarding the safe use of tools and materials (S4) and the presentation and justification of work outcomes (S6) (Table 2). Although these dimensions are not entirely absent, they do not constitute a dominant part of the proposed activities (Fig. 3).

4 Conclusion

The findings suggest that 3D pen integration provides a specific context for STEAM-oriented didactic reasoning, particularly within the planning and enactment dimensions of instructional design. At the same time, reflective–regulatory components appear less systematically embedded, indicating a selective rather than uniformly balanced pattern of curricular alignment. The study identifies a structured profile of didactic reasoning among pre-service teachers, characterized by strong emphasis on project implementation, creativity, and functional problem-solving. Rather than uniformly saturating all curricular dimensions, the analysed activities reveal a dominant engineering-oriented pattern that prioritizes construction and product realization. For teacher education, these findings imply the need to move beyond purely technological training toward a more comprehensive didactic framework. Future preparation should emphasize not only the creative and constructive use of 3D technologies but also the systematic integration of reflective, communicative, and regulatory components—such as safety protocols, evaluation strategies, and presentation skills—to ensure balanced curricular implementation.

4.1 Limitation

A primary limitation of this study is the sample of ten collaborative projects, which may not capture the full range of individual student reasoning. Consequently, the findings do not provide direct evidence of how these activities would unfold during real-time interaction with pupils in a classroom setting, nor do they account for the unpredictable variables of school practice. Furthermore, although inter-rater agreement reached fair

levels when weighted measures were applied, the moderate reliability suggests that future studies should refine coding rubrics or incorporate calibration sessions prior to independent coding.

4.2 Future Research Direction

Future research should transition from the design phase to the implementation phase, validating these STEAM activities directly within primary education classrooms. A promising direction involves a comparative analysis between the planned didactic potential and the learning outcomes achieved by pupils. Additionally, longitudinal studies are recommended to track how systematic training in 3D technologies influences the long-term development of pre-service teachers' engineering thinking and their sustained willingness to innovate technical education according to the new National Curriculum requirements.

5 Discusion

The findings of the analysis provide an answer to the research question concerning how practical experiences with 3D pen integration manifest in the didactic reasoning of pre-service teachers within the context of STEAM education [13] and the new National Curriculum [5]. The Sankey visualisation does not merely quantify relationships between goals and standards, rather, it serves as an analytical tool for reconstructing the implicit pedagogical logic underlying the proposed activities. By implicit pedagogical logic, we refer to the underlying, not always explicitly articulated, structure of pedagogical priorities and decision-making that becomes visible through the alignment patterns embedded in instructional designs. The strong concentration of connections around goals G3 (implementation of a technical project), G4 (creativity and personal imagination), and G8 (practical functionality) suggests that pre-service teachers conceptualise the 3D pen primarily as a tool for project-based and constructivist learning. Their didactic reasoning appears to emphasise the active creation of a tangible outcome, problem-solving processes, and the practical application of knowledge in authentic contexts. Such a configuration corresponds to a constructivist-oriented model of technical education, in which learning occurs through active construction, experimentation, and engagement with real-world tasks [12]. This orientation aligns closely with STEAM principles [11], which stress the integration of creativity, engineering thinking, and contextualised problem-solving. Furthermore, performance standards S1 (design and implementation of a technical project) and S3 (identification and resolution of a technical problem) function as integrative nodes connecting multiple curricular goals simultaneously. This pattern indicates that the students' didactic reasoning tends towards a synthetic rather than fragmented understanding of technical education. The technological tool is not perceived as an isolated instructional add-on, but as a medium through which multiple competence domains can be developed concurrently. It can therefore be inferred that pre-service teachers perceive the 3D pen as a comprehensive pedagogical instrument capable of integrating constructional thinking, applied knowledge, and creative design within a coherent educational framework. Conversely, the comparatively lower intensity

observed in areas related to safety (G5) and presentation and communication (G6) indicates that regulatory and reflective dimensions of technical education are not equally foregrounded in the proposed activities. This does not imply their absence, rather, it suggests that the dominant orientation of the instructional designs prioritises product development and technical problem-solving, while metacognitive, communicative, and normative aspects are less systematically articulated. In other words, the engineering and construction dimension prevails over explicit reflective and argumentative components of learning. Within the broader STEAM framework, these findings may be interpreted as evidence of a pronounced engineering orientation, characterised by emphasis on design, construction, and functional optimisation. Practical experience with the 3D pen appears to activate primarily creative and project-based modes of thinking. At the same time, comprehensive curricular integration requires deliberate pedagogical structuring to ensure balanced representation of all targeted dimensions. From the perspective of teacher education, the results suggest that technology-oriented tasks can significantly contribute to the development of complex didactic reasoning. The 3D pen emerges as a suitable medium for fostering STEAM-oriented pedagogical thinking [8], particularly through its support of project planning, authentic problem situations, and functionally oriented design processes. Moreover, the analysis suggests its alignment within selected goals and performance standards of the new National Curriculum, as the proposed activities demonstrate substantial alignment between project implementation, creativity, and applied technical knowledge in accordance with curricular requirements.

Acknowledgement. KEGA 077UK-4/2025 “Edu-Steam laboratory in primary education teaching”.

Erasmus+ 2023-1-CZ01-KA220-HED-000160664: “Accelerating STEAM-related Knowledge and Skills via 3D Modelling and 3D Printing”.

References

1. Chun, H.: A study on the design education method using 3D pen in an era of manufacturing change. Nanotechnol. Environ. Eng. **7**(1), 243–251 (2022). https://doi.org/10.1007/s41204-021-00174-5
2. Smith3D: 3D Pen for Young Creators. https://www.smith3d.com/3d-pen-for-young-creators/ (2024), last accessed 2026/02/09
3. Takahashi, H., Kim, J.: 3D pen + 3D printer: exploring the role of humans and fabrication machines in creative making. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13. ACM, New York, NY (2019). https://doi.org/10.1145/3290605.3300525
4. Wei, H., Bos, R., Drijvers, P.: Developing functional thinking: from concrete to abstract through an embodied design. Digit. Exp. Math. Educ. **10**(3), 323–351 (2024). https://doi.org/10.1007/s40751-024-00142-z
5. State Educational Programme: Educational Standards – Man and the World of Work. Ministry of Education, Science, Research and Sport of the Slovak Republic, Bratislava. https://www.minedu.sk/data/files/11817_clovek-a-svet-prace.pdf (2023), last accessed 2026/01/08
6. Papadopoulou, E.A.: Advancements in STEAM education for 21st century learners. Int. J. Educ. **16**(4), 39–70 (2024). https://doi.org/10.5296/ije.v16i4.22270

7. Kassimova, M., Akhatayeva, U., Sydykova, Z., Jetpisbayeva, S., Yesnazar, A.: Formation of creative activities of primary school students: STEAM technology. J. Educ. Soc. Res. **15**(1), 139–152 (2025). https://doi.org/10.36941/jesr-2025-0139
8. Wendt, J., Beach, J., Wendt, S.: 3D printing: practical applications for K-16 education. In: Contemporary Approaches to Research in Learning Innovations, pp. 1–21. Brill (2020). https://doi.org/10.1163/9789004415133_001
9. Saldaña, J.: The Coding Manual for Qualitative Researchers, 2nd edn. SAGE Publications, London (2013)
10. Chappell, K., Hetherington, L., Juillard, S., Aguirre, C., Duca, E.: A framework for effective STEAM education: pedagogy for responding to wicked problems. Int. J. Educ. Res. Open. **8**, 100474 (2025). https://doi.org/10.1016/j.ijedro.2025.100474
11. Ulbrich, E., Tejera, M., Schmid, A., Sabitzer, B., Lavicza, Z.: Fostering the four C's: a gendered perspective on technology use in STEAM education. Educ. Sci. **15**(5), 528 (2025). https://doi.org/10.3390/educsci15050528
12. Korenova, L., Schmid, A., Tejera, M., Lavicza, Z.: 3D technologies in teacher education: reflections and didactic potential. In: Proceedings of the 24th European Conference on e-Learning (ECEL 2025), pp. 210–218
13. Shulman, L.S.: Those who understand: knowledge growth in teaching. Educ. Res. **15**(2), 4–14 (1986)
14. Mishra, P., Koehler, M.J.: Technological pedagogical content knowledge: a framework for teacher knowledge. Teach. Coll. Rec. **108**(6), 1017–1054 (2006)

Students' Intrinsic Motivation When Using Moodle in Informatics Classes

Ladislav Kalous[1] and Miloslava Cerna[2](✉)

[1] Faculty of Science, University of Hradec Kralove, Hradecká 1285, 500 03 Hradec Králové, Czech Republic

[2] Faculty of Informatics and Management, University of Hradec Kralove, Hradecká 1249/6, 500 03 Hradec Králové, Czech Republic

miloslava.cerna@uhk.cz

Abstract. This study examined upper secondary students' intrinsic situational motivation when using the Moodle learning management system in Informatics classes. The theoretical framework was Self Determination Theory (SDT), which highlights autonomy, competence, and relatedness as key conditions for intrinsic motivation. To capture motivation in a specific learning situation, students completed the Situational Motivation Scale, focusing on the intrinsic motivation subscale. The sample comprised 110 students (57 in Grade 1 and 53 in Grade 2). Data were collected at the beginning of the school year during a task based Moodle activity within regular class time. Group differences were tested using Welch's t test. Results showed a statistically significant grade level difference in favor of Grade 1, $t(107.67) = 2.08$, $p = .04$, with a medium effect size ($d = 0.40$). The findings indicate that intrinsic motivation in an LMS is sensitive to situational and instructional factors such as novelty, habituation, and perceived task demands. Implications for blended learning design include supporting autonomy through meaningful choice, competence through clear structure and formative feedback, and relatedness through guided peer interaction.

Keywords: intrinsic motivation · e-learning; Moodle · blended learning · SIMS · Self-Determination Theory (SDT)

1 Introduction

Digital technologies have reshaped educational practices over the last two decades, and learning management systems (LMS) have become a central infrastructure supporting both blended and online learning [1, 2].

In the Czech context, the development of e-learning accelerated notably at the beginning of the new millennium and gradually became a standard component of educational practice, particularly in upper-secondary and higher education [3]. Syntheses of Czech research highlight the stable institutional embedding of e-learning and its natural integration into both face-to-face and blended modes of study [4]. E-learning is often implemented not as a fully distance-based model but rather as part of blended learning, which

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 153–162, 2027.
https://doi.org/10.1007/978-3-032-32104-6_14

connects classroom instruction with online support and enables more flexible organization of learning activities [5]. Since 2000, the Czech educational system has gradually shifted toward developing digital competences and adopting online education. Analytical studies document this trend across all school levels. The current situation confirms that digital education is firmly established in the Czech Republic, supported in part by stable LMS infrastructure and sustained strategic investments in educational digitalization. Comparable trends are observed in the broader Central European context, where modern educational technologies and blended learning environments are increasingly emphasized as key components of contemporary instruction [6]. E-learning in the Czech Republic is not only widely implemented but also continuously reflected in scholarly research, which can be understood as an indirect indicator of its institutional consolidation and a stable technological base relying on LMS platforms [7].

One of the most widely used systems for supporting e-learning is Moodle, an open-source LMS platform adopted across educational institutions and maintained by a large international community. Beyond its technological infrastructure, Moodle is supported by an extensive ecosystem of pedagogical resources, training courses, and professional development opportunities [8]. The system provides a structured environment for distributing learning materials, assigning tasks, administering tests, facilitating communication, and conducting both formative and summative assessment. These affordances support individualized learning and enable systematic monitoring of learners' progress [9, 10]. In the Czech context, Moodle is well established not only at universities [11] but also in secondary schools, where it contributes to the development of students' digital literacy and their competence for autonomous learning [12]. At the same time, research consistently emphasizes that the mere presence of technology does not guarantee higher-quality learning: the key factors are course design, the nature of tasks, and the type of pedagogical support provided in the online environment [7, 12, 13].

Against this background, examining students' motivation in LMS-supported learning is particularly relevant, as motivational processes shape engagement and the quality of learning in blended environments. Student motivation emerges as a crucial variable influencing engagement, persistence, and depth of cognitive processing [14, 15]. This is especially relevant in e-learning, where students often work independently and their success depends heavily on self-regulation. Motivation can therefore determine whether the e-learning platform fosters active, meaningful learning or functions merely as a passive repository of materials [16].

Self-Determination Theory (SDT) offers a robust theoretical lens for studying motivational quality. SDT proposes that intrinsic motivation is supported when the needs for autonomy, competence, and relatedness are satisfied [17, 18]. In LMS-based learning, these needs can be shaped by design decisions such as meaningful options, clear structure and scaffolding, timely formative feedback, and opportunities for collaboration.

To capture motivation within a specific learning situation, it is appropriate to use tools designed to measure situational motivation. For this study, the Situational Motivation Scale (SIMS) [19] is employed. Developed within the SDT framework, SIMS enables measurement of momentary motivational states during a particular activity. The instrument is brief, practical for use in school settings, and its psychometric properties have been validated in multiple educational contexts [20]. Given the focus of this research,

attention is concentrated on the dimension of intrinsic motivation, which within SDT represents an especially adaptive form of motivation associated with interest, enjoyment, and deeper learning. In this study, intrinsic motivation was examined through four SIMS items targeting students' interest and enjoyment, specifically: "Because you feel good while doing these activities," "Because you think these activities are pleasant," "Because you find these activities fun," and "Because you find these activities interesting."

In Informatics education, LMS-based learning often involves task-based practice, e.g., spreadsheets, data handling, algorithmic thinking, or introductory programming. Such tasks can vary in perceived difficulty; if students experience competence frustration due to unclear instructions or steep task difficulty, intrinsic motivation may decline even when the technology is stable. Conversely, well-structured tasks with achievable challenge and feedback can sustain intrinsic interest.

This paper investigates intrinsic situational motivation of upper-secondary students while using Moodle in Informatics classes at the beginning of the school year. The early-year measurement captures motivation before extensive routine develops, when novelty effects may still be present.

Research question (RQ): Does the level of intrinsic motivation of upper-secondary students working in the Moodle e-learning system during informatics classes differ between Grade 1 and Grade 2 students?
Hypothesis (H1): We expect a statistically significant difference in intrinsic motivation between the two grade levels.

The study draws on data collected using the SIMS questionnaire from 110 students (57 Grade 1 students and 53 Grade 2 students) at the beginning of the school year while working in the Moodle environment. The following section therefore describes the research design, the instrument used, the participant selection, the data collection procedure, and the statistical analysis employed to compare the two groups.

2 Methodology

2.1 Research Design

The study employed a quantitative cross-sectional design with data collected at a single time point. The design was appropriate for comparing two independent groups (Grade 1 and Grade 2) within the same instructional context [21].

2.2 Context and Setting

The research was conducted at a Czech upper-secondary grammar school in regular Informatics classes supported by Moodle. Students worked with a Moodle-based activity embedded in the course environment, which consisted of a questionnaire created in Google Forms.

2.3 Participants and Sampling

The sample consisted of 110 students (57 Grade 1; 53 Grade 2). The sampling strategy was purposeful; participants represented the full accessible cohort in the two grades within the school. The students were 16–17 years old, and the gender distribution was approximately balanced between boys and girls. While Grade 1 students had no prior experience with Moodle, Grade 2 students had already worked with the platform in previous courses.

2.4 Instrument and Scoring

We assessed intrinsic motivation with the Situational Motivation Scale (SIMS), a 16-item instrument grounded in Self-Determination Theory [19]. The SIMS captures four theoretically distinct types of situational motivation, each represented by a four-item subscale. Intrinsic motivation reflects engaging in an activity for inherent interest and enjoyment, which in this study was assessed through four items asking whether students participate "Because you feel good while doing these activities," "Because you think these activities are pleasant," "Because you find these activities fun," and "Because you find these activities interesting." Identified regulation refers to engaging in the activity because it is personally valued or perceived as important. External regulation captures behaviour driven by external demands, rewards, or pressures. Finally, amotivation reflects a lack of intention or perceived competence, characterized by feelings of disinterest or meaninglessness.

In this study, only the intrinsic motivation subscale was analysed. Responses were provided on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). The intrinsic motivation score was computed as the mean of the four items.

2.5 Procedure

The SIMS items were administered electronically within Moodle. Participation was voluntary and anonymous, and the completion time was approximately 5–10 min. Importantly, Grade 1 students had no prior experience with the Moodle LMS, whereas Grade 2 students had already completed one year of work with the platform, including studying course materials, taking tests, watching instructional videos, and accessing external links.

2.6 Data Analysis

The analysis proceeded in two stages with accompanying diagnostic checks. First, descriptive statistics means and standard deviations (M, SD) were computed separately for Grade 1 and Grade 2 to summarise the distribution of intrinsic motivation scores and to inform subsequent inference.

To test whether group means differed, an independent-samples *t*-test was planned as the primary procedure [22, 23]. Assumptions were examined prior to interpretation. When the homogeneity-of-variance assumption was not met, we applied Welch's correction, which provides more robust control of Type I error under unequal variances and/or

unequal sample sizes [24, 25]. Alongside *p*-values, we reported Cohen's *d* to quantify the practical significance of the between-group difference [26]. The significance threshold was set at $\alpha = .05$, consistent with conventions in educational and psychological research [26].

2.7 Ethical Considerations

The study followed standard ethical principles for school-based survey research. Students were informed about the purpose of the study, voluntary participation, and anonymity of responses. No personally identifying information was collected, and the data were used solely for research purposes.

3 Results

3.1 Descriptive Statistics

Grade 1 students reported higher intrinsic motivation (M = 3.50, SD = 1.43, n = 57) than Grade 2 students (M = 2.97, SD = 1.26, n = 53). Table 1 presents descriptive statistics for both groups.

Table 1. Descriptive statistics of intrinsic motivation (SIMS intrinsic subscale) by grade.

Group	n	M	SD
Grade 1	57	3.50	1.43
Grade 2	53	2.97	1.26

Scale: 1 = strongly disagree; 7 = strongly agree. M and SD refer to the SIMS intrinsic motivation subscale.

3.2 Inferential Statistics

Welch's independent-samples *t*-test indicated that the grade-level difference was statistically significant, $t(107.67) = 2.08$, $p = 0.04$. The effect size was medium (Cohen's $d = 0.40$) (Fig. 1).

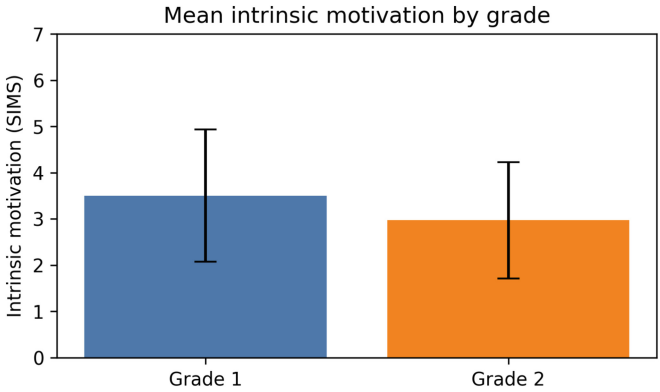


Fig. 1. Mean intrinsic motivation by grade (error bars represent SD; scale 1–7). Grade 1: n = 57; Grade 2: n = 53.

4 Discussion

4.1 Summary of Key Findings

This study investigated whether intrinsic situational motivation during Moodle-based learning in Informatics differs between Grade 1 and Grade 2 upper-secondary students. The results showed a statistically significant difference favouring Grade 1. In e-learning and blended learning contexts, where motivation is closely tied to engagement, persistence, and the quality of task execution, this difference is educationally meaningful [15].

4.2 Interpreting Group Differences Through SDT

Self-Determination Theory (SDT) provides a coherent explanation for the observed pattern. SDT posits that the quality of motivation depends on the satisfaction of three basic psychological needs, namely autonomy, competence, and relatedness, and that autonomous forms of motivation, including intrinsic motivation, are associated with deeper learning and greater persistence [18].

For Grade 1 students, the novelty of the LMS and the initial learning situation can stimulate interest and exploratory engagement while supporting a sense of choice and meaningful involvement. In this grade, students primarily encounter introductory algorithmic concepts and the basics of programming in Python, and when these tasks are presented in a clearly structured and appropriately challenging way, they can foster competence satisfaction and sustain intrinsic motivation.

By contrast, Grade 2 students typically engage with more abstract and cognitively demanding content, including more advanced programming topics and robotics. If scaffolding is insufficient, rising task demands may undermine perceived competence. This, in turn, can reduce intrinsic motivation, even when the technological environment remains constant [17]

4.3 Moodle Affords Design—Not Motivation Per Se

The findings should not be read as a technology effect. Although Moodle provides the necessary infrastructure, the motivational outcomes are ultimately shaped by instructional design and specific contextual factors. In line with SDT, motivation is enhanced when the design of activities strengthens competence through clarity, progressive challenge, worked examples, and formative feedback. It is also supported when learners are provided with meaningful choice and an appropriate degree of autonomy and control [12]. This underscores treating LMS course construction primarily as motivational design, not mere content delivery.

4.4 Methodological Note on Measurement (SIMS)

Using the SIMS instrument is appropriate in this context because it captures situational motivation at the point of performing specific LMS-based activities rather than only general attitudes towards e-learning. Developed within the SDT framework and validated across contexts, SIMS enhances the interpretability of the findings. Focusing on the intrinsic-motivation subscale targets a form of motivation that is particularly relevant to learning quality and sustained engagement in online environments [27].

4.5 Practical Implications for LMS-Supported Informatics

The results translate into concrete guidance for blended learning in secondary education, particularly for higher grades where novelty naturally declines and cognitive demands rise:

- Support autonomy: Offer meaningful options via optional tasks, alternative solution paths, and choice of topic or output format.
- Support competence: Ensure a clear course structure and instructions, calibrate difficulty gradually, provide model materials, and deliver regular formative feedback so that progress is visible.
- Support relatedness: Use discussion forums, collaborative tasks, and peer feedback to counteract isolation and foster a sense of classroom community.

These recommendations align with SDT principles and leverage commonly available Moodle features, thereby strengthening intrinsic motivation in digital learning environments [28].

4.6 Limitations

Several constraints temper the conclusions. The cross-sectional design captures a single moment in students' motivational experience, preventing causal inference and obscuring developmental trajectories over time. The single-school setting limits generalisability to other institutions. Intrinsic motivation was assessed solely via self-report; despite robust psychometrics, such measures reflect subjective perceptions rather than behavioural indicators. Finally, between-grade differences may partly reflect curricular variation that was not controlled in detail.

4.7 Future Research

Future work should adopt longitudinal designs to track motivation across the school year, particularly as novelty fades and task complexity increases. Combining quantitative approaches with qualitative reflections, interviews, or think-aloud protocols would help identify which specific Moodle features most strongly support or hinder the needs for autonomy, competence, and relatedness. Such a mixed-methods perspective would make it possible to determine how particular elements of the platform influence students' motivational experiences. Such evidence can guide more targeted, developmentally responsive LMS design.

5 Conclusion

The study showed that intrinsic situational motivation during Moodle-based Informatics lessons differs between Grade 1 and Grade 2 upper-secondary students at the beginning of the school year. The difference was statistically significant and of medium practical magnitude, favouring Grade 1 ($t(107.67) = 2.08$, $p = .04$, $d = 0.40$). Taken together with the Discussion, these results indicate that learners engage with the same LMS environment in systematically different ways across grade levels, reflecting variations in novelty, task demands, and the satisfaction of basic psychological needs.

Crucially, the findings do not point to a technology effect. Moodle provides the infrastructure, but motivational outcomes are contingent on pedagogical design and contextual conditions. Course designs that draw on Self-Determination Theory are well positioned to sustain intrinsic motivation as curricular complexity increases. They do so by strengthening competence through clear structure, progressive challenge, and formative feedback, and by supporting autonomy through meaningful choice and an appropriate degree of learner control.

Overall, the evidence underscores that thoughtful, SDT-informed course development is central to fostering intrinsic motivation in Moodle-supported Informatics learning. Such design not only facilitates content mastery but also enhances students' engagement, confidence, and long-term interest in the subject.

References

1. Niaz, S., Memon, S., Khokhar, S.: Development of e-learning: a historical review with global perspective. Int. Res. J. Arts Hum. **49**(49) (2021) https://sujo.usindh.edu.pk/index.php/IRJAH/article/view/3817
2. Tibaná-Herrera, G., Fernández-Bajón, M.T., De Moya-Anegón, F.: Categorization of E-learning as an emerging discipline in the world publication system: a bibliometric study in Scopus. Int. J. Educ. Technol. High. Educ. **15**(21), 1–23 (2018). https://doi.org/10.1186/s41239-018-0103-4
3. Eger, L., Egerová, D.: e-Learning trends in Central Europe: the case of the Czech Republic. Knowl. Manag. E-Learn. **5**(3), 375–387 (2013)
4. Hrtonová, N., Kohout, J., Rohlíková, L., Zounek, J.: Factors influencing acceptance of e-learning by teachers in the Czech Republic. Comput. Hum. Behav. **51**, 873–879 (2015)

5. Zormanová, L.: Models of distance learning in higher education. Int. J. Res. E-Learn. IJREL. **3**(1), 49–61 (2017)
6. Pokorný, M., Voštinár, P.: Moderné technológie vo vzdelávaní. Belianum – Vydavateľstvo Univerzity Mateja Bela v Banskej Bystrici (2025) ISBN 978-80-557-2299-3
7. Poulová, P., Černá, M., Simonová, I.: eLearning – 15 years at Czech universities. Int. J. Appl. Math. Inf. **16** (2022). https://doi.org/10.46300/91014.2022.16.4
8. Moodle Academy: Moodle Academy – official learning hub for Moodle educators. https://moodle.academy/ (2025)
9. Gamage, S.H.P.W., Ayres, J.R., Behrend, M.B.: A systematic review on trends in using Moodle for teaching and learning. Int. J. STEM Educ. **9**, 9 (2022). https://doi.org/10.1186/s40594-021-00323-x
10. Al-Fraihat, D., Alshahrani, A.M., Alzaidi, M., Shaikh, A.A., Al-Obeidallah, M., Al-Okaily, M.: Exploring students' perceptions of the design and use of the Moodle learning management system. Comput. Hum. Behav. Rep. **18**, 1–15 (2025)
11. Kordek, D., Kopeček, M.: The use of LMS Moodle and Moodle mobile app in educational process at the Faculty of Medicine in Hradec Králové. MEFANET J. **6**(1), 18–22 (2019)
12. Maněna, V., Maněnová, M., Šín, M., Myška, K.: Moderně s Moodlem: Jak využít e-learning ve svůj prospěch? 1. vydání edn. CZ.NIC (2015)
13. Stal, J., Paliwoda-Pękosz, G.: The possibility of using the VLE moodle to support CLIL. In: Ministr, J., Tvrdíková, M. (eds.) Information Technology for Practice 2013: Proceedings of the 16th International Conference IT Practice 2013, Ostrava, Czech Republic, Frýdek-Místek, VŠB-Technické Univerzita Ostrava, pp. 148–156 (2013) Available at SSRN: https://ssrn.com/abstract=2683565
14. Deci, E.L., Ryan, R.M.: Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press (1985)
15. Vallerand, R.J.: Toward a hierarchical model of intrinsic and extrinsic motivation. Adv. Exp. Soc. Psychol. **29**, 271–360 (1997). https://doi.org/10.1016/S0065-2601(08)60019-2
16. Harandi, S.R.: Effects of E-learning on students' motivation. Procedia Soc. Behav. Sci. **181**, 423–430 (2015). https://doi.org/10.1016/j.sbspro.2015.04.905
17. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. **55**(1), 68–78 (2000). https://doi.org/10.1037/0003-066X.55.1.68
18. Ryan, R.M., Deci, E.L.: Self-determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. Guilford Press (2017). https://doi.org/10.1521/978.14625/28806
19. Guay, F., Vallerand, R.J., Blanchard, C.: On the assessment of situational motivation: the situational motivation scale (SIMS). Motiv. Emot. **24**(3), 175–213 (2000). https://doi.org/10.1023/A:1005614228250
20. Standage, M., Duda, J.L., Ntoumanis, N.: A model of contextual motivation in physical education: using constructs from self-determination and achievement goal theories. J. Educ. Psychol. **95**(1), 97–110 (2003). https://doi.org/10.1037/0022-0663.95.1.97
21. Hunziker, S., Blankenagel, M.: Cross-sectional research design. In: Research Design in Business and Management, pp. 187–199. Springer (2024). https://doi.org/10.1007/978-3-658-34357-6_10
22. Ross, A., Willson, V.L.: Independent Samples T-Test. In: Basic and Advanced Statistical Tests, pp. 31–40. SensePublishers, Rotterdam (2017). https://doi.org/10.1007/978-94-6351-086-8_3
23. Akpan, E.E., Clark, J.L.: Independent t test statistics: It's relevance in educational research. Int. J. Eminent Sch. **10**(1), 79–88 (2023)
24. West, R.M.: Best practice in statistics: use the Welch t-test when testing the difference between two groups. Ann. Clin. Biochem. **58**(4), 267–269 (2021). https://doi.org/10.1177/0004563221992088

25. Dohm, M.R.: Digging Deeper into T Test plus the Welch Test. Statistics LibreTexts (2024)
26. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Routledge (1988). https://doi.org/10.4324/9780203771587
27. Černá, M.: Deployment of cognitive and affective determinants in blended learning - case study. In: Cheung, S., Kwok, L., Ma, W., Lee, L.K., Yang, H. (eds.) Blended Learning. New Challenges and Innovative Practices. ICBL 2017 Lecture Notes in Computer Science, vol. 10309. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59360-9_41
28. Yahiaoui, F. et al.: The impact of e-learning systems on motivating students and enhancing their outcomes during COVID-19: a mixed-method approach. Front. Psychol. **13**, Article 874181 (2022). https://doi.org/10.3389/fpsyg.2022.874181

Reading and Writing as Enabling Practices in Technology-Enhanced Learning for Students with Special Educational Needs

Maja Kerneža[1,2](✉), Maja Vičič Krabonja[3], and Andrej Flogie[2]

[1] Faculty of Education, University of Maribor, Koroška street 160, 2000 Maribor, Slovenia
maja.kerneza1@um.si

[2] Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška street 160, 2000 Maribor, Slovenia

[3] Secondary School of Economics and Grammar School Maribor, Trg Borisa Kidriča 3, 2000 Maribor, Slovenia

Abstract. Reading and writing are integral but often implicit components of technology-enhanced learning scenarios for students with special educational needs within adapted and special educational programmes. Although digital technologies are frequently discussed in terms of innovation and learner autonomy, less attention is paid to the linguistic conditions that enable meaningful participation in technology-supported activities. The aim of the study is to explore how reading and writing are didactically embedded in technology-enhanced learning scenarios and what roles they play within these contexts. The study is based on a qualitative document analysis of teaching scenarios developed by educational professionals across different educational programmes and subject areas for students with special educational needs. The analysis focused on identifying how reading and writing activities are planned and how they function in relation to the use of digital technologies. The results show that reading and writing are consistently present as functional and supportive practices rather than as explicit learning objectives. Reading primarily supports understanding and orientation within activities, while writing appears as a structured and limited response enabling task completion. These findings highlight the importance of reading and writing as key enabling practices in inclusive, technology-enhanced learning environments.

Keywords: digital competence · functional literacy · special education · teaching scenarios · technology-enhanced learning

1 Introduction

Digital technologies are frequently presented in contemporary education as tools that promote learner autonomy, individualisation, and active participation. Discussions of educational technology often focus on platforms, applications, and innovative instructional formats, while the didactic conditions that make participation in technology-enhanced activities possible receive less attention. This issue becomes particularly salient in educational contexts involving students with special educational needs, where the ability to

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 163–174, 2027.
https://doi.org/10.1007/978-3-032-32104-6_15

act independently, understand instructions, and navigate digital environments is often limited and closely dependent on structured pedagogical support. Research in special education consistently shows that the effective use of digital technology depends not only on access to tools but also on teacher competence, pedagogical design, and the characteristics of learners themselves [1].

Within the Slovenian context, students with special educational needs are formally defined by the Placement of Children with Special Needs Act (ZUOPP-1), which includes students with intellectual disabilities, visual impairments, hearing impairments, speech and language disorders, physical disabilities, chronic illnesses, specific learning difficulties, autism spectrum disorders, and emotional and behavioural disorders. These students are educated in adapted educational programmes, including programmes with a lower educational standard and special education programmes, where learning goals, content, and instructional approaches are systematically adjusted to their developmental and educational characteristics.

Reading and writing inevitably emerge within technology-enhanced learning activities in these contexts. Learners are expected to read short instructions, interpret prompts, recognise symbols, and enter written responses in order to complete tasks. Although these practices are essential for participation, they are rarely defined as explicit learning objectives. Instead, they function as enabling practices that support orientation, understanding, and response within digitally structured activities. As such, reading and writing operate as functional literacy practices embedded within digital participation rather than as separate instructional targets.

Existing research on technology-enhanced learning for students with special educational needs has primarily focused on the effectiveness of digital tools, accessibility features, and measurable learning outcomes. Reviews and meta-analyses have examined the impact of digital interventions on specific reading skills such as fluency [2, 3], while intervention studies on writing have explored technologies such as speech-to-text as supports for written production [4]. Although these studies provide valuable insights into technological interventions, less attention has been given to the didactic conditions that enable learners to participate in technology-enhanced activities in the first place. In particular, the role of reading and writing as embedded conditions within digitally structured learning scenarios remains underexplored.

Empirical findings from Slovenia further indicate that the development of digital competence in adapted and special education programmes primarily focuses on foundational digital skills such as operating devices, accessing information, and ensuring digital safety, while more advanced forms of digital creation and specialised assistive tools are less frequently implemented [5]. In practice, teachers often rely on general-purpose digital tools and develop instructional adaptations independently, which highlights the central role of pedagogical design in shaping learners' opportunities for digital participation.

This study is based on the assumption that teaching scenarios function as didactic documents that reflect how educators conceptualise technology-enhanced learning and embed enabling practices such as reading and writing within instructional design. The

aim of this study is to examine how reading and writing are embedded in technology-enhanced learning scenarios for students with special educational needs and to analyse their didactic function within these contexts.

Accordingly, the study addresses the following research questions:

1. How are reading and writing activities embedded within technology-enhanced learning scenarios in adapted and special education programmes?
2. What functions does reading perform in enabling orientation, understanding, and participation in digitally structured activities?
3. What functions does writing perform in task completion, digital content creation, and digital communication?
4. How are reading and writing practices supported through didactic scaffolding and multimodal instructional strategies?

By focusing on didactic planning rather than technological effectiveness, the study complements existing intervention-oriented research and contributes to a deeper understanding of how literacy practices enable participation in digitally mediated learning environments.

The study was conducted within the broader framework of the DigComp PP project, which investigates the development of digital competence in adapted educational programmes and provides insight into instructional practices in these contexts [5].

2 Theoretical Framework

2.1 Technology-Enhanced Learning as a Didactically Structured Environment

Technology-enhanced learning is often associated with learner autonomy, flexibility, and self-regulated learning. Digital environments are commonly understood as tools that enable learners to search for information, construct knowledge, and manage their learning independently. However, this perspective assumes that learners possess sufficient cognitive, linguistic, and executive functioning skills to engage with digital environments without substantial support.

In special education contexts, these assumptions often do not hold. Digital technology does not function as an independent learning agent but as part of a structured pedagogical environment requiring clear sequencing, instructional scaffolding, and continuous teacher guidance. Rather than reducing the role of the teacher, digital environments often increase the need for pedagogical mediation, as learners depend on explicit instruction to participate meaningfully.

Empirical findings support this interpretation. Börnert-Ringleb et al. [1] showed that digital learning in special education is strongly influenced by teacher competence, institutional support, and learner characteristics such as self-regulation. Similarly, research conducted in Slovenia indicates that teachers primarily use general digital tools such as interactive whiteboards and tablets, while specialised assistive technologies remain less common [5]. These findings reinforce the understanding of digital learning as a didactically structured environment rather than an autonomous technological process.

The European Digital Competence Framework DigComp 2.2 provides an important conceptual reference point in this context [6]. DigComp defines digital competence as a combination of knowledge, skills, and attitudes required to use digital technologies effectively. Core competence areas include information search, digital communication, and content creation. These activities inherently involve reading and writing, as participation in digital environments requires understanding instructions, interpreting information, and producing textual responses. Digital competence is therefore inseparable from literacy practices.

2.2 Reading and Writing as Enabling Literacy Practices

Within digitally structured learning environments, reading and writing function as enabling literacy practices that support participation rather than as isolated instructional objectives. Functional literacy refers to the use of language in meaningful contexts, where reading and writing serve as tools for orientation, understanding, and action.

Students must read instructions, interpret digital interfaces, recognise symbols, and enter written responses to complete tasks. These practices are typically simplified, structured, and supported through visual or procedural scaffolding. Research on assistive technologies has demonstrated that text-to-speech tools can improve reading comprehension for students with intellectual disabilities [7], while speech-to-text technologies can support written production and reduce barriers associated with handwriting [4, 8].

In adapted educational programmes, digital competence development is closely linked to foundational forms of participation, such as navigating interfaces, understanding prompts, and responding to digital tasks [5]. These activities rely on literacy practices even when literacy itself is not defined as the primary instructional focus.

Reading and writing thus function as enabling mechanisms that make participation in digital learning possible.

2.3 Teaching Scenarios as Representations of Instructional Design

Teaching scenarios represent structured descriptions of learning activities and reflect instructional decision-making. In research contexts, they can be analysed as didactic artefacts that reveal how educators conceptualise learning, technology, and support [9].

Scenarios developed within competence-based frameworks such as DigComp 2.2 include defined learning objectives, activity sequences, and expected learner responses. This structure makes them suitable for analysing embedded literacy practices and understanding how reading and writing function within technology-enhanced learning environments.

Document analysis of teaching scenarios provides insight into instructional planning, even though it does not capture actual classroom implementation.

3 Methodology

3.1 Research Design

The study employed a qualitative document analysis of teaching scenarios as artefacts of instructional design. The aim was to identify and interpret how reading and writing practices were embedded within technology-enhanced learning activities. The analysis focused on the didactic function of literacy practices rather than their frequency or measurable outcomes.

3.2 Materials

The dataset consisted of 62 teaching scenarios developed within the DigComp PP project during the 2024 to 2025 school year. Rather than representing a sample, the corpus included all teaching scenarios submitted as part of the project by educational professionals working across mixed subject areas in adapted and special education programmes. Within the project framework, the preparation of a teaching scenario formed part of the participants' professional engagement and was therefore a required component of participation.

All scenarios were expected to align the selected DigComp 2.2 competence area with curricular content, instructional goals, and planned learning activities. The scenarios were developed across mixed subject areas and educational contexts, including programmes with a lower educational standard and special education programmes. In addition to the descriptions of instructional planning, the scenario template also included sections for teacher reflection and, in many cases, learner reflection. Although these reflective elements do not constitute independent interview data, they provided additional insights into perceived difficulties, learner responses, and the pedagogical reasoning surrounding implementation.

3.3 Analytical Procedure

The analysis was conducted in three phases. First, all teaching scenarios were read in full in order to identify segments in which reading and writing were mentioned or implicitly required within the description of learning activities, instructions, digital tool use, assessment criteria, and reflective sections. Second, these segments were coded according to their didactic function, with attention to how reading and writing supported orientation, comprehension, task completion, digital content creation, communication, and participation. Third, the codes were compared across the corpus and grouped into broader thematic categories that captured recurring patterns in the role of literacy practices in technology-enhanced learning. The coding framework focused on the didactic functions of literacy, particularly reading and writing for orientation, comprehension, task completion, digital content creation, communication, and supported participation.

The coding process was carried out by multiple researchers. Initial coding was followed by comparison of codes, discussion of differences in interpretation, and refinement of category definitions until a shared analytical framework was established. This comparative procedure increased the consistency of the analysis and supported the development

of conceptually robust thematic categories. Throughout the analysis, the focus remained on literacy as an enabling condition of participation in digitally mediated learning rather than on frequency counts or formal curricular status.

The study therefore analysed teaching scenarios primarily as documents of planned instructional design. At the same time, because the scenario template included teacher reflection and learner reflection, the corpus also offered limited insight into how planned activities were perceived during implementation. These reflective elements were treated as part of the document corpus, not as stand-alone interview material.

4 Results

The collected teaching scenarios form a coherent empirical corpus that enables analysis of how digital competences are enacted in adapted and special educational programmes, including adapted programmes with a lower educational standard and special education programmes, as well as in connection with subject areas such as language education, social studies, general knowledge, social learning, visual arts, and digital literacy instruction. The scenarios primarily address DigComp 2.2 components 1.1 browsing, searching, and filtering, and 3.1 developing digital content, while digital communication and sharing (2.1 and 2.2) appear less frequently.

Overall, digital competence emerges as an interplay of technical procedures, literacy processes, and didactic support. Progress depends less on the presence of digital tools than on conditions of accessibility, the level of learner independence, and the quality of scaffolding, including structured guidance, modelling, and the use of visual and procedural supports.

4.1 Reading as an Operational Condition for Digital Activity

Across scenarios targeting browsing, searching, and filtering, digital activity is consistently mediated by literacy. Tasks such as searching for information about authors, identifying characteristics of topics, or exploring subject-specific content require the formulation or recognition of keywords, interpretation of search results, and selection of relevant information. These processes unfold within digital environments where meaning is conveyed through combinations of text, icons, menus, and images.

In such activities, participation depends not only on technical operation but on the ability to interpret short instructions, navigate interface elements, and evaluate the relevance of information. Even when image-based search is foregrounded, reading remains embedded in recognising categories, understanding labels, and orienting within the digital environment. The difficulty is also explicitly reflected in teacher observations embedded in the scenarios, where learners were described as “struggling to formulate appropriate keywords” and to distinguish relevant from irrelevant search results, particularly when confronted with a large amount of textual information that was “too demanding and not adapted to their reading abilities”. In some cases, teachers noted that learners relied on visual cues, as written content often exceeded their reading abilities.

A recurring pattern across scenarios is that difficulties arise less from handling devices and more from interpreting information. Challenges include understanding key

terms, maintaining focus in environments with multiple alternatives, and identifying relevant content among competing results. Digital activity thus becomes an entry point into early forms of critical literacy, where selecting, refining, and evaluating information are necessary conditions for task completion.

4.2 Writing as Both a Constraint and a Developmental Mechanism

Writing appears as a structurally central component of digital activity, taking two primary forms. Functional writing includes entering search queries, login information, file names, and short textual inputs required for navigation and execution. Compositional writing involves producing short texts, organizing information, and constructing digital artefacts such as presentations, descriptions, or simple formal messages.

Many scenarios combine language-related objectives with digital production, for example in tasks where learners write structured descriptions, compose short texts in word-processing programmes, or format documents according to given criteria. These activities require the integration of orthographic knowledge with procedural operations such as text selection, formatting, saving, and file naming.

Functional writing frequently represents a point of difficulty. Inaccurate input leads to unsuccessful searches or procedural errors, which can interrupt task progression. At the same time, these constraints support development. When writing is structured into manageable steps and supported through prompts, modelling, and visual cues, it becomes a mechanism for learning, enabling the gradual acquisition of both digital vocabulary and procedural understanding.

In more advanced activities, writing extends beyond isolated input towards meaning-making. Learners construct messages, combine textual and visual elements, and participate as authors within digital environments, although this process remains closely dependent on instructional support.

This is particularly evident in scenarios where learners copied or produced short texts in a word-processing programme, combining orthographic requirements with digital operations. Teacher reflections indicate that learners often required substantial support in using input devices and writing accurately, as they "required substantial support and guidance", with difficulties arising "due to limited writing abilities", which directly affected task completion.

4.3 Multimodality as a Condition of Accessibility

Digital activity in the analysed scenarios is consistently realized through multimodal engagement. Learners work with combinations of images, short written text, symbols, sound, and video, while extended written text appears less frequently. Multimodality functions primarily as a condition of accessibility rather than a supplementary feature.

Participation is often enabled through visual or auditory channels, such as searching for images, working with video materials, or expressing meaning through graphical elements and symbols. Activities in which learners produce digital artefacts, such as visually supported presentations or simple animated stories, illustrate how meaning is constructed through the integration of multiple modes. Learner reflections embedded in

the scenarios further suggest that visual elements were often perceived as more accessible and engaging than written text, particularly in contexts "where they do not need to read extended text", and where learners "enthusiastically explored the functions of the programme."

Multimodality also operates as a compensatory strategy in situations where reading presents a barrier. However, the scenarios indicate that effective learning does not rely on replacing text with images, but on structuring relationships between modes. Visual content is most productive when linked to naming, explanation, and integration into a coherent product, thereby supporting rather than bypassing literacy engagement.

4.4 Didactic Scaffolding as a Condition for Developing Digital Competence

Didactic scaffolding emerges as a central condition for participation and competence development. Four recurring forms of scaffolding can be identified: segmentation of tasks into clearly defined steps, modelling of procedures and expected outcomes, use of visual supports such as icons, screenshots, and prompts, and continuous individualised guidance. The need for such support is evident in scenarios where learners relied on visual aids for tasks such as logging into digital environments, using keyboard functions, or identifying interface elements. Teacher reflections indicate that without such structured support, learners would not be able to complete even basic procedural steps independently.

These forms of support are embedded in activities that require learners to navigate digital environments, use specific functions, or combine multiple elements into a coherent output. Visual aids, model texts, and step-by-step instructions enable learners to perform actions such as accessing digital tools, applying functions, or completing multi-step tasks that would otherwise exceed their level of independent functioning. In some cases, learners completed tasks "with the help of a password template" or relied on "visual keyboard support", while difficulties were also observed in basic operations such as "appropriate mouse handling (double-click)" and acquiring "vocabulary related to digital language."

Monitoring is closely integrated into instruction and functions as an ongoing diagnostic process. Teachers identify whether difficulties are related to motor skills, conceptual understanding, language comprehension, attention regulation, or task interpretation, and adjust support accordingly. Digital competence thus develops through guided participation rather than through exposure to technology alone.

4.5 Development of Learner Independence

The scenarios reveal a clear developmental progression in learner independence. Initial stages are characterised by the successful execution of individual procedural steps, such as opening applications, entering input, selecting elements, and saving outputs. At later stages, learners engage in more strategic actions, including refining search queries, evaluating information, selecting relevant content, and integrating digital processes into broader tasks.

Independence is not understood as the absence of support, but as the ability to act within a structured environment where expectations are clear and assistance remains

available. Teacher reflections indicate that structured guidance supports sustained attention and reduces distraction, while less structured environments increase demands on self-regulation. The development of independence is therefore closely linked to the balance between guidance and autonomy.

Teacher reflections further indicate that independence develops gradually and remains closely tied to the availability of support, as learners often require guidance when encountering unfamiliar tasks or when attention is disrupted by competing digital stimuli; in such cases, it was noted that learners "will need a great deal of practice" and that digital environments can easily "distract them from the task."

4.6 Digital Communication and Sharing

Digital communication appears as both a motivating and demanding component of digital competence. Activities such as sending messages, sharing files, or presenting digital products introduce additional procedural and literacy demands, including entering addresses, naming files, and following sequences of actions required for successful transmission.

When embedded within meaningful tasks, communication supports participation and engagement. Learners experience the sharing of digital products as purposeful and socially relevant, which contributes to their sense of competence. At the same time, these activities highlight the complexity of seemingly simple actions, as successful communication depends on the coordination of multiple literacy and procedural elements.

In scenarios involving sharing digital products, teachers observed that tasks such as sending an e-mail or attaching a file required careful step-by-step guidance, as errors in entering addresses or naming files frequently disrupted the process despite learners' high motivation to share their work; in some cases, teachers explicitly noted that "we had difficulties saving the recording."

5 Discussion

The findings of this study show that reading and writing are consistently present in technology-enhanced learning scenarios within adapted educational programmes with a lower educational standard and special education programmes. However, they are rarely defined as explicit didactic objectives. Instead, their primary role is to enable participation in digitally structured learning activities. In this sense, literacy functions as an enabling condition of digital participation rather than as a separate learning outcome. This role aligns with the concept of functional literacy, where language serves as a means for action within meaningful contexts rather than as an isolated instructional goal. At the same time, this pattern reflects the inherent structure of digital environments, which are fundamentally text- and symbol-based and therefore place implicit literacy demands on learners.

The analysis further shows that digital activities such as searching for information, creating presentations, composing short texts, and sharing digital products cannot be carried out without fundamental literacy processes. This finding extends previous research

within the DigComp PP project [5] by demonstrating that literacy is not only present alongside digital competence, but constitutes the underlying mechanism that enables activities to take place. Digital competence development in adapted programmes can therefore be more accurately understood as literacy-mediated participation in digitally structured environments.

Reading emerges as a particularly critical condition for orientation and participation in digital environments. Learners must interpret short instructions, recognise interface elements, understand search results, and evaluate their relevance. The findings suggest that difficulties in digital tasks are often rooted not in technical operation, but in challenges related to comprehension and meaning-making. This reinforces the view that digital competence cannot be reduced to technical skills but must be conceptualized as a literacy-mediated practice. Such an interpretation is consistent with studies highlighting the importance of structured digital texts and assistive reading technologies in facilitating access to written information for learners with special educational needs [10, 11].

Writing plays a similarly important, although more constrained, role. While writing is often limited to structured inputs and short responses, it also represents an important developmental space. Digital environments enable learners to construct messages, organize ideas, and participate as authors, even within highly scaffolded contexts. This dual role, as both constraint and opportunity, highlights the importance of carefully designed support. These findings resonate with intervention studies showing that assistive technologies such as speech-to-text can support written expression and increase participation among learners with special educational needs [4, 8].

Another important dimension concerns the relationship between literacy and multimodality. Multimodal design enhances accessibility and enables participation by reducing reliance on extended written text. However, the findings indicate that multimodality is most effective when it supports, rather than replaces, literacy practices. This suggest that the didactic challenge lies not in simplifying content by removing text, but in designing meaningful connections between visual, textual, and symbolic modes.

The results also demonstrate that digital competence does not emerge spontaneously through exposure to technology. Instead, it develops through carefully structured didactic support. Teaching scenarios consistently include segmentation of tasks, modelling, visual supports, and individualised guidance, which enables learners to gradually develop independence. This confirms existing research showing that the effectiveness of digital learning in special education depends not on technology alone, but on pedagogical design and support structures that take learner characteristics into account [1].

Taken together, the findings suggest that digital technologies do not replace literacy practices but restructure and mediate them. Reading and writing function as foundational mechanisms that enable participation in digital activities, while digital environments reshape the forms and conditions under which literacy is enacted. Digital competence therefore emerges not as a purely technical ability, but as an integrated configuration of linguistic, cognitive, and pedagogical processes.

By focusing on teaching scenarios as documents of didactic planning, this study contributes a perspective that complements the predominantly intervention-oriented research [2–4]. Rather than examining the effects of specific technologies, it highlights how literacy demands are embedded within the design of digital learning activities themselves.

Understanding reading and writing as enabling practices thus represents an important foundation for the design of inclusive technology-enhanced learning environments.

A limitation of the study is that the analysis is based primarily on teaching scenarios as documents of planned instructional design. The corpus therefore provides insight into how educational professionals structured literacy-supported digital participation, but not into classroom interaction in the full sense. Although many scenarios included teacher reflection and short learner reflections, these remained embedded within the written scenario template and did not provide the depth of interview or observational data. Future research should therefore complement document analysis with teacher interviews, classroom observation, and learner perspectives in order to examine how these literacy practices function in enacted teaching and learning.

6 Conclusions

This study examined how reading and writing are embedded in technology-enhanced learning scenarios for students with special educational needs from the perspective of didactic planning. The analysis showed that literacy practices are consistently present but rarely formulated as explicit learning objectives. Instead, reading and writing function as enabling conditions that allow learners to understand instructions, navigate digital environments, and participate in digitally mediated activities. Digital competence therefore does not emerge independently of literacy but is fundamentally mediated by it.

The findings further demonstrate that digital competence in adapted educational programmes develops through the interaction of literacy practices, structured pedagogical support, and digital tools. Reading enables learners to interpret digital information and orient themselves within tasks, while writing allows them to respond, create content, and participate in digital communication. Technology does not replace these processes but reshapes their form and conditions. As a result, digital competence should be understood not as a purely technical ability, but as a literacy mediated competence that depends on didactic structuring and support.

These findings have important implications for both research and practice. They highlight the need to recognise literacy as a central component of digital competence development and to address reading and writing explicitly within the design of technology-enhanced learning environments. At the same time, the study shows that access to digital learning depends not only on the availability of technology but on pedagogical conditions that enable meaningful participation. Future research should extend this work by examining how literacy practices function in classroom implementation and how learners experience and develop digital competence in authentic educational settings.

Acknowledgments. The publication of this article is co-funded by the Ministry of Education and the European Union – NextGeneration EU, within the DigComp PP project (Co-funding contract no.: C3350-24-927006).

References

1. Börnert-Ringleb, M., Casale, G., Hillenbrand, C.: What predicts teachers' use of digital learning in Germany? Examining the obstacles and conditions of digital learning in special education. Eur. J. Spec. Needs Educ. **36**(1), 80–97 (2021)
2. Alqahtani, S.S.: Technology-based interventions for children with reading difficulties: a literature review from 2010 to 2020. Educ. Tech. Res. Dev. **68**, 3495–3525 (2020)
3. Mize, M., Park, Y., Martin, M.: Technology-assisted reading fluency interventions for students with reading difficulties: evidence from a meta-analytic approach of single case design studies. Disabil. Rehabil. Assist. Technol. **18**(8), 1544–1554 (2023)
4. Kambouri, M., Simon, H., Brooks, G.: Using speech-to-text technology to empower young writers with special educational needs. Res. Dev. Disabil. **135**, 104466 (2023)
5. Kerneža, M., Čampelj, B., Flogie, A., Vičič Krabonja, M., Zemljak, D.: Digital competence in learners with disabilities. In: Šumak, B., Pušnik, M., Grašič, S. (eds.) Digital Inclusion in Information Society. Proceedings of the 28th International Multiconference Information Society – IS 2025, vol. H, pp. 7–11. Institut Jožef Štefan, Ljubljana (2025)
6. Vuorikari, R., Kluzer, S., Punie, Y.: DigComp 2.2: The Digital Competence Framework for Citizens: With New Examples of Knowledge, Skills and Attitudes. Publications Office of the European Union, Luxembourg (2022)
7. Svensson, I., Nilsson, S., Fälth, L., Selenius, H., Sand, C.: Text-to-speech as reading aid for students with intellectual disabilities: a single-subject design. J. Spec. Educ. Technol. (2025)
8. Sand, C., Svensson, I., Nilsson, S., Selenius, H., Fälth, L.: Speech-to-text intervention to support text production for students with intellectual disabilities. Disabil. Rehabil. Assist. Technol. **20**(2), 408–415 (2025)
9. Laurillard, D.: Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology. Routledge, London (2012)
10. Svensson, I., et al.: Effects of assistive technology for students with reading and writing disabilities. Disabil. Rehabil. Assist. Technol. **16**(2), 196–208 (2021)
11. Alison, C., Root, J.R., Browder, D.M., Wood, L.: Technology-based shared story reading for students with autism who are English-language learners. J. Spec. Educ. Technol. **32**(2), 91–101 (2017)

AI Literacy of Teachers in the Czech Republic in the Context of DigComp 3.0

Kristyna Kilianova, Petra Kockova, and Katerina Kostolanyova(✉)

University of Ostrava, Dvorakova 7, 701 03 Ostrava, Czech Republic
{kristyna.kilianova,petra.kockova,katerina.kostolanyova}@osu.cz

Abstract. This study examines the level of AI literacy among teachers in the Czech Republic in the context of DigComp 3.0. With the rapid expansion of generative artificial intelligence (GenAI), educators increasingly need to use AI tools effectively, evaluate outputs critically, and apply them ethically. The study aimed to assess GenAI adoption in school practice, teachers' self-confidence in key competence areas, and barriers to AI integration.

Data were collected between May 2025 and January 2026 through an online questionnaire survey of 2,153 pre-primary, primary and secondary schools' teachers. Results show that 58.4% of respondents use GenAI tools at least occasionally, mainly for preparing teaching materials and administrative documents. ANOVA confirmed significant differences in usage depending on teaching experience and school type, with the highest adoption among early-career teachers and secondary school educators.

Teachers reported the highest confidence in working with digital resources and content creation, and the lowest in assessing AI-supported student outputs and ensuring academic integrity. The most common barriers were lack of methodological support, legal uncertainty, and ethical concerns. The study concludes with recommendations for educational policy focusing on strengthening critical AI evaluation skills and supporting experienced teachers in the digital transformation process.

Keywords: AI literacy · Artificial Intelligence · DigComp 3.0 · Generative AI

1 Introduction

Artificial intelligence has become a driving force of digital transformation in recent years. Its influence has long exceeded the boundaries of industry, healthcare, or public administration and is increasingly penetrating the field of education. Education systems therefore face a unique situation: on the one hand, AI offers powerful tools for personalized learning and reduces teachers' administrative workload; on the other hand, it raises complex issues related to data security, ethics, and the ability to critically evaluate information.

With the emergence of generative AI and its easy accessibility, these technologies have become a common part of everyday life for pupils and students. In this context,

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 175–186, 2027.
https://doi.org/10.1007/978-3-032-32104-6_16

the need to systematically develop competencies enabling individuals not only to use AI, but also to understand it, critically evaluate its outputs, and reflect on its impact on individuals and society is growing. In academic literature, this set of knowledge, skills, and attitudes is increasingly referred to as AI literacy. We understand AI literacy as "*a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace*" [1].

The importance of these skills is also confirmed by the European Commission in the fifth edition of the European Digital Competence Framework, also known as DigComp 3.0. This framework responds to current technological trends and emphasizes areas that are gaining significance with the rise of AI. Whether it concerns cybersecurity, digital rights, digital wellbeing, or combating disinformation, DigComp 3.0 integrates AI throughout the framework and calls for a critical and responsible approach to technologies in school, work, and everyday life [2].

The growing importance of AI literacy is further highlighted by European strategic and legislative documents, such as the Artificial Intelligence Act (AI Act). This regulation requires that people working with high-risk AI systems have an appropriate level of knowledge [3]. The development of AI literacy concerns not only students, but especially teachers. Teachers play a key role as guides who help learners manage new technologies, although they often must respond to rapid developments without clear methodological guidance.

Inspiration can also be drawn from international models, such as the **AI Literacy Framework** developed by Digital Promise, which states that "*AI literacy is critical to safely and effectively participate in an increasingly digital world. Our expanded AI Literacy Framework depicts the relationship between AI Literacy Practices, Core Values, Modes of Engagement, and Types of Use that support the knowledge and skills to enable people to critically understand, evaluate, and use AI systems and tools*" [4] (Fig. 1).

Fig. 1. Expanded AI Literacy Framework [4]

A forthcoming joint project of the European Commission and the OECD aims to unify the definition of AI literacy for primary and secondary education [5].

In the Czech environment, digital technologies have been incorporated into curricula mainly through the revision of the Framework Educational Programme in informatics. However, the extent to which Czech teachers are prepared to integrate AI into teaching remains unclear. Therefore, the aim of this study is to map the level of AI literacy among teachers in the Czech Republic and identify the main challenges associated with implementing AI in education.

1.1 Current State of Teachers' AI Literacy in the Czech Republic (Data 2024–2025)

The approach of Czech education to generative artificial intelligence (GenAI) significantly changed in 2024 and 2025. Schools moved from initial spontaneous experimentation toward more systematic work. A key milestone was the publication of the Ministry of Education's methodological recommendation on the use of AI in schools [6], followed in 2024 by the National Pedagogical Institute (NPI ČR), which provided practical guidance aligned with the DigCompEdu framework.

A 2024 survey conducted by the initiative AI dětem and the Scio company showed that approximately 30–35% of Czech teachers had tried a GenAI tool (most commonly ChatGPT, Gemini, or Claude). Teachers primarily use AI for:

- Designing curricula and worksheets.
- Preparing tasks for group work.
- Translating and simplifying foreign-language texts.

A significant role in awareness-raising and professional development is played by the non-profit sector, for example through the AI Accelerator for Schools program [7]. This program helps schools establish rules for AI use, create ethical codes, and train teaching staff in effective prompting of AI tools.

However, the Czech School Inspectorate (ČŠI) notes in its thematic report (2024) that the level of AI literacy differs substantially among teachers. While teachers are improving in working with digital resources (Area 2 of DigCompEdu), uncertainty remains in student assessment (Area 4 of DigCompEdu) [8]. The main challenge continues to be how to objectively assess assignments produced with AI assistance and how to ensure originality and academic honesty in the era of generative models.

2 Methodology

2.1 Research Aim and Research Questions

The main objective of the research was to evaluate the level of AI literacy among teachers in the Czech Republic. For this purpose, a questionnaire survey was conducted, based on the DigComp 2.2 and DigCompEdu frameworks, while the collected data were interpreted in the context of the new DigComp 3.0 framework. The specific objectives were defined as follows:

- To identify the areas in which teachers feel least confident when working with AI.

- To analyze the relationship between teaching experience, school type, and the frequency of AI tool use.
- To identify the main obstacles preventing effective integration of AI into teaching.

The research was guided by the following research questions:

- RQ1: To what extent do teachers at different educational levels use generative AI tools?
- RQ2: In which DigComp competence areas do teachers feel most confident, and where do they perceive the greatest gaps?
- RQ3: What barriers do teachers most frequently mention in relation to implementing AI in teaching?

2.2 Research Design

The research was designed as a quantitative online questionnaire survey conducted between May 2025 and January 2026. This period allowed the study to capture the situation at a time when schools already had access to initial methodological guidance (from MŠMT, NPI ČR, or the AI dětem initiative) and teachers had gained more experience with generative tools. Therefore, the results reflect not only first impressions, but also the phase of gradual adaptation of schools to new technologies.

The study represents an extended cross-sectional survey mapping the level of AI literacy among teachers in the Czech Republic during 2025. Data were analyzed both overall and with respect to the time of data collection (spring vs. autumn phase), enabling the identification of possible developmental trends.

2.3 Timing of Data Collection in Relation to DigComp 3.0

Data collection took place during the transition to the updated DigComp 3.0 framework. The questionnaire was therefore based primarily on DigCompEdu and DigComp 2.2, which were the valid standards at the time. DigComp 3.0 was used for the interpretation of the findings, allowing the results to be situated within a broader context of AI-related competencies across education. This shift is illustrated in Table 1.

Table 1. Comparison of Selected Digital Competencies from the Perspective of DigComp 2.2 and 3.0

Competence Area	DigComp 2.2 (traditional approach)	DigComp 3.0 (AI perspective)
Information and data literacy	Searching and filtering data and digital content	Critical evaluation of AI-generated answers and detection of manipulation
Communication and collaboration	Sharing data and collaborating through technologies	Effective interaction with AI (prompting) and understanding algorithmic principles

(*continued*)

Table 1. (*continued*)

Competence Area	DigComp 2.2 (traditional approach)	DigComp 3.0 (AI perspective)
Digital content creation	Creating and editing content in various formats	Co-creation with AI and addressing copyright issues in AI-generated outputs
Safety	Protecting devices, data, and digital identity	Protecting privacy when entering data into AI models and understanding AI-related risks

2.4 Research Sample

The target group consisted of teachers from pre-primary, primary, and secondary schools in the Czech Republic. Between May 2025 and January 2026, the online questionnaire collected a total of 2,438 initiated responses. After removing incomplete or duplicate entries, the final analysis included N = 2,153 valid responses.

The sample was relatively balanced across school types (pre-primary 18.6%, primary lower 29.8%, primary upper 30.9%, secondary 20.7%). The average teaching experience was 14.7 years (SD = 8.8), with representation across all career stages (up to 5 years 19.4%, 6–15 years 34.7%, 16–25 years 27.3%, over 25 years 18.6%) (Fig. 2).

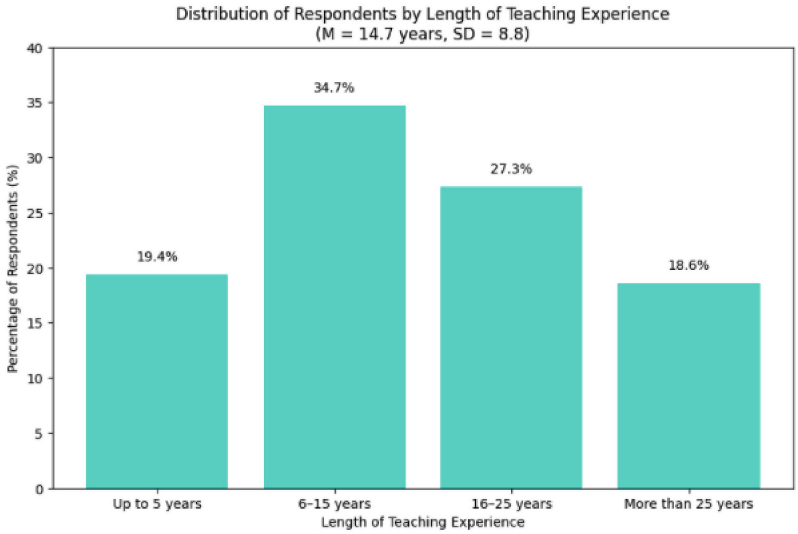


Fig. 2. Distribution of Respondents by Length of Teaching Experience

Due to the nature of data collection, the sampling method was non-probabilistic and based on self-selection. The questionnaire was distributed nationwide through professional networks, methodological centers, and online communities focused on digital education. The high number of respondents was achieved through long-term data collection and the involvement of teachers across all regions of the Czech Republic.

2.5 Research Tool

A structured online questionnaire was developed for data collection. Its design was based on the DigComp 2.2 framework [9] and the DigCompEdu framework for educators [10], which at the time defined European standards for digital competence.

When developing individual items, inspiration was drawn from international AI literacy models (e.g., the Digital Promise AI Literacy Framework) and from current methodological recommendations by MŠMT and NPI ČR regarding the use of generative AI in schools. The questionnaire consisted of four main sections:

1. **Demographic Data** – This section collected basic respondents' characteristics, including school type, teaching experience, taught subjects, and region of employment.
2. **Level of AI Use** – Teachers reported how often they use tools such as ChatGPT, Gemini, or Copilot on a five-point Likert scale (1 = never to 5 = very often). The questionnaire also explored specific purposes of AI use, including material preparation, differentiation, task design, assessment, and administration.
3. **Self-Assessment of AI Competencies** – This key section followed the structure of five DigComp 2.2 competence areas: Information and Data literacy, Communication and Collaboration, Digital Content Creation, Safety, Problem Solving

Items were adapted to reflect the specifics of working with AI, with a particular focus on critical evaluation of AI outputs, personal data protection, and ethical content creation. Respondents rated their level in each area on a scale from 1 to 5 (very low to very high).

The questionnaire consisted of several sections focusing on key areas, including the use of generative AI teaching, self-assessed AI competencies, and attitudes toward AI in education. The items were measured using a Likert-type scale. Each section included multiple items designed to capture specific aspects of teachers' experience with AI. For example, respondents evaluated statements such as *"I am able to critically assess AI-generated outputs"* or *"I use AI tools to support my teaching practice"*. The full instrument is available from the authors upon request.

2.6 Barriers to AI Implementation

The final section investigated the most significant barriers perceived by teachers (e.g., lack of methodology, ethical concerns, legal uncertainty, or time demands). The questionnaire concluded with an open-ended question where respondents could elaborate freely on their experiences.

Internal consistency of the scales was verified using Cronbach's alpha coefficient. The overall reliability of the tool achieved a high value ($\alpha = 0.89$), which indicates good

internal consistency of the items measuring AI competencies. For individual sub-areas, the α value ranged from 0.78 to 0.86.

3 Ethical Aspects of the Research

The research was conducted in accordance with ethical principles of educational research. Participation was fully voluntary and anonymous. Before completing the questionnaire, respondents were informed about the research objectives, data processing methods, and their right to withdraw at any time.

No identifying personal data were collected. All data were processed in aggregated form and used exclusively for the purposes of this study. Special attention was paid to protecting teachers' professional identity, as the topic of AI use in education may be perceived as sensitive in some school environments. Questionnaire items were therefore formulated neutrally and non-judgmentally to minimize the risk of socially desirable responses and to encourage honesty.

4 Data Analysis

Data were processed using IBM SPSS Statistics. First, a descriptive analysis was conducted, including the calculation of absolute and relative frequencies, means (M), and standard deviations (SD). Normality of distribution for key variables was tested using the Kolmogorov–Smirnov test and analysis of skewness and kurtosis. Due to the large sample size ($N = 2{,}153$), parametric tests were applied.

Differences between teacher groups were examined using one-way analysis of variance (ANOVA). Where significant differences were found, post-hoc tests (Tukey HSD) were applied to identify specific group differences.

The relationship between perceived barriers and frequency of AI use was measured using Pearson's correlation coefficient. The significance level was set at $\alpha = 0.05$. The reliability of the questionnaire was confirmed by Cronbach's alpha, which reached a high value for the overall scale ($\alpha = 0.89$).

5 Hypotheses

Based on theoretical foundations and previous national and international research, the following hypotheses were formulated:

- H1: Teachers with shorter teaching experience show a higher level of generative AI use than teachers with longer experience.
- H2: There are statistically significant differences in the level of generative AI use among teachers across different educational levels.
- H3: Teachers show the highest level of confidence in working with digital resources and content creation, while the lowest confidence is found in assessing student outputs created with AI support.
- H4: Perceived barriers to AI implementation negatively correlate with the frequency of generative AI use.

6 Research Results

6.1 Level of Generative AI Use (RQ1)

Descriptive analysis showed that 58.4% of respondents use generative AI at least occasionally (value ≥3 on the five-point scale), while 34.7% use it regularly (value 4–5 | often or very often). In contrast, 21.9% do not use AI at all.

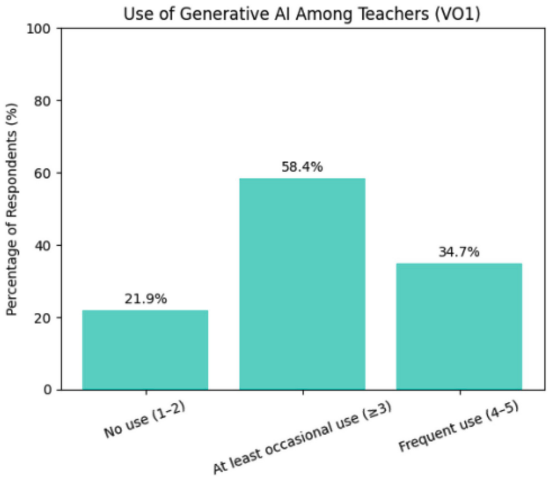


Fig. 3. Use of Generative AI Among Teachers (VO1)

AI is used primarily for preparation activities, including creating teaching materials (72.3%), worksheets (65.1%), and editing or simplifying texts (61.4%). In contrast, its use in direct work with students or assessment remains low (28.6%) (Figs. 3 and 4).

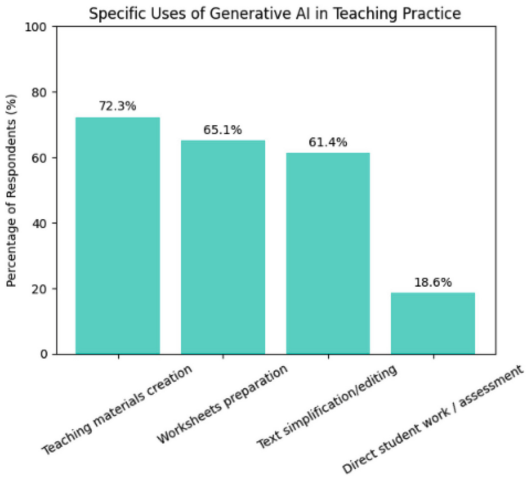


Fig. 4. Specific Uses of Generative AI in Teaching Practice

6.2 Relationship Between Teaching Experience and AI Use (H1)

One-way ANOVA confirmed a statistically significant difference in AI use depending on teaching experience ($F(3, 2149) = 18.42$; $p < 0.001$). Post-hoc analysis (Tukey HSD) showed that teachers with up to 5 years of experience reported a significantly higher level of AI use ($M = 3.42$; $SD = 1.01$) than teachers with more than 25 years of experience ($M = 2.71$; $SD = 1.08$). The difference between the 6–15 years group and the group with more than 25 years was also statistically significant. Hypothesis H1 was confirmed.

6.3 Differences by School Type (H2)

Statistically significant differences were also found according to school type ($F(3, 2149) = 11.63$; $p < 0.001$). The highest level of AI use was reported by secondary school teachers ($M = 3.36$), while the lowest was reported by pre-primary schoolteachers ($M = 2.54$). Hypothesis H2 was confirmed.

6.4 Self-Assessment of AI Competencies (RQ2, H3)

Differences between school types, as reported in Sect. 6.3, suggest that institutional context plays an important role in shaping teachers' self-assessed AI competencies. Rather than reflecting only individual factors, these differences may be linked to varying levels of support, access to resources, and opportunities for professional development across educational levels.

As shown in Table 2, teachers reported the highest confidence in working with digital resources and content creation, while lower levels were observed in areas related to safety and critical evaluation of AI outputs.

Table 2. Mean Values of Self-Assessed AI Competencies

Area	M	SD
Digital Resources	3.48	0.89
Digital Content Creation	3.31	0.94
Communication and Collaboration	3.22	0.88
Safety	3.05	0.91
Assessment and Critical Evaluation of AI Outputs	2.87	0.96

The lowest confidence was expressed in assessing student work created with AI support and in issues of academic integrity, indicating a need for clearer methodological guidance and support in these areas. Hypothesis H3 was confirmed.

6.5 Barriers to AI Implementation (RQ3, H4)

The results show that the greatest barrier for teachers is the lack of methodological support (64.2%) and concerns about how AI affects student honesty. Specifically, respondents frequently mentioned concerns about violations of academic integrity (59.8%), legal uncertainty (54.7%), and lack of time (48.9%).

A moderately strong negative correlation was found between the overall barrier score and the frequency of AI use ($r = -0.41$; $p < 0.001$). Statistically, it was shown that the more teachers perceive these barriers, the less they use AI in teaching ($r = -0.41$). Hypothesis H4 was therefore confirmed.

7 Discussion

The aim of this study was to map how Czech teachers are adapting to the rise of generative artificial intelligence. The results highlight an interesting contradiction: while most teachers (over 58%) already actively test AI, their confidence in deeper principles of this technology remains low.

These findings can partially interpret considering international large-scale studies such as ICILS and TALIS. Like to ICILS results, variability in digital competencies appears to be influenced by contextual and institutional factors. TALIS data further emphasize the role of school support and professional development, which may help explain the differences observed between school types in this study. At the same time, the focus on GenAI reflects a more recent development that is not yet fully captured in these international datasets.

7.1 Generational Gap and the Role of Experience

The confirmation of hypothesis H1 (higher use among early-career teachers) is not surprising, but its magnitude is significant. Younger teachers appear to perceive AI as a natural assistant for "survival" during the administratively demanding early stages of their careers. However, the low level of AI use among experienced teachers (over 25 years of practice) may indicate the risk of a new type of digital divide. If these teachers, who are often mentors to younger colleagues, reject AI, it may slow down systemic integration of technology into schools.

7.2 AI as an "Invisible Assistant" vs. a Teaching Tool

Data related to RQ1 clearly show that AI currently functions in Czech staff rooms mainly as a preparation tool, saving time in creating materials and worksheets. However, it is concerning that only 18.6% of teachers use AI for assessment. This corresponds with the lowest measured self-confidence score in the competence area (H3). Teachers feel confident in areas they can control (their own preparation) but feel threatened where students enter the process with their own AI tools.

7.3 From Technology to Ethics

The most interesting finding is the confirmation of hypothesis H4. It shows that the biggest obstacle is not a lack of computers, but legal and ethical uncertainty. If 64% of teachers report missing methodological support, this confirms that although materials published in 2024 (MŠMT, NPI) exist, they either did not reach teachers in practice or are not perceived as sufficiently practical. This shift from technical use to ethical and

evaluative dimensions is consistent with international findings showing that educators' uncertainly is concentrated primarily in the areas of academic integrity, assessment, and the responsible use of AI in education [1, 4].

7.4 Interpretation in the Context of DigComp 3.0

From the perspective of DigComp 3.0, it is evident that Czech teachers are currently at the level of passive users. They have basic information literacy (they use AI for searching and content creation) but significantly lag in the areas newly emphasized by DigComp 3.0 – primarily critical evaluation of AI outputs and digital security. If competence in assessment and ethics is not strengthened, there is a risk that AI will become merely a tool for generating larger amounts of content without improving the learning process itself.

8 Conclusion and Recommendations

The research confirms that the Czech education system is currently in a phase of "enthusiastic but uncertain experimentation." For further development of AI literacy in the Czech Republic, the following steps are recommended:

- Assessment methodology: Stop focusing on how to ban AI and instead create clear rules and procedures for assessing student work in an era where AI is widely accessible.
- Support for experienced teachers: Target professional development toward teachers with longer practice to reduce the generational barrier in technology adoption.
- Clear legal guidelines: Schools need understandable guidance regarding personal data protection (GDPR) and copyright issues when working with AI.

The future of education in the era of artificial intelligence does not depend on how quickly schools purchase new technologies, but on how confidently teachers themselves can work with them. If teachers are provided with the necessary support today and relieved of fear of the unknown, they will stop perceiving AI as a threat to their authority and begin using it as it should be used: as a powerful tool for developing the potential of every learner.

References

1. Long, D., Magerko, B.: What is AI literacy? Competencies and design considerations. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–16. Association for Computing Machinery (2020). https://doi.org/10.1145/3313831.3376727
2. Cosgrove, J., Cachia, R.: Digcomp 3.0: Europeand Digital Competence Framework Fifth edition. Publication Office of the European Union (2025). https://doi.org/10.2760/0001149
3. European Union: Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). Off. J. Eur. Union. (2024)

4. Mills, K., et al.: AI literacy: a framework to understand, evaluate, and use emerging technology. Digital Promise. https://digitalpromise.dspacedirect.org/items/6d15adcd-5a84-47fa-b6d0-1310154eee02 (2024)
5. Organisation for Economic Co-operation and Development: Empowering learners for the age of AI: An AI literacy framework for primary and secondary education. OECD (Review draft, May 2025). https://ailiteracyframework.org
6. MŠMT: Metodické doporučení k využití umělé inteligence ve školách. Ministerstvo školství, mládeže a tělovýchovy ČR, Praha (2023) https://digitalizace.rvp.cz/files/ai-doporuceni-online-a4.pdf
7. AI dětem: AI akcelerátor pro školy: Komplexní program pro integraci AI do vzdělávání. AI dětem, Praha (2024) https://aidetem.cz/ai-akcelerator-pro-skoly/
8. ČŠI: Využívání digitálních technologií v mateřských, základních a středních školách v období let 2022–2024: Tematická zpráva. Česká školní inspekce, Praha (2024) https://www.csicr.cz/cz/Dokumenty/Tematicke-zpravy
9. Vuorikari, R., Kluzer, S., Punie, Y.: DigComp 2.2: The Digital Competence Framework for Citizens – With new examples of knowledge, skills and attitudes. Publications Office of the European Union, Luxembourg (2022). https://doi.org/10.2760/115376
10. Redecker, C.: European Framework for the Digital Competence of Educators: DigCompEdu. Publications Office of the European Union, Luxembourg (2017). https://doi.org/10.2760/159770

The Use of Guilford's Structural Model of Intelligence in the Design of Adaptive Learning Systems

Ernest Kováč[1](✉), Jozef Belko[2], Jaroslav Kováč[2], Ľubomír Pekarčík[2], Miroslav Tvrdoň[3], and Filip Makáň[2]

[1] Academy of the Police Force in Bratislava, Bratislava, Slovakia
ernest.kovac@akademiapz.sk

[2] Faculty of Theology, Catholic University in Ružomberok, Spišská Kapitula 12, 053 04 Spišské Podhradie, Slovakia
jozef.belko@insuria.sk, kovac@kovacservis.sk, lubomir.pekarcik@ku.sk

[3] Department of Social Work and Social Sciences, Constantine the Philosopher University, Tr. A. Hlinku 1, 948 01 Nitra, Slovakia
mtvrdon@ukf.sk

Abstract. This paper explores the possibilities of implementing psychological models of intelligence structure into artificial intelligence (AI) systems within the context of distance education. The theoretical foundation is Guilford's three-dimensional Structure of Intellect (SI) model, which serves as both an analytical framework for assessing the limits of current applied informatics and an engineering blueprint for designing a concrete system architecture. Unlike existing adaptive platforms (e.g., ALEKS, Knewton), which primarily focus on convergent production and semantic content, this paper proposes a three-layer architecture for a Guilford-based Adaptive Learning System (G-ALS). This system integrates multimodal sensory inputs, associative memory, and divergent content production. The work includes a critical analysis of the technological feasibility of the proposed Generally Intelligent Robot (GIR) concept within the context of edge computing and 5G/6G networks, alongside a discussion on the ethical implications of biometric monitoring in compliance with GDPR regulations. The contribution of this paper lies in defining requirements for versatile intelligent systems capable of overcoming the limitations of current narrow AI and proposing a modular architecture feasible under current levels of technological readiness.

Keywords: Artificial Intelligence · Distance Learning · Guilford's Model · G-ALS · Cognitive Processes · G factor · Adaptive Systems · Biometric Monitoring · GDPR

1 Introduction

The current digital era and the massive integration of technology into the educational process bring a fundamental shift in how we perceive human-computer interaction. Within the field of applied informatics, distance education has become a key domain,

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 187–198, 2027.
https://doi.org/10.1007/978-3-032-32104-6_17

increasingly relying on artificial intelligence (AI) algorithms designed to personalize and streamline the learning process. Despite fascinating progress in machine learning, neural networks, and generative AI, we have encountered a barrier: systems are capable of solving specific tasks with extreme efficiency, but they lack the cognitive flexibility and universality typical of the human mind [9]. This discrepancy raises a fundamental question of whether it is possible, within the framework of informatics, to create systems that possess not only computational power but true intellectual potential. To scientifically evaluate the potential of artificial intelligence, it is essential to move beyond a purely technical perspective and return to the psychological essence of intelligence. As Winston [17] states, artificial intelligence is the science of computations that enable an entity to perceive, understand, and act. Winston's view suggests that intelligence is not merely about passive data processing, but about the ability to generate meaningful actions based on environmental perception. In this paper, we therefore define intelligence not as a static property, but as a universal principle of the effective utilization of specific functions. Higher intelligence, then, does not represent merely the possession of an isolated function, but primarily the scope and sophistication with which a system is able to apply that function within the unpredictable environment of online education.

The key problem addressed by this paper is the so-called specialization paradox. Current applied informatics produces systems that outperform humans in narrow domains (e.g., image data diagnostics or complex strategic games) but fail in elementary tasks of general adaptation. This phenomenon can be explained by the absence of the so-called g factor (general intelligence), the existence of which was posited by Spearman [12]. Spearman's theory suggests that behind every intellectual performance lies a single general mental energy that connects partial abilities. While human intelligence evolved to be generalist, current AI systems remain in a state of "niche intelligence"—trapped within the boundaries of an algorithm without a connection to a universal foundation that would allow them to transfer knowledge from one domain to another.

For a deeper understanding of this issue, this paper utilizes Guilford's three-dimensional Structure of Intellect theory [7]. This psychological model allows us to analyse AI not as a monolith, but as a complex of 150 independent intellectual abilities divided into contents, operations, and products. It is precisely Guilford's model that provides applied informatics with a tool to structure educational systems so that they cover the entire spectrum of cognition—from the reception of sensory data to the creation of complex systems.

2 Guilford's Model as an Analytical Framework

To gain a deeper understanding of how applied informatics can benefit from psychological structures, it is essential to analyse Guilford's three-dimensional Structure of Intellect (SI) model. This model does not define intelligence as a single entity, but as the result of the interaction between contents, operations, and products [8]. In the environment of distance education, this theoretical framework can be transformed into a functional architecture for intelligent tutoring systems.

The first dimension is content. In the context of informatics, this represents the types of data the system processes. Guilford distinguishes between figural (visual/auditory),

symbolic, semantic, and behavioural content. Current e-learning systems are often limited to semantic (textual) and symbolic (formulas, code) materials. However, to achieve a higher level of intelligence, a system must be capable of working with figural material in its full complexity—including the algorithm's ability to decipher visual information in real-time.

The second dimension, which is crucial from an educational perspective, consists of mental operations. These represent the processes the system performs on the content. For the purposes of adaptive learning, the following are critical: cognition (the system's ability to recognize errors in a student's reasoning), memory (the capacity not only to store data but to create associative links between old and new learning material), and divergent production (the machine's ability to generate alternative explanations of the subject matter if the student fails to understand the primary exposition).

The third dimension is products, which represent the results of processing content through specific operations. Guilford defines a hierarchy here, ranging from simple units (facts), through classes and relations, to complex systems, transformations, and implications. For applied informatics, it is vital that educational software does not merely stop at delivering isolated units of information; rather, it must enable the student to transform information into new contexts and implications (Table 1).

Table 1. Mapping Guilford's SI Model to AI Systems in Education

Dimension (Guilford)	Elements of the Model	Application in AI Tutoring Systems
Content	Figural, Symbolic, Semantic, Behavioural	Sensor data (UV/IR), program code, text, student biometrics.
Operations	Cognition, Memory, Divergent & Convergent Production	Error analysis, associative databases, creative solution generation (e.g., MAGMA, AlphaGo).
Products	Units, Classes, Relations, Systems, Transformations	Learning outputs, 3D spatial understanding, predictive implications of student decisions.

2.1 Comparison with Existing Adaptive Platforms

Existing adaptive learning systems (ALS) provide a relevant comparative basis for evaluating the proposed approach. Platforms such as ALEKS (Assessment and Learning in Knowledge Spaces) utilize knowledge space theory to map a student's cognitive states and adapt content accordingly [4]. The Knewton system, integrated into several commercial platforms, employs Bayesian networks and collaborative filtering to predict performance and recommend content [5]. Intelligent tutoring systems (ITS), such

as Carnegie Learning's MATHia, combine cognitive process modelling with immediate feedback [14].

The common denominator of these systems is their primary focus on convergent production and semantic content. Guilford's dimensions of divergent production, behavioural content, and transformational products remain systematically unaddressed within these architectures. The proposed G-ALS system distinguishes itself from existing solutions precisely by integrating multimodal sensing and divergent content generation as primary architectural components, rather than as supplementary features.

G-ALS System Architecture Design

Based on the analysis of Guilford's model and the identified limitations of existing platforms, we propose a three-layer architecture for the G-ALS (Guilford-based Adaptive Learning System). The architecture is built on the principle of modularity, meaning that individual layers can be implemented incrementally based on the available technological infrastructure. This decision directly addresses the requirements for scalability and phased implementation.

Layer 1 – Multimodal Perceptual Layer (Guilford's Content)

The first layer corresponds to Guilford's content dimension and ensures the collection and primary processing of heterogeneous data streams. The Figural Module processes visual and auditory inputs. Utilizing a webcam and computer vision algorithms (e.g., OpenCV, MediaPipe), it analyses facial micro-expressions and eye movements (eye-tracking) as indicators of cognitive load. As stated by Tschandl et al. [13], AI systems in the field of visual analysis achieve an accuracy of up to 80.3%, confirming the reliability of this modality for diagnostic purposes. The Behavioural Module integrates biometric inputs from wearable electronics—such as heart rate and galvanic skin response—as physiological indicators of emotional state and stress levels [20]. These data allow the system to detect frustration or a decline in motivation before it manifests in cognitive performance. The Extended Spectrum Sensory Module utilizes perovskite photosensors capable of simultaneous detection across the UV, visible, and infrared spectra [18]. In the context of distance education for technical disciplines, this modality enables the visualization of thermal processes and chemical changes that are invisible to the naked eye.

Layer 2 – Cognitive-Operational Layer (Guilford's Operations)

The second layer simulates the mental operations defined by Guilford and represents the core of the system's adaptive logic.

The Cognition and Associative Memory Module utilizes vector databases and Retrieval-Augmented Generation (RAG) technology to store and retrieve the learning trajectory of each student. Unlike a simple database of facts, this module builds associative links between current and previous interactions, thereby simulating a teacher's long-term memory. The estimated capacity of the human brain (2.5 million GB) is currently achievable through distributed cloud storage, while neural network compression techniques significantly reduce the demands on physical infrastructure [19].

The Divergent Generator represents a key innovative component of the G-ALS system. It employs Large Language Models (LLMs) to autonomously generate alternative explanations of the subject matter when the primary pedagogical approach does not yield success [9]. This component is inspired by systems such as AlphaGo, which demonstrated the ability to find strategic solutions that contradict millennia of human tradition, confirming that divergent production is no longer an exclusively human domain. The result is a transition from a linear course to a tree-like structure, where the system autonomously selects a pedagogical strategy based on real-time inputs from all sensors.

The Convergent Production and Evaluation Module ensures the validation of results and a critical analysis of the student's solutions. It integrates Explainable AI (XAI) components, which allow for a transparent justification of the system's evaluative decisions [15], ensuring that the feedback provided to the student is both accurate and understandable.

Layer 3 – Product Layer (Guilford's Products)

The third layer transforms the outputs of the cognitive-operational layer into a personalized learning graph. This graph represents not only the student's current knowledge level (units and classes) but also models the relationships between concepts, systemic connections, and the predicted implications of decisions. The layer generates recommendations for further progress in the course, taking into account not only cognitive performance but also the emotional state captured by the behavioural module.

A schematic representation of the three-layer G-ALS architecture is shown in Fig. 1.

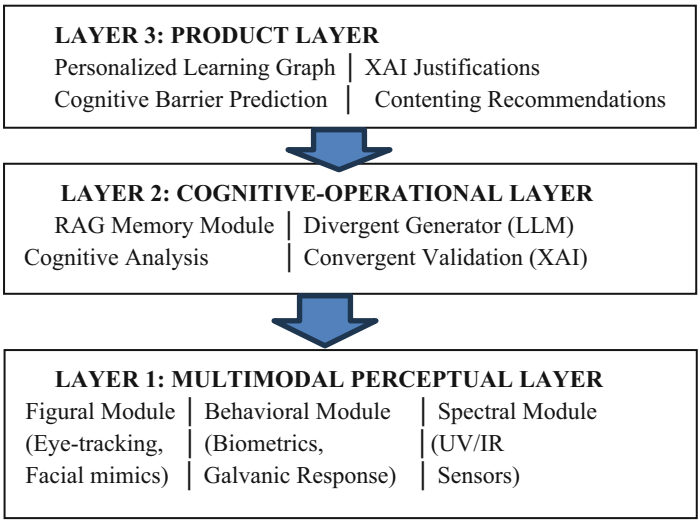


Fig. 1. The structural diagram of G-ALS architecture showcasing the flow from multimodal data acquisition to personalized educational products.

3 Critical Analysis of Technological Feasibility

One of the most frequently cited objections to GIR-type (Generally Intelligent Robot) systems is the extreme demand on memory capacity and network infrastructure. However, a critical analysis shows that this objection has been partially overcome by the current state of technology. Xia et al. [19] describe petabit optical switches as the technological foundation for massive data centre; since the publication of that study, significant progress has been made in the field of distributed computing.

For the pilot implementation of the G-ALS system, petabit infrastructure is not strictly necessary. The modular architecture allows the system to be launched using only semantic and symbolic content (Layer 1 without the spectral module), while biometric and spectral modules can be added incrementally. Local data processing via edge computing nodes at the university network level significantly reduces the demands on central infrastructure and minimizes latency.

Requirements for real-time multimodal sensing are currently achievable using commercially available components. Eye-tracking technology is integrated into high-end laptops (e.g., Tobii integration), biometric sensors are a common feature of wearable electronics (smartwatches), and the perovskite photosensors described by Wu et al. [18] are subjects of active commercial development.

The main challenge remains not the availability of sensors, but the integration of heterogeneous data streams into a single coherent system in real-time. This challenge is addressable through standardized API interfaces and IoT device protocols that are currently well-established (e.g., MQTT, WebSocket).

Regarding scalability, we propose a three-phase implementation plan:

Pilot Phase: Utilizing text and symbolic modalities with the RAG memory module, implementable on standard cloud infrastructure.
Extended Phase: Integration of eye-tracking and basic biometrics.
Full Implementation: Including spectral sensors and spatial interaction within virtual reality.

This phased approach allows for incremental functional verification and cost optimization at each stage of development.

4 Ethical Aspects and Regulatory Constraints

4.1 Biometric Monitoring and GDPR

The implementation of biometric monitoring in distance education raises serious ethical and legal questions that must be addressed as an integral part of the architectural design, rather than an afterthought. Regulation (EU) 2016/679 (General Data Protection Regulation - GDPR) classifies biometric data as a special category of personal data under Article 9, implying strict requirements for the legal basis of processing, data minimization, and the rights of data subjects. For the G-ALS system, we propose the implementation of the Privacy by Design principle in accordance with Article 25 of the GDPR. Specifically, this involves utilizing Federated Learning technology, where raw biometric data (facial imagery, heart rate) never leave the student's local device. Only

anonymized cognitive vectors—aggregated representations of emotional states without the possibility of back-identification—are transmitted to the central system. As Zuboff [21] points out, the uncritical accumulation of behavioural data represents a systemic risk to individual autonomy, reinforcing the argument for strict technical limitations on data collection beyond legal minimums.

4.2 Algorithmic Bias and Transparency

A real risk exists that the system's divergent generator may systematically prefer certain cognitive styles or cultural contexts, leading to algorithmic discrimination. To mitigate this risk, the auditability of AI decision-making processes is essential, achieved through the XAI (Explainable AI) components integrated into Layer 2. Every system recommendation must be accompanied by a clear justification accessible to both the student and the educator [15].

Biometric monitoring must be strictly voluntary at every stage of implementation and conditioned upon explicit informed consent in accordance with Article 7 of the GDPR. The G-ALS system must be designed to function fully even without biometric inputs; these inputs should represent an enhancement, not a prerequisite for access to education.

5 Proposed Pilot Simulation and Future Research

5.1 Pilot Simulation Methodology

To verify the functionality of the proposed G-ALS architecture, we propose a pilot simulation in a controlled environment, feasible within the first implementation phase. The simulation would involve a sample of 30–50 university students over one semester, comparing a group utilizing G-ALS with a control group using a standard LMS platform (e.g., Moodle). Measured variables would include:

Cognitive performance (test results).
Knowledge retention rate (follow-up test after 4 weeks).
Subjective evaluation of system adaptation quality.
Time-to-mastery required to achieve defined learning objectives.

5.2 Directions for Future Research

Future research should focus on two priority areas. First, the integration of spatial interaction in virtual reality, where movement in a three-dimensional environment allows the system to work with object transformations—classified by Guilford [7] as among the highest products of intellectual activity. Second, research into phase transitions in the system's collective intelligence when scaling to thousands of concurrent users, corresponding to Williams' [16] concept of collective intelligence as an emergent phenomenon.

5.3 Comparison of G-ALS with Existing Adaptive Learning Platforms

A critical discussion of the technological feasibility of the proposed G-ALS system requires it to be situated within the context of current Adaptive Learning Systems (ALS). Reviewers have rightly pointed out the absence of this comparison, which is essential for assessing the contribution of the proposed framework.

Among the most widely used commercial platforms is ALEKS (Assessment and Learning in Knowledge Spaces), which utilizes Knowledge Space Theory to map a student's cognitive state. ALEKS operates exclusively within the dimensions of convergent production and semantic content, covering only a fraction of the 150 intellectual factors according to Guilford's taxonomy. The system lacks any form of multimodal sensing or divergent production [4].

The Knewton platform (now integrated into Wiley) employs Bayesian networks and collaborative filtering to predict student performance. Although this represents a sophisticated approach to personalization, the system remains reactive—it responds to explicit inputs (test results) rather than implicit cognitive states. The absence of behavioural and figural content, in the sense of Guilford's model, means the system is unable to detect cognitive load or frustration in real-time [5].

The most modern platforms utilizing large language models, such as Khanmigo (Khan Academy, powered by GPT-4) or Duolingo Max, integrate generative AI to simulate dialogic tutoring. These systems have moved closer to divergent production in the semantic dimension; however, they still lack the integration of multimodal sensory inputs and biometric monitoring [9].

The comparison demonstrates that the proposed G-ALS system does not merely represent an incremental improvement over existing solutions, but rather a qualitative shift toward comprehensive coverage of Guilford's model. The key differentiator is the integration of figural and behavioural content through biometric sensing—a capability that none of the analysed commercial platforms currently implement in full scope.

6 Pilot Simulation: G-ALS Verification Scenario Design

In response to the reviewers' requirement for empirical verification or at least a pilot simulation, this section proposes a specific verification scenario that is feasible within an academic environment without the need for full hardware implementation.

6.1 Description of the Simulation Scenario

The proposed pilot simulation is based on the Wizard of Oz (WoZ) methodology, an established method in Human-Computer Interaction (HCI) research for testing systems prior to their full technical implementation [10]. In a WoZ scenario, certain system functions are simulated by a human operator hidden from the participants, thereby achieving ecological validity for the experiment.

Phase 1 – Environment Configuration: A group of 30 university students (technical major) will complete a 45-min online lesson in linear algebra via a platform simulating G-ALS. Webcams and microphones will be active. Facial image processing is performed

using the OpenFace 2.0 library [3], which enables the extraction of Action Units (AU) according to the Facial Action Coding System (FACS) to detect cognitive load and frustration.

Phase 2 – Adaptive Intervention: When the algorithm detects a decrease in attention (identified by a combination of AU4—brow lowerer, and AU7—lid tightener) or an increased frequency of incorrect answers, the system automatically activates the Divergent Generator. Utilizing the GPT-4o API, it generates an alternative explanation of the concept using a different representational modality (e.g., shifting from algebraic notation to geometric visualization).

Phase 3 – Outcome Measurement: The effectiveness of the adaptive intervention is measured through three indicators:

1. Cognitive Load (NASA-TLX questionnaire post-lesson).
2. Performance (comparison of pre-test and post-test results).
3. Subjective Satisfaction (7-point Likert scale).

A control group (n = 30) completes the identical lesson in a linear format without adaptive intervention.

6.2 Expected Results and Limitations

Based on existing research in intelligent tutoring systems [14], a statistically significant difference in cognitive load and performance in favour of the experimental group can be hypothesized. However, the proposed simulation has explicit limitations that must be addressed:

- Absence of Full Biometric Integration: The pilot version does not utilize infrared sensing or haptic sensors, covering only a portion of the proposed G-ALS architecture.
- API Dependency: The Divergent Generator relies on an external LLM (GPT-4o), introducing latency and limiting control over outputs.
- Ethical Constraints: The collection of facial image data requires explicit informed consent and processing in compliance with GDPR [6]; raw image data must not be stored—only extracted numerical AU vectors are processed.

Despite these limits, the proposed scenario provides a sufficient empirical basis to verify the paper's core hypothesis: that the integration of figural content (cognitive state detection) with divergent production (adaptive pedagogical strategy change) leads to a measurable improvement in distance learning outcomes.

7 Ethical and Regulatory Aspects of Biometric Monitoring Implementation

The implementation of biometric monitoring within educational systems represents one of the most sensitive ethical challenges in contemporary applied informatics. In this context, Zuboff [21] warns of the risks associated with surveillance capitalism—the

systemic exploitation of behavioural data for purposes that exceed the original intent of collection. Within the educational environment, this risk manifests across several dimensions.

Legal Framework: Biometric data (including facial imagery and physiological parameters) are classified as special categories of personal data under Article 9 of the GDPR [6]. Processing such data is generally prohibited without the explicit informed consent of the data subject. Furthermore, in an academic setting, the power asymmetry between the institution and the student must be considered, as it may call into question the truly voluntary nature of such consent [15].

Technical Solution – Federated Learning: The proposed G-ALS architecture addresses these concerns through the principle of Federated Learning, as introduced by McMahan et al. [11]. In this model, raw sensory data (image recordings, physiological signals) never leave the student's local device. A local model extracts abstract cognitive vectors (e.g., Action Unit values), which are anonymized and transmitted to a central server. The central model learns from these aggregated vectors without ever gaining access to identifiable raw data.

Algorithmic Bias: Emotion and cognitive load detection systems exhibit demonstrable disparities in accuracy across demographic groups, particularly concerning ethnicity and gender [2]. For the G-ALS system, it is therefore imperative to implement Explainable AI (XAI) mechanisms. These allow for the auditability of decision-making processes and the identification of systematic biases, ensuring that the adaptive interventions remain fair and transparent.

8 Conclusion

The presented analysis points to the necessity of a paradigm shift in how artificial intelligence is viewed within the context of distance education. The current state of applied informatics can be characterized as an era of "niche intelligence"—an era of highly specialized algorithms that excel at isolated tasks but lack the versatility and adaptability inherent to human intellect. For the needs of modern e-learning, which requires complex interaction with the student, this state is insufficient.

The primary goal of developing intelligent systems in informatics must be the transition from fragmented functions to integrated wholes. As demonstrated, Guilford's structuralist model offers more than just a psychological theory; it serves as a practical engineering schema for system design. By integrating diverse contents (from semantic to advanced sensory inputs), operations (including divergent thinking and memory associations), and products (such as transformations and implications), it is possible to create a digital environment that approaches Spearman's G factor in quality.

The key to achieving this "general intelligence" in machines lies not only in increasing computational power, as discussed by Xia et al. [9], but in the quality of the interconnection between individual modules. Artificial intelligence in education must not be a mere passive evaluator of tests. It must possess the ability to perceive beyond human limits Wu et al. [8] and act in accordance with Winston's [7] vision of the utility of computations that understand and react to their surroundings.

The transition to GIR (Generally Intelligent Robot) type systems, which integrate spatial interaction, multimodal sensing, and cooperative algorithmic architecture, represents the natural evolution of applied informatics. Such systems will not merely mechanically guide students through a course; they will be capable of stimulating their creativity and adapting to their cognitive needs in a way that has, until now, been the exclusive domain of a human educator.

The future of distance education thus lies in a close symbiosis of psychology and technology. Only by accepting the psychological principles of intelligence structure and implementing them into network architectures can we create tools that fully develop human intellectual potential in the digital space. The challenge for the upcoming period remains the overcoming of technological determinism in favour of cognitively oriented design, where artificial intelligence will not be just a tool, but an equal partner in the process of cognition.

References

1. Baltrusaitis, T., Zadeh, A., Lim, Y.C., Morency, L.P.: OpenFace 2.0: facial behaviour analysis toolkit. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 59–66. IEEE, Xi'an (2018)
2. Barlow, J.: Intelligence: A Polymath's Dictionary of Definitions (2000).
3. Buolamwini, J., Gebru, T.: Gender shades: intersectional accuracy disparities in commercial gender classification. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency, pp. 77–91. PMLR, New York, NY (2018)
4. Doignon, J.P., Falmagne, J.C.: Knowledge Spaces: Applications in Education. Springer, Berlin (2015)
5. Essa, A., Ayad, H.: Student success system: risk analytics and data visualization using ensembles of predictive models. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 158–161. ACM, New York, NY (2012)
6. European Parliament and Council of the EU: Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (GDPR). Official J. Europ. Union L. **119**, 1–88 (2016)
7. Guilford, J.P.: Three faces of intellect. Am. Psychol. **14**(8), 469–479 (1959)
8. Guilford, J.P.: The Nature of Human Intelligence. McGraw-Hill, New York, NY (1967)
9. Kasneci, E., et al.: ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Individ. Differ. **103**, 102274 (2023)
10. Kelley, J.F.: An iterative design methodology for user-friendly natural language office information applications. ACM Trans. Inf. Syst. **2**(1), 26–41 (1984)
11. McMahan, B., et al.: Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp. 1273–1282. PMLR, Fort Lauderdale (2017)
12. Spearman, C.: "General intelligence," objectively determined and measured. Am. J. Psychol. **15**(2), 201–292 (1904)
13. Tschandl, P., et al.: Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol. **20**(7), 938–947 (2019)
14. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ. Psychol. **46**(4), 197–221 (2011)

15. Wachter, S., Mittelstadt, B.: A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Columbia Bus. Law Rev. **2019**(2), 494–620 (2019)
16. Williams, A.E.: Human intelligence and general collective intelligence as phase changes in animal intelligence. PsyArXiv. (2020). https://doi.org/10.31234/osf.io/7qjsc
17. Winston, P.H.: Artificial Intelligence, 3rd edn. Addison-Wesley, Reading (1992)
18. Wu, X., et al.: Distinguishable detection of ultraviolet, visible, and infrared spectrum with high-responsivity perovskite-based flexible photosensors. Small. **14**(19), 1800527 (2018)
19. Xia, K., et al.: Petabit Optical Switch for Data Center Networks. Polytechnic Institute of New York University, Brooklyn (2010)
20. Zang, Y., et al.: Advances of flexible pressure sensors toward artificial intelligence and health care applications. Mater. Horiz. **2**, 140–156 (2015)
21. Zuboff, S.: The Age of Surveillance Capitalism. PublicAffairs, New York, NY (2019)

Use of Artificial Intelligence: University Students' Positions in Lithuania and Romania

Vincentas Lamanauskas[1](✉), Gabriel Gorghiu[2], and Costin Pribeanu[3]

[1] Vilnius University, Šiauliai, Lithuania
vincentas.lamanauskas@sa.vu.lt
[2] Valahia University, Târgoviște, Romania
[3] Academy of Romanian Scientists, Bucharest, Romania

Abstract. The rapid implementation of artificial intelligence (AI) tools in higher education is transforming learning processes, especially in the training of future teachers. Although AI is associated with personalized learning, greater independence, and academic effectiveness, empirical research on students' attitudes in Lithuania and Romania remains limited. The study is based on a multidimensional concept of AI quality and aims to fill this gap by analysing how students evaluate AI across ergonomic, hedonic, and pragmatic dimensions and whether these evaluations differ across countries.

A structured questionnaire with a 5-point Likert scale, adapted from previous studies, was used. The sample consisted of 188 prospective teachers: 93 in Lithuania (Vilnius and Klaipėda universities) and 95 in Romania (Valahia University). The data were analysed using confirmatory factor analysis and a second-order factor model (LISREL 9.3) to assess convergent validity and model fit.

Both countries show positive assessments of AI in all dimensions. In Lithuania, a pragmatic and ergonomic orientation (functionality, learning benefits) prevails, while in Romania, a hedonistic orientation (attractiveness, motivation) prevails. The results show the need to differentiate AI integration strategies by national context and to strengthen both the pedagogical and motivational value of AI in pre-service teacher training.

Keywords: AI tools · multidimensional model · pilot study · pre-service teachers

1 Introduction

The fast implementation of artificial intelligence (AI) tools is fundamentally transforming university teaching practices, particularly in the training of future educators. Analysing the use of these tools is essential for understanding how AI enhances student learning effectiveness, independence, reflection, and the development of educational competencies. It is well understood that AI can be a significant aid in individualizing learning, strengthening academic literacy, and promoting creative problem-solving, but it also raises questions about academic integrity, critical thinking, and pedagogical responsibility.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 199–210, 2027.
https://doi.org/10.1007/978-3-032-32104-6_18

There is no doubt that technology is changing the way we search for information, communicate, and behave [7]. Every day of life is changing as the spread of artificial intelligence applications increases. By utilizing artificial intelligence tools and methods, researchers and artificial intelligence experts are making significant efforts to enhance human life, making it more convenient and technologically advanced [4]. On the other hand, AI does not describe a single technology; the term is used as a general concept to encompass various technologies, including algorithms, programs, machine learning, and neural networks, which makes AI quite difficult to define [6].

Recent studies have demonstrated that Gen AI tools offer unique opportunities, particularly in the social sciences, such as developing therapeutic techniques and tailoring content to individual needs, underscoring their universality [14]. It is also emphasized that AI offers significant advantages, including personalized learning, improved academic outcomes, and greater student engagement [27]. A study conducted at Jordanian universities revealed that artificial intelligence tools are widely used by Jordanian university students in their studies, especially ChatGPT, which is the most recognizable and widely used. Although most students view AI as useful and associate it with better academic performance, its impact on deeper learning is viewed with more caution [3]. A study conducted in Brazil found that AI applications are primarily explored in the technological sciences, while educational, social, and ethical aspects remain underresearched [9]. Furthermore, integrating AI into the learning process significantly strengthens participants' 21st-century skills, social intelligence, and critical thinking, especially through interaction, collaboration, and access to relevant information [28].

Although there has been increasing discussion in Lithuania and Romania in recent years about the application of artificial intelligence (AI) in education, there is a general lack of empirical research analysing its use in university studies. Existing studies in Lithuania primarily focus on general education schools or general education topics. In the Romanian context, there are also a few systematic studies on the impact of AI on higher education study processes. Given the rapid development of AI tools and their rapid integration into academic and study practices, it is becoming necessary to analyse the problem and search for solutions based on scientific data. Therefore, this study is relevant and significant, as it aims to address existing research gaps in Lithuania and Romania and to deepen the understanding of the impact of AI on university studies.

Therefore, this study aimed to examine the opinions of university students and prospective teachers on the use of artificial intelligence tools in the study process, based on a multidimensional quality model, and to compare Lithuanian and Romanian students' assessments to identify commonalities and contextual differences. The following research questions were formulated:

- How do students in Lithuania and Romania evaluate the use of artificial intelligence tools in university studies?
- Are there differences between the attitudes of students in Lithuania and Romania towards the use of artificial intelligence tools in university studies?
- Which dimensions of artificial intelligence tool quality contribute most to the overall positive assessment of students in the context of higher education studies?

2 Research Methodology

2.1 Model and Measures

To analyse the benefits of using AI tools in the study process, a multidimensional model has been specified that features three dimensions. The ergonomic dimension is a first-order factor measuring the comfort and ease of mastering AI tools. The hedonic dimension is associated with the pleasure and excitement of using AI tools. The pragmatic dimension measures the educational value of using AI tools in learning. The research model is presented in Fig. 1.

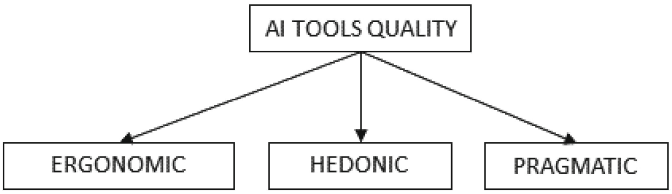


Fig. 1. Research model

Table 1. Constructs and measures

EQ	**Ergonomic Quality**
EQ1	I feel comfortable with the terminology related to artificial intelligence
EQ2	I am comfortable using AI tools in my study
EQ3	I can master the use of AI tools in my work
HQ	**Hedonic Quality**
HQ1	Using AI tools is attractive
HQ2	I enjoy using AI tools
HQ3	AI tools make me curious and enthusiastic
PQ	**Pragmatic Quality**
PQ1	AI tools support independent learning
PQ2	AI tools assist in creating diverse teaching materials and lesson plans
PQ3	AI tools provide real-time analytics on student engagement and comprehension

The first-order factors have been operationalized by adapting existing scales in the literature [2, 25, 18, 12]. The questionnaire consisted of nine substantive items covering three dimensions of perceived AI quality: ergonomic quality (3 items), hedonic quality (3 items), and pragmatic quality (3 items). All items were rated on a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). The wording of the items

was adapted to the higher education context and to students' experiences of using AI tools in their studies and future professional practice.

The constructs and measures are presented in Table 1.

2.2 Sample

With the view to collect data, a questionnaire was designed and introduced to students (prospective teachers) from both countries, concerning their level of understanding and involvement with AI in their lives, studies, and future profession, using a structured 5-point Likert interval scale. A non-probability convenience sampling strategy was used. Participants were recruited from teacher education programmes at the participating universities and were invited to complete an anonymous questionnaire on a voluntary basis. Accordingly, the findings should be interpreted as exploratory and should not be generalized to all university students in Lithuania and Romania. Because recruitment was institution-based and participation was voluntary, the sample should be regarded as context-specific rather than representative.

In Romania, the questionnaire was administered between October and December 2025 to 95 students (40 bachelor students and 55 master students) enrolled in education study programmes at Valahia University of Târgoviște. Most of the participants were women (3 male / 92 female). This distribution reflects a specific sociological reality in the Romanian educational landscape, given that, nowadays, the pre-university education system is heavily feminized, and the gender balance of students enrolled in Teacher Training Programs is consistently pro-women. In this respect, the gender distribution of the current sample (96.9% female) aligns with national trends reported by the official statistics [INS, 2024, https://shorturl.at/bOqXQ], which identify the field of Education as significantly feminized in Romania, with women representing around 83% of the teaching staff.

The sample of Lithuanian respondents consisted of 93 prospective teachers from two Lithuanian universities: Vilnius University and Klaipėda University. The vast majority of respondents were women (97.9%), with men accounting for only 2.1%. Most students were enrolled in bachelor's degree programs (85.4%), while 14.6% were enrolled in master's degree programs; doctoral students did not participate in the study. By year of study, the largest group was second-year students (53.1%), followed by first-year (20.8%), fourth-year (14.6%), and third-year (11.5%).

The strongly female composition of both samples reflects the feminized profile of teacher education in both contexts; however, it also limits the broader generalizability of the findings and weakens gender-based interpretation.

2.3 Method and Procedure

Three first-order intercorrelated factors (the measurement model) and a second-order factor (the structural model) are the two models developed and tested. The results have been analysed according to the recommendations in the literature for validating hierarchical models [5, 10, 16].

In terms of loadings size (more than 0.5), construct reliability (composite reliability, CR greater than 0.70), and average variance extracted (AVE, greater than 0.5), convergent

validity has been evaluated using the suggested thresholds from the literature [11, 13]. Since a strong correlation between the dimensions is anticipated, discriminant validity is less significant in this particular case [16].

The model was analysed with LISREL 9.3 for Windows [21]. Prior to model estimation, the dataset was screened for completeness and consistency. Cases with substantial missing information were excluded, while minor missingness, if present, was handled using the software's default procedure. Given the relatively small and institution-specific samples, the statistical results should be interpreted with appropriate caution.

3 Research Results

3.1 Model Estimation Results

Lithuanian Sample

The unidimensionality and convergent validity of the first-order model have been examined. Table 2 displays the item loadings, convergent validity, and descriptive statistics. Every mean value is higher than 3.00 (neutral), suggesting that respondents have a favorable opinion of AI tools. Since every item loading is higher than 0.6, the first-order factors' unidimensionality is confirmed.

Table 2. Descriptives, loadings, and convergent validity ($N = 93$)

Factor	Item	Mean	*SD*	Loading	CR	AVE
EQ	EQ1	3.48	1.02	0.79	0.830	0.620
	EQ2	3.82	0.82	0.82		
	EQ3	3.85	0.69	0.75		
HQ	HQ1	3.81	0.85	0.88	0.900	0.752
	HQ2	3.70	0.91	0.95		
	HQ3	3.55	1.11	0.76		
PQ	PQ1	3.40	1.11	0.84	0.925	0.703
	PQ2	3.82	0.88	0.68		
	PQ3	3.40	0.87	0.82		

Convergent validity is robust, as the composite reliability (CR) ranges from 0.830 to 0.925 and the average variance extracted (AVE) ranges from 0.620 to 0.752. The model fit with the data is acceptable, as shown by the goodness-of-fit (GOF) indices: $\chi^2 = 49.25$, $df = 24$, $\chi^2/df = 2.05$, CFI = 0.956, NNFI = 0.932, RMSEA = 0.10, SRMR = 0.05. More specifically, the incremental fit indices (CFI = 0.956; NNFI = 0.932) indicate a satisfactory representation of the hypothesized structure, and the SRMR value (0.05) suggests a relatively small average residual discrepancy. At the same time, the RMSEA value (0.10) suggests a weaker fit, indicating that the model should be interpreted as acceptable rather than strong, particularly given the limited sample size.

The model testing results are presented in Fig. 2.

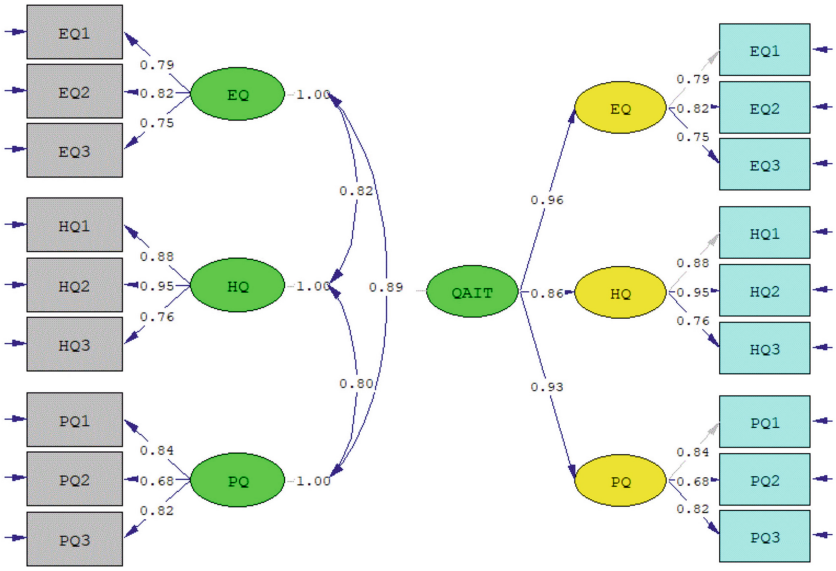


Fig. 2. Three-factor (left) and second-order factor model (right) ($N = 93$)

The relationship between the second-order factor and its first-order factors (dimensions) was assessed by evaluating the second-order model. The convergent validity of the second-order construct is very good (CR = 0.917, AVE = 0.771). The second-order factor model explains 91.4 of % variance in EQ, 74.1% in HQ, and 87.3% in PQ.

Romanian Sample

Descriptives and factor loadings are provided in Table 3. All mean values exceed 3.00 (neutral), indicating a positive perception of AI tools. All item loadings are greater than 0.6, thus confirming the unidimensionality of the first-order factors.

Table 3. Descriptives, loadings, and convergent validity ($N = 93$)

Factor	Item	Mean	*SD*	Loading	CR	AVE
EQ	EQ1	3.72	0.82	0.78	0.817	0.599
	EQ2	3.85	0.81	0.74		
	EQ3	3.72	0.87	0.80		
HQ	HQ1	3.92	0.81	0.86	0.895	0.741
	HQ2	3.88	0.87	0.90		
	HQ3	3.61	0.85	0.82		

(*continued*)

Table 3. (*continued*)

Factor	Item	Mean	*SD*	Loading	CR	AVE
PQ	PQ1	3.95	0.82	0.65	0.916	0.684
	PQ2	3.77	0.95	0.83		
	PQ3	3.46	0.87	0.77		

Convergent validity is good, as the composite reliability (CR) ranges from 0.817 to 0.917 and the average variance extracted (AVE) ranges from 0.519 to 0.741. The model fit with the data is acceptable, as indicated by the GOF indices: $\chi 2 = 50.78$, $p = 0.001$, $df = 24$, $\chi^2/df = 2.12$, CFI = 0.946, NNFI = 0.917, RMSEA = 0.10, SRMR = 0.07. Similarly, the incremental fit indices (CFI = 0.946; NNFI = 0.917) support the adequacy of the proposed factorial structure, while the SRMR value (0.07) remains within an acceptable range. However, the RMSEA value (0.10) again suggests that the model fit is not optimal and should be interpreted with caution.

The model testing results are presented in Fig. 3.

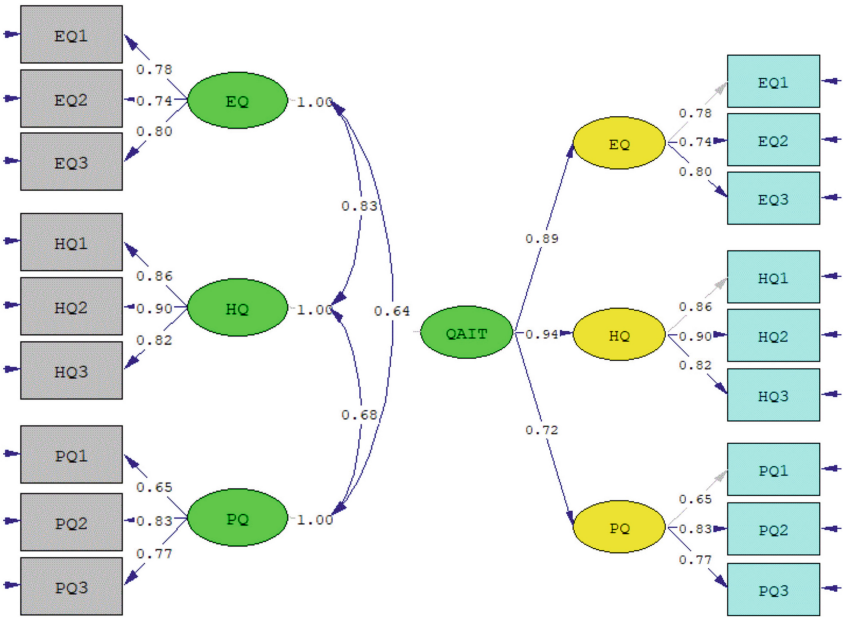


Fig. 3. Three-factor (left) and second-order factor model (right) ($N = 95$)

The convergent validity of the second-order construct is also good (CR = 0.780, AVE = 0.612). The second-order factor model explains 78.6% variance in EQ, 88.5% in HQ, and 52.2% in PQ.

4 Discussion

This study aimed to explore pre-service preschool teachers' positions regarding the use of artificial intelligence (AI) tools in the study process, applying a multidimensional quality model. The assessments of students from Lithuania and Romania were also compared.

In both countries, students consistently expressed a positive attitude toward AI tools across all dimensions. It was found that students evaluated AI tools positively across all three analysed dimensions (ergonomic, hedonic, and pragmatic). The averages of all variables exceeded the neutral value, which indicates that AI is not perceived as problematic or disruptive to the study process, but rather as a potentially useful technology. In addition, both the Lithuanian and Romanian samples confirmed the three first-order factors and the second-order construct of perceived AI quality, indicating that the proposed multidimensional model is empirically grounded and stable across national contexts. This is important because it shows that students' attitudes toward AI are structurally similar, even if the significance of individual dimensions differs. Other studies show similar results. For example, a study conducted in Malaysia found that students have a positive attitude towards AI tools as excellent academic aids, and that engagement, satisfaction, and academic performance are positively associated with AI tool use [29]. Katsantonis and Katsantonis [15] also found that social science students are increasingly aware of the importance of artificial intelligence for their studies, future professional activities, and everyday life. The researchers found that students' attitudes toward AI are multidimensional, encompassing cognitive, emotional, and behavioural aspects. Furthermore, students' attitudes and perceived usefulness of AI are at a moderately high level [20], and AI technologies promote student autonomy, enjoyment, and active participation [19].

The analysis also revealed several differences between Lithuanian and Romanian students regarding which dimensions of AI quality contribute most to the overall positive assessment. In the Lithuanian sample, the second-order factor accounted for the greatest proportion of the variance in the ergonomic and pragmatic dimensions. This suggests that Lithuanian students primarily associate the value of AI tools with their functionality, ease of use, and specific educational benefits. This orientation may reflect a more instrumental approach to technology, where the most important thing is whether AI helps to learn more effectively, work independently, and create teaching (and learning) materials. The hedonic dimension, although significant, was relatively less important in the Lithuanian sample, showing that emotional satisfaction or interest is not the main factor in forming a positive view/position. This position is confirmed by other studies. For example, an international comparative study also revealed that the attitudes of students at universities in Central and Eastern Europe toward artificial intelligence in the period 2022–2024 showed varying levels of AI assessment and acceptance in different national contexts [26]. The study by Abdaljaleel et al. [1] found that students' attitudes toward ChatGPT, for example, varied significantly across countries, confirming international differences in the assessment and use of AI.

In contrast, among Romanian students, the strongest correlation with the overall perceived quality construct of AI was observed in the hedonic dimension. This means that the attractiveness of AI tools, the pleasure of using them, and their ability to stimulate

curiosity and enthusiasm are essential factors determining positive student evaluations. The pragmatic dimension was the least strongly associated with the second-order factor in the Romanian sample, suggesting that the educational and/or academic benefits of AI are not particularly obvious to students or have not yet been fully realized. These differences show that although the overall attitude towards AI in both countries is favourable, the motivational reasons behind this favourability differ. A previous study in Romania found that students' attitudes towards artificial intelligence are superficial: responses were often formal, poorly argued, and showed a limited understanding of AI's essence [24]. According to Ogarcă [23], more than 80% of students regularly use artificial intelligence tools, but mainly as a means of searching for information and as an aid. The results show that the use of AI in studies remains largely passive, with little focus on personalized, creative, or active learning.

The results are consistent with previous studies showing that AI tools can increase student engagement, motivation, and learning effectiveness. It has been found that the use of AI tools is strongly correlated with greater student engagement, motivation, and independent learning [8]. Students using AI achieved twice the learning growth in half the time, with greater motivation and engagement [22]. Overall, most studies show significant improvements in academic achievement, student engagement, and knowledge retention when AI tools are used [17].

This study complements existing academic literature by showing not only a generally positive attitude toward AI, but also that students in different countries evaluate AI for different reasons. It is also important to note that the study involved future teachers, making the results particularly relevant to teacher training, where the acceptance of technology can have a long-term impact on future teaching practices.

From a practical point of view, these results suggest that integrating AI into university studies should not be uniform. In the Lithuanian context, it is appropriate to emphasize the educational and pragmatic value of AI tools, such as their potential to strengthen independent learning, support differentiated education, and prepare teaching materials. Meanwhile, in the Romanian context, it makes sense to make greater use of AI's motivational and creative potential to increase student interest and active involvement in the study process. Such a differentiated strategy could contribute to a more effective application of AI in the university study process.

Despite its significant insights, the study has some limitations. First, the samples are relatively small, so the results should be interpreted with caution. Second, the data is based on students' subjective opinions, which may be influenced by their limited experience with AI tools. Future studies should use larger, more diverse samples, and examine the relationships among perceived AI quality, actual AI usage frequency, and academic performance.

Overall, the CFA results support the theoretical distinction among ergonomic, hedonic, and pragmatic dimensions, while the second-order model suggests that these dimensions can also be interpreted as components of a broader latent construct of perceived AI quality. Nevertheless, the fit indices indicate that the model is acceptable rather than definitive, so the factorial structure should be retested in larger, more heterogeneous samples.

5 Conclusions and Implications

The study revealed that students in both Lithuania and Romania perceive artificial intelligence tools positively in university studies. Although the structure of perceived AI quality is identical in both countries and comprises ergonomic, hedonic, and pragmatic dimensions, significant differences were found in the relative importance of these dimensions. In the Lithuanian context, pragmatic and ergonomic rationality predominates, indicating that students primarily value AI as a functional tool that supports learning and provides tangible academic benefits. In contrast, among Romanian students, the hedonic–motivational aspect is more prominent, reflecting the attractiveness of AI tools, emotional engagement, and their capacity to stimulate interest and motivation. These findings suggest that cultural and educational context influences not the overall positive perception of artificial intelligence, but rather the underlying reasons why these tools are evaluated favourably.

Based on the analysed multidimensional model, pre-service teachers' positions reveal a meaningful balance between professional utility and personal engagement, with a consistently positive attitude toward AI tools across all dimensions. The ergonomic dimension, reflected in accessibility and ease of use, is associated with intuitive interfaces, user comfort, and effective human–technology interaction, indicating that pre-service teachers are more favourable toward AI tools when they are seamlessly integrated into existing academic and professional workflows. The hedonic dimension highlights the importance of emotional fulfillment, stimulation, and enjoyment of use, positioning AI as a potential catalyst for creativity and innovation in educational contexts. Although this dimension appeared relatively less prominent in the Lithuanian sample, it nevertheless represents an important pathway for fostering engagement, professional pride, and originality. Finally, the pragmatic dimension underscores AI's instrumental role in supporting concrete pedagogical goals. Pre-service teachers tend to adopt a cautiously pragmatic stance, viewing AI as a highly efficient assistant rather than a replacement for human educators, and expressing stronger approval when AI tools clearly contribute to students' learning and developmental outcomes.

Ethical Statement. An anonymous, voluntary questionnaire survey given to university students served as the basis for the present study. The aim of the study, the voluntary nature of participation, the confidentiality of the data, and the participants' freedom to suspend participation at any moment prior to submission were all explained to them at the outset of the questionnaire. No personally identifiable data was gathered or kept for examination. No additional ethical issues were found, and formal ethics committee clearance was not needed in compliance with institutional guidelines for anonymous, non-interventional survey research.

References

1. Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N.A., Abazid, H., Malaeb, D., et al.: A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. Sci. Rep. **14**, Article 1983 (2024)

2. Agrawal, M., Parvez, M.: Enhancement of quality education and academic performance through mobile technology: a conceptual framework. Int. J. Adv. Educ. Res. **3**(2), 261–264 (2018). https://doi.org/10.22271/educationjournal.2018.v3.i2.72
3. Al Mashagbeh, M., Alsharqawi, M., Tudevdagva, U., Khasawneh, H.J.: Student engagement with artificial intelligence tools in academia: a survey of Jordanian universities. Front. Educ. **10**, Article 1550147 (2025). https://doi.org/10.3389/feduc.2025.1550147
4. Allam, H., Dempere, J., Akre, V., Flores, P., Parakash, P., Lakshmanna, K.: Artificial intelligence in education (AIED): implications and challenges. In: Al Kaabi, H. (ed.) Proceedings of the HCT International General Education Conference (HCT-IGEC 2023), pp. 126–140. Atlantis Press (2023). https://doi.org/10.2991/978-94-6463-286-6_12
5. Anderson, J.C., Gerbing, D.W.: Structural equation modeling in practice: a review and recommended two-step approach. Psychol. Bull. **103**(3), 411–423 (1988). https://doi.org/10.1037/0033-2909.103.3.411
6. Baker, T., Smith, L., Baker, N.: Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. Nesta. https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf (2019)
7. Chassignol, M., Khoroshavin, A., Klimova, A., Bilyatdinova, A.: Artificial intelligence trends in education: a narrative overview. Procedia Comput. Sci. **136**, 16–24 (2018). https://doi.org/10.1016/j.procs.2018.08.233
8. Das, A., Chaudhery, U., Das, N.Y.: Augmenting learning with artificial intelligence: an empirical examination of generative AI tools and student outcomes in higher education. Int. J. Res. Sci. Innovat. (IJRSI). **12**(7), 1–14 (2025). https://doi.org/10.51244/IJRSI.2025.120700041
9. Durso, S.D.O., Arruda, E.P.: Artificial intelligence in distance education: a systematic literature review of Brazilian studies. Probl. Educ. 21st Cent. **80**(5), 679–692 (2022). https://doi.org/10.33225/pec/22.80.679
10. Edwards, J.R.: Multidimensional constructs in organizational behavior research: an integrative analytical framework. Organ. Res. Methods. **4**(2), 144–192 (2001). https://doi.org/10.1177/109442810142004
11. Fornell, C., Larcker, D.F.: Evaluating structural equations models with unobservable variables and measurement error. J. Mark. Res. **18**(1), 39–50 (1981). https://doi.org/10.2307/3151312
12. Grájeda, J.A., Burgos, J., Córdova, P., Sanjinés, A.: Assessing student-perceived impact of using artificial intelligence tools: construction of a synthetic index of application in higher education. Cogent Educ. **11**(1), Article 2287917 (2024). https://doi.org/10.1080/2331186X.2023.2287917
13. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L.: Multivariate Data Analysis, 6th edn. Prentice Hall (2006)
14. Ibieta, A., Aguilera, S., Constanza-Medina, M., Sepúlveda, M.I., Castellanos-Alvarenga, L.M.: Academic use of artificial intelligence tools by university students. In: Tomczyk, Ł. (ed.) New Media Pedagogy: Research Trends, Methodological Challenges, and Successful Implementations, pp. 349–365. Springer (2025). https://doi.org/10.1007/978-3-031-95627-0_23
15. Katsantonis, A., Katsantonis, I.G.: University students' attitudes toward artificial intelligence: an exploratory study of the cognitive, emotional, and behavioural dimensions of AI attitudes. Educ. Sci. **14**(9), Article 988 (2024). https://doi.org/10.3390/educsci14090988
16. Koufteros, X.A., Babbar, S., Kaighobadi, M.: A paradigm for examining second-order factor models employing structural equation modeling. Int. J. Prod. Econ. **120**(2), 633–652 (2009). https://doi.org/10.1016/j.ijpe.2009.04.010
17. Kwak, M.: The effectiveness of AI-driven tools in improving student learning outcomes compared to traditional methods. Issues Inf. Syst. **26**(4), 233–247 (2025). https://doi.org/10.48009/4_iis_2025_120

18. Lamanauskas, V., Makarskaitė-Petkevičienė, R., Gorghiu, G., Santi, E.A., Pribeanu, C.: Exploring the students' perceived academic value in an exclusive online learning environment. Revista Romaneasca Pentru Educatie Multidimensionala. **13**(4), 1–16 (2021). https://doi.org/10.18662/rrem/13.4/467
19. Liang, H., Reiss, M.J.: The associations between students' attitudes toward AI and learning engagement: serial mediating roles of perceived autonomy and learning enjoyment. Front. Psychol. ***16***, Article 1681635 (2025). https://doi.org/10.3389/fpsyg.2025.1681635
20. Maberah, S. et al.: Students' attitudes and perceived usefulness of artificial intelligence (AI) tools in physical education. Int. J. Inf. Educ. Technol. **15**(4), 767–773 (2025). https://doi.org/10.18178/ijiet.2025.15.4.2282
21. Mels, G.: LISREL for Windows: Getting Started Guide. Scientific Software International (2006)
22. Mollick, E.R., Mollick, L.: Assigning AI: seven approaches for students, with prompts (The Wharton School Research Paper No. 2023-02). SSRN (2023). https://doi.org/10.2139/ssrn.4475995
23. Ogarcă, R.F.: Romanian students' perceptions of artificial intelligence use in higher education: benefits and challenges. Revista de Științe Politice. **88**, 91–110 (2025)
24. Pisică, A.I., Ioan, R., Bucur, L.M., Popa, A., Zaharia, R.M.: Romanian students' opinions on implementing artificial intelligence in higher education: a qualitative approach. Transform. Bus. Econ. **23**(2), 21–35 (2024)
25. Sánchez-Prieto, J.C., Cruz-Benito, J., Therón Sánchez, R., García-Peñalvo, F.J.: Assessed by machines: development of a TAM-based tool to measure AI-based assessment acceptance among students. Int. J. Interact. Multimedia Artif. Intell. **6**(4), Article 80 (2020)
26. Skalka, J. et al.: Cross-national survey data on student attitudes toward artificial intelligence. Data Brief. **62**, Article 112022 (2025). https://doi.org/10.1016/j.dib.2025.112022
27. Vieriu, A.M., Petrea, G.: The impact of artificial intelligence (AI) on students' academic development. Educ. Sci. **15**(3), Article 343 (2025). https://doi.org/10.3390/educsci15030343
28. Zorlu, Y.: The role of artificial intelligence (AI)-assisted question generation on preservice science teachers' 21st-century teaching competencies. J. Balt. Sci. Educ. **24**(5), 984–1005 (2025). https://doi.org/10.33225/jbse/25.24.984
29. Khairuddin, Z., Shahabani, N.S., Ahmad, S.N., Ahmad, A.R., Zamri, N.A.: Students' perceptions of the artificial intelligence (AI) tools as academic support. Malays. J. Soc. Sci. Hum. (MJSSH). ***9***(11), Article e003087 (2024). https://doi.org/10.47405/mjssh.v9i11.3087

Digital Inclusive Technologies in the Literacy Education of Primary School Pupils

Filip Makan[1], Andriana Shyshak[2](✉), Rostyslav Rudenskyi[2], Silvia Jakabová[3], Miroslav Tvrdoň[4], and Beáta Akimjaková[5]

[1] Faculty of Theology, Theological Institute, The Catholic University in Ružomberok, Spišská Kapitula 12, 053 04 Spišské Podhradie, Slovak Republic

[2] Department of Educology and Pedagogy, West Ukrainian National University, 46009 Ternopil, Ukraine
{a.shyshak,r.rudenskyi}@wunu.edu.ua

[3] University DTI, Sládkovičova 533/20, 018 41 Dubnica nad Váhom, Slovak Republic
jakabova@dti.sk

[4] Department of Social Work and Social Sciences, Constantine the Philosopher University, Tr. A. Hlinku 1, 948 01 Nitra, Slovakia
mtvrdon@ukf.sk

[5] Faculty of Theology, Department of Social Sciences, The Catholic University in Ružomberok, Hrabovská cesta 1A, 034 01 Ružomberok, Slovak Republic

Abstract. The purpose of the article is to define the essence of digital inclusive technologies in the context of implementing universal design in primary education, to study their potential and practical application for teaching literacy to primary school pupils. The study uses the following methods: analysis of scientific and pedagogical literature, synthesis, generalisation and systematisation of theoretical material and practical experience in the use of digital inclusive resources in the educational process to develop reading and writing skills in primary school children.

The use of digital inclusive technologies in primary school is based on the principles of accessibility, flexibility and multi-channel communication. These tools provide personalised learning paths, reduce psycho-emotional stress.

The use of platforms such as Edugames and Epic!, and the creation of personalised HTML games using artificial intelligence tools (ChatGPT, Google Gemini), makes it possible to differentiate tasks according to the individual needs of each child. These technologies can support student autonomy through synchronous feedback, gamification elements, and tolerance for mistakes, creating a situation of success for every learner in today's educational environment.

Keywords: Digital Inclusive Technologies · Literacy Training · Reading Skills · Writing Skills · Primary School Pupils · Universal Design in Education · Primary School · Artificial Intelligence · Digital Resources

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 211–221, 2027.
https://doi.org/10.1007/978-3-032-32104-6_19

1 Introduction

Digital transformation is taking place in all areas of public life, which logically necessitates the use of digital tools in education in general and in primary education in particular [1, 2]. This increases pupils' motivation to learn, keeps their attention on teaching materials for longer, and stimulates everyone's overall involvement in the active educational process.

In addition to the trend towards digitalisation, the principle of inclusiveness must be implemented throughout modern education. There is a need to seek educational solutions that will ensure equal access to knowledge for every student. The introduction of digital inclusive technologies is effective in this context. They make it possible to create an educational environment that is flexible enough to constantly consider the individual characteristics of each child.

Inclusiveness is not only defined by creating conditions for people with health problems. It is about ensuring equal access to education for everyone in all circumstances. For example, during COVID-19, the ability to switch to distance learning, minimising face-to-face contact, was also evidence of the inclusiveness of the education system. Education continued despite the pandemic; it simply changed from face-to-face to online due to the events taking place in the world at that time [2–4].

The first stage – the period of literacy training – is particularly important in the process of obtaining an education. Teaching children basic reading and writing skills using traditional tools often does not allow for an individual approach to each primary school pupil. Therefore, there is a need to apply digital inclusive technologies that are multimedia, interactive and multi-channel. It is precisely this comprehensive approach that forms the basis for developing sustainable reading and writing skills, reducing psychological and emotional difficulties in the process. In addition, it confirms the values of education on which the upbringing of the new generation is based. These include the value of equal access to education, child-centredness, psychological safety and emotional well-being, inclusion and diversity, active participation and learning motivation, sustainable development of the educational environment, and much more. Adherence to these values improves the quality of life as a complex phenomenon [5, 6].

The relevance of this topic provides an opportunity to consider the theoretical foundations of digital inclusive technologies and explore their potential for teaching literacy to younger schoolchildren. This confirms the need to study practical cases on the use of digital inclusive technologies for literacy teaching in primary schools.

2 Research Methodology

The research is theoretical, analytical and conceptual in nature. The source base comprises articles, systematic reviews, meta-analyses, materials from scientific and practical conferences, and regulatory documents. Chronologically, the study covers mainly works from 2017 to 2025, reflecting the current state of research into digital inclusive technologies in primary education; however, some earlier publications have also been included in the analysis.

Sources were selected based on the following criteria: thematic relevance to the issue of digital inclusive technologies in primary education; the inclusion of provisions regarding universal design for learning, accessibility of the educational environment, and digital tools for developing reading and writing skills; highlighting the pedagogical potential of interactive, mobile and gaming resources; the presence of generalised or empirically supported conclusions regarding the effectiveness of the relevant technologies.

The study employed the following methods: analysis – to examine academic approaches to the nature of digital inclusive technologies, their characteristics and educational potential; synthesis – to combine findings from various academic sources into a coherent understanding of the role of digital technologies in literacy education; comparison – to contrast approaches regarding the pedagogical appropriateness of using digital resources to ensure the accessibility of educational materials for all children; generalisation and systematisation – to formulate conclusions regarding the use of digital inclusive technologies for teaching literacy to younger schoolchildren.

In view of the above, the research methodology is aimed at the academic analysis, systematisation and conceptualisation of existing approaches to the use of digital inclusive technologies for teaching literacy to younger primary school pupils.

3 Characteristics of Digital Inclusive Technologies in the Context of Implementing Universal Design in Primary Education

In contemporary research on pedagogy and educational technologies, digital inclusive technologies are increasingly being considered in the context of universal design for learning (UDL). Universal design in education is "the design of objects, environments, educational programmes and services that ensures their maximum suitability for use by all persons without the need for adaptation or special design" [7]. This term is also understood as a set of principles for developing educational programmes, within which equal opportunities for learning are ensured for every individual, regardless of their individual characteristics [8].

In the context of implementing the principle of universal design in education, the use of digital inclusive technologies involves creating educational solutions that are accessible and effective for the widest possible range of students without the need for further individual adaptation [9]. According to the results of systematic reviews, the inclusiveness of digital technologies lies in their ability to provide variability in the presentation of educational material, methods of activity and forms of demonstration of learning outcomes [10].

Digital technologies designed in accordance with UDL principles enable students with different educational needs, levels of readiness and learning styles to work with the same learning material while maintaining their own pace and rhythm [11]. This pedagogical flexibility is important when organising learning in primary school. After all, there is often significant heterogeneity among children, particularly in terms of their level of linguistic, cognitive or regulatory skills. In this case, digital inclusive technologies provide the basis for organising an educational process focused on the diversity of students, promoting their independence in learning [12].

When characterising digital inclusive technologies in the context of implementing universal design in primary education, we believe that they should comply with the main principles of universal design itself and its educational context [8, 13]. Digital inclusive technologies used to work with primary school children should ensure:

- equality and accessibility of use (digital inclusive technologies should be equally accessible to all primary school pupils; there is no need to establish the existence of special educational needs or levels of cognitive abilities, as this does not affect the effectiveness of the tool);
- flexibility of use (within the digital resource, there should be a variety of ways to present educational material that can be selected through settings: learning pace, level of difficulty, approach to presenting information, etc.);
- simplicity and intuitiveness of the digital learning environment (the interface and functionality of digital educational tools should be understandable and logical for primary school children, regardless of their previous experience with such technologies, their level of digital literacy or cognitive characteristics);
- multichannel display of educational information (digital inclusive technologies must be adaptable to a variety of learning conditions; this is achieved by providing information in various forms – text, audio, visual, interactive and combined);
- tolerance for mistakes (the digital educational environment in which child users work should reduce the risks of negative consequences of their mistakes or accidental actions and provide for the possibility of quick correction; digital resources should have an option to repeat an action, receive a hint, and get feedback);
- minimising physical and psycho-emotional stress (digital inclusive technologies should be designed considering the age characteristics of primary school children, ensuring ease of use and optimal duration of interaction; they should be ergonomic and reduce excessive physical and cognitive stress during the organisation of educational activities);
- appropriate digital space for interaction and learning (the digital educational environment should provide sufficient virtual and physical space for interaction with other participants in the educational process, as well as with content and digital devices) [8, 13].

4 The Potential of Digital Inclusive Technologies for Developing Reading and Writing Skills in Primary School

Within the framework of the educational process in primary school, teachers ensure that children learn the material through its presentation and organisation of interaction. Teachers provide comprehensive pedagogical support in this process. In the context of inclusive primary education, pedagogical support is a system of actions aimed at ensuring the success of each child's learning without stigmatising them [14].

Digital inclusive technologies perform several functions in the educational process, namely: regulatory (tools should be aimed at organising educational activities), compensatory (resources should reduce the impact of difficulties on learning to read or write) and motivational (tools are designed to maintain interest in learning) [15]. For example, functions such as text-to-speech, synchronous feedback, or automatic error visualisation

allow students to independently monitor their own learning, which promotes autonomy and confidence in their own abilities.

Scientific research on literacy skills development confirms the potential of digital technologies in the development of reading and writing skills. However, their use must, of course, be pedagogically sound and methodologically balanced. In particular, the results of analyses and reviews show that computer programmes can have a positive impact on the automation of reading skills and the development of written language [16, 17]. Teachers can adjust the level of difficulty of tasks depending on the language proficiency level of the children they are working with. The use of online platforms and digital tools allows students to work through the material at their own pace and to repeat difficult elements multiple times [18].

The ability to combine text, audio support, visual cues and interactive elements within digital inclusive technologies is particularly important for teaching literacy to primary school pupils, as it promotes a deeper understanding of the learning material [19]. Digital resources enable students not only to perceive words visually, but also to hear their pronunciation and see the corresponding visual context, which improves the quality of memorisation and understanding of lexemes. This stimulates the development of associative thinking, based on which a stable reading skill is formed, which is especially important for engaging children who have difficulty mastering language through traditional methods [18].

To develop speaking skills, it is advisable to use the following resources: trainers for practical consolidation of structures; testers for identifying weaknesses; interactive textbooks that combine theory with practical tasks [20].

Among the digital technologies researchers distinguish between hardware and software tools [21]. After analysing them, we conclude that some of them are inclusive. For example, devices that provide visual and tactile interaction and are actively used in the educational process in primary schools include interactive whiteboards and displays, tablets, and smartphones.

The interactive whiteboard and display enable teachers to demonstrate teaching material using visual elements (images or videos). Students control objects displayed on the monitor by touching them with their fingers and hands or using a marker [22]. When younger schoolchildren do this, they prepare their hands for writing: for example, primary school pupils no longer practise writing letters in the air but draw them on an interactive whiteboard or interactive display. The use of letter templates or their elements is also useful in teaching writing. Such templates can be displayed on the monitor, and students learn to write them using a marker. This facilitates frontal work and the ability to explain synchronously how to overcome difficulties. The large diagonal of the monitor allows you to clearly see all the details of writing symbols when using computer presentations or video clips with algorithms of actions.

In the educational process of primary school, tablets and smartphones are personalised educational tools. They make it possible to individualise literacy training for younger pupils [23]. For example, applications downloaded onto these devices can be effective in developing basic reading skills. Such software is usually game-based, and children, eager to win the game, try to guess what is written in the hints or tasks, learning the basics of reading in the process.

Interestingly, mobile apps provide constant access to learning materials anytime, anywhere. They adapt to the learner's interests and level of knowledge (for example, apps that use artificial intelligence algorithms can automatically suggest corrections in real time, which helps learners master grammar rules). The presence of gamification elements, such as points, levels, rewards, etc., significantly increases students' motivation to learn grammar and vocabulary. In addition, mobile applications provide the ability to immediately receive personal results and assessments. This helps children learn from their mistakes and not worry about other people's evaluative judgements [20].

Digital applications and web services have the potential to create an immersive environment for children learning to read and write. Interactive books and textbooks contain audio accompaniment, animation and video, which ensures active interaction during reading. Online whiteboards allow users to comment on text together and work with stickers and formulas in real time. QR code creation services enable quick transitions between web pages, which helps to optimise the use of various additional materials for lessons and, as a result, differentiate the complexity of reading texts for each child according to their characteristics [21]. Resources for creating interactive visualisations (graphs or maps) provide a space for systematising vocabulary that children have already learned [24]. The use of virtual and augmented reality creates immersive lessons in which children can "dive" into the plot of a work [21, 22].

5 Practical Examples of Using Digital Inclusive Technologies for Literacy Training

Analysis of practical research shows that there is a wide variety of digital inclusive technologies that are used in literacy teaching in primary schools. These include interactive educational platforms, digital gaming environments, programmes for developing reading and writing skills, and tools for supporting written language, in particular technologies for converting speech into text and vice versa [25, 26]. In pedagogical practice, such technologies are implemented using interactive exercises for learning sound-letter correspondences, digital texts with the ability to change display parameters, and educational games aimed at automating reading and writing skills. It is important that these tools are equally accessible to all students in the class and do not require a modified inclusive usage scenario, which is in line with current approaches to inclusive education [27]. Here are some examples of digital inclusive technologies that are useful for teaching literacy in primary school. The tools were selected based on a qualitative analysis of academic sources and open educational resources dedicated to literacy teaching in primary schools. They were considered as representative examples based on criteria such as accessibility, pedagogical appropriateness, multimodality, suitability for primary schools, and support for diverse educational needs.

The Edugames platform (https://edugames.rozumniki.com/) from the Rozumnyky publishing house is an effective digital inclusive resource for teaching literacy in Year 1. It provides a multisensory approach and individualisation of the educational process. The inclusive potential of the platform is realised through its ability to adapt to the different cognitive needs of students. The tasks include text, audio accompaniment, images and animation. This is important for children with speech impairments or concentration

difficulties, as it simultaneously engages both visual and auditory channels of information perception. The characters in the games react emotionally to the user's actions, allowing them to see the results for themselves and correct mistakes at their own pace. The presence of a game plot and a reward system help children who find it difficult to work with traditional textbooks to concentrate [18]. In general, the game-based format of the exercises ensures that the learning process is organised with a focus on the child and their age and psychological characteristics [28].

Edugames for first graders offers tasks:

- games for recognising sounds in words (the player is shown pictures of objects (apricot, pineapple, etc.), and their task is to click on the images whose names begin with a certain sound; this game develops phonemic awareness, which is the basis for developing reading skills);
- exercises for constructing syllables and words (the student must drag the letters into the appropriate cells to form the name of the object shown on the screen; animation indicates whether the direction of movement is correct, which serves as a hint for children with impaired fine motor skills or spatial orientation);
- games for recognising letters (younger schoolchildren must collect crops only from baskets with words containing a certain letter written on them, or put together a puzzle from parts of letters, which helps them learn to recognise them visually);
- exercises for matching sound patterns and words (the student must select the correct sound pattern, composed of vowels or consonants, for the spoken or written word; these tasks are voiced, which allows the child to work independently, even if they have not yet fully mastered the skill of reading);
- sentence construction tasks (the player must create a diagram for a sentence spoken by the computer, which allows them to learn sentence structure without yet knowing the letters);
- educational interactive videos (younger schoolchildren press a button and the platform tells them about sounds and which letters represent them when reading and writing; the entire text is voiced and accompanied by images of bright objects and letters, including arrows to indicate the direction of elements when writing; this provides multi-channel information delivery and flexible control over the pace of learning); and many others.

For example, during a literacy lesson, a teacher can use the Edugames platform to set interactive tasks and then analyse the pupils' results using the teacher's tools. This makes it possible to organise differentiated learning and take each pupil's pace into account.

Given the above, the use of Edugames allows teachers to differentiate tasks. This may help every child, regardless of their abilities, to master reading and writing skills, which confirms the inclusiveness of this web resource.

The Epic! Digital Library (https://www.getepic.com/) is also a digital inclusive technology. It is designed for children aged 5 to 12. The platform provides equal access to quality content for learners with different levels of language proficiency. The service has over 40,000 resources, including traditional texts, illustrated publications, and audiobooks. This allows children with visual impairments or dyslexia to effectively perceive information through the auditory channel, or those who are not yet proficient readers to acquire the lexical richness of the language by ear [21, p. 43].

The inclusive potential of the Epic! library is realised through personalisation and support for student autonomy. The platform offers educators tools to ensure a differentiated approach. After registering, teachers can select literature according to the individual interests and current reading level of each student. Thanks to the ability to choose books independently and the absence of fear of making mistakes, children develop an internal motivation to read and build their own progress in this area. Teachers have access to monitoring tools that allow them to track the content children choose and the time they spend reading.

The practical implementation of inclusivity in Epic! is ensured by several specialised features, such as the Read-To-Me format. This option provides audio accompaniment with synchronous highlighting of each word, which is an effective tool for developing reading skills in children. High-quality visualisation facilitates the semantics of new lexemes. The thematic range of the library allows even students with specific interests to actively engage with the book.

An example of how the Epic! platform can be used in a reading lesson is that the teacher can invite pupils to work with books on the platform that match their reading level, whilst pupils with reading difficulties can use the Read-To-Me feature. This ensures that the learning process is tailored to individual needs.

Artificial intelligence tools are now actively entering the sphere of use by both adults and children. Experience in developing original teaching materials using ChatGPT (https://chatgpt.com/) and Google Gemini (https://gemini.google.com/) to create interactive tasks by generating HTML codes aimed at teaching literacy to younger schoolchildren allows us to consider artificial intelligence tools as a digital inclusive technology. After all, they implement the principle of inclusiveness in primary education – they may facilitate accessible, adaptable, and effective learning for all learners.

HTML games that use drag-and-drop technology to match graphic images of letters are effective. An example of such tasks is a game like "Match the Pairs", in which the child matches handwritten and printed letters. If the task is completed correctly, the student receives positive feedback from the game (verbal and symbolic confirmation of success), and if a mistake is made, the element automatically returns to its initial position without a negative assessment. This logic of the game helps to form a stable graphic image of the letter. Psychologically, this reduces anxiety in younger schoolchildren, which is especially important for those who find it difficult to master reading and writing.

It should be noted that in the process of creating games with handwritten symbols, it is often necessary to repeatedly correct their form in accordance with the requirements of current regulations, which is fundamentally important for preventing the formation of false graphic representations. However, the final version of such interactive tasks allows you to learn the correct form of handwritten letters. Repeated playing of the game creates conditions for their comprehensive consolidation.

Another example is games for recognising and locating letters in words, provisionally called "The Lost Letter". In such games for first-graders, artificial intelligence tools are used to create interactive tasks with minimal text content and a predominance of audio accompaniment. Instructions are provided in audio form. Players perform learning activities by selecting or dragging letters to the appropriate position. This makes the game

accessible to children who have not yet developed reading skills or who experience difficulties in this type of speech activity.

Another important aspect of inclusivity in the interactive tasks we have created using artificial intelligence tools is the multi-channel presentation of educational material. In other words, these games combine visual, auditory and motor components. An example is a game task such as "Listen and choose", in which the child listens to an audio accompaniment (the name of a letter or sound) and matches it with the corresponding graphic image. This forms phonemic perception and the basis for the development of reading competence.

For example, when preparing for a literacy lesson, a teacher can use the tools mentioned above to create an interactive HTML game for recognising letters or sounds, tailored to the level of a particular class. This makes it possible to take into account each child's individual pace of learning.

6 Conclusion

Thus, the use of digital inclusive technologies in the educational process of primary school makes it possible to create an educational environment based on universal design. Digital tools allow students to independently choose the pace, level of difficulty, or method of information perception. A multisensory approach (the presence of visual images, sound, interaction, etc.) makes the literacy learning process accessible to children, reducing the risk of their stigmatisation within the group.

The practical application of resources such as the Edugames interactive platform and the Epic! digital library demonstrates potential for developing reading and writing skills through specific game mechanics and the possibility of synchronous feedback. For educators, artificial intelligence tools (ChatGPT and Google Gemini) are environments for creating personalised teaching materials, such as interactive HTML games. This approach can facilitate the assimilation of educational material. In addition, it significantly reduces the level of psychological and emotional stress for younger schoolchildren, as each mistake becomes another step in learning something new.

In summary, we conclude that the use of digital inclusive technologies can be beneficial for literacy teaching in primary schools. It can support personalised learning, foster autonomy, and create favourable conditions for a sense of achievement among learners. However, this article is limited by its conceptual nature, so further research is needed to assess the effectiveness of these technologies in real-world learning environments.

Acknowledgement. This work was supported by the Cultural and Educational Grant Agency (KEGA) of the Ministry of Education, Research, Development and Youth of the Slovak Republic based on project number 014KU-4/2026: Family Mediation as an Effective Tool for Supporting the Cohesion of Family Relationships and the Constructive Resolution of Conflict Situations.

References

1. Alam, S., Hameed, A., Balogh, Z., Haque, S.A.: Digitalisation of education: challenges and issues in virtual pedagogical competence. J. Educ. Cult. Soc. **16**(1), 165–177 (2025). https://doi.org/10.15503/jecs2025.2.165.177

2. Kildė, L.: ESL teachers' approaches towards the acceptance of educational technology integration in non-formal education: A case from Kenya. J. Educ. Cult. Soc. **14**(1), 634–649 (2023). https://doi.org/10.15503/jecs2023.1.634.649
3. Pavlikova, M., Maturkanič, P., Akimjak, A., Mazur, S., Timor, T.: Social interventions in family in the post-COVID pandemic period. J. Educ. Cult. Soc. **14**(1), 106–123 (2023). https://doi.org/10.15503/jecs2023.1.106.123
4. Lesková, A., Lenghart, P.: Post-Covid media behaviour patterns of the generation Z members in Slovakia. J. Educ. Cult. Soc. **14**(1), 503–513 (2023). https://doi.org/10.15503/jecs2023.1.503.513
5. Kobylarek, A., Maturkanič, P.: Value order decomposition as an educational challenge. J. Educ. Cult. Soc. **16**(2), 7–14 (2025). https://doi.org/10.15503/jecs2025.3.7.14
6. Petrovič, F., et al.: Measuring the quality of life university students and its predictors as a basis for its applied research. Cross-country evidence. Sustain. Fut. **10**, 101214 (2025). https://doi.org/10.1016/j.sftr.2025.101214
7. Zakon Ukrainy "Pro osvitu" [Law of Ukraine "On Education"]. Vidomosti Verkhovnoi Rady 38–39, art. 380 (2017). https://zakon.rada.gov.ua/laws/show/2145-19#Text
8. Nosenko, Y.H.: Tekhnolohii pidtrymky universalnoho dyzainu v osviti [Technologies for supporting universal design in education]. In: Proceeding of the Scientific Reporting Conference of the Institute of Information Technologies and Learning Tools of the NAES of Ukraine. Institute of Information Technologies and Learning Tools of the NAES of Ukraine, Kyiv (2018) https://lib.iitta.gov.ua/id/eprint/711688/
9. Nochovna, N.V., Riabokin, N.O., Kukharchuk, O.: P.: Inkliuzyvni mozhlyvosti tsyfrovykh platform u protsesi navchannia inozemnoi movy [inclusive possibilities of digital platforms in the process of learning a foreign language]. In: Novitni tendentsii navchannia inozemnoi movy za profesiinym spryamuvanniam, pp. 167–168, Kherson State Maritime Academy, Odesa. https://digimar.si/wp-content/uploads/2025/11/Proceedings_Conference_KSMA_25-1.pdf#page=168 (2025)
10. Navas-Bonilla, C.R., Guerra-Arango, J.A., Oviedo-Guado, D.A., Murillo-Noriega, D.E.: Inclusive education through technology: a systematic review of types, tools and characteristics. Front. Educ. **10**, 1–22 (2025). https://doi.org/10.3389/feduc.2025.1527851
11. Almeqdad, Q.I., Alodat, A.M., Alquraan, M.F., Mohaidat, M.A., Al-Makhzoomy, A.K.: The effectiveness of universal design for learning: a systematic review of the literature and meta-analysis. Cogent Educ. **10**(1), 1–24 (2023). https://doi.org/10.1080/2331186X.2023.2218191
12. Diputra, Y., Husin, M., Fadhilah, Karimi, A., Abubakar Yunusa, A., Muthma'innah, M.: Factors shaping student readiness for online learning: technology, perception, and lecturer roles. J. Educ. Cult. Soc. **16**(1), 365–387 (2025). https://doi.org/10.15503/jecs2025.2.365.387
13. Mace, R.L., Hardie, G.J., Place, J.P.: Accessible Environments toward Universal Design. Center for Universal Design, Raleigh, NC (1991)
14. Nordström, T., Andersson, U.B., Fälth, L., Gustafson, S.: Teacher inquiry of using assessments and recommendations in teaching early reading. Stud. Educ. Eval. **63**(3), 9–16 (2019). https://doi.org/10.1016/j.stueduc.2019.06.006
15. Pidaeva, S.B., Agzamova, Z., Izatullaeva, X., Xidirova, Z.P., Kakharova, N.I.: Game technologies in teaching a foreign language. Int. J. Early Childhood Spec. Educ. **14**(4), 1536–1540 (2022) https://www.int-jecse.net/media/article_pdfs/1536-1540.pdf
16. Cheung, A.C.K., Slavin, R.E.: The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: a meta-analysis. Educ. Res. Rev. **9**, 88–113 (2013). https://doi.org/10.1016/j.edurev.2013.01.001
17. Jamshidifarsani, H., Garbaya, S., Lim, T., Blazevic, P., Ritchie, J.M.: Technology-based reading intervention programs for elementary grades: an analytical review. Comput. Educ. **128**, 427–451 (2019). https://doi.org/10.1016/j.compedu.2018.10.003

18. Krokhmalna, H., Halaiko, M.: Formuvannia leksychnoi kompetentnosti shkoliariv u konteksti novoi ukrainskoi shkoly: intehratsiia tradytsiinykh i suchasnykh pedahohichnykh tekhnolohii [Formation of lexical competence of schoolchildren in the context of the New Ukrainian School: integration of traditional and modern pedagogical technologies]. Molodyi vchenyi. **4**(135), 78–83 (2025). https://doi.org/10.32839/2304-5809/2025-4-135-17
19. Coyne, P., Pisha, B., Dalton, B., Zeph, L.A., Smith, N.C.: Literacy by design: a universal design for learning approach for students with significant intellectual disabilities. Remedial Spec. Educ. **33**(3), 162–172 (2012). https://doi.org/10.1177/0741932510381651
20. Lebedieva, N., Perva, A.-O., Romaniuk, K.: Mobilni dodatky yak instrument dlia pokrashchennia hramatychnykh navychok uchniv [Mobile applications as a tool for improving students' grammatical skills]. J. Cross-Cult. Educ. **3**, 45–54 (2024). https://doi.org/10.31652/2786-9083-2024-3-45-54
21. Shvardak, M.V.: Tsyfrovi interaktyvni tekhnolohii v osvitnomu protsesi pochatkovoi shkoly [Digital interactive technologies in the educational process of primary school]. Sci. J. Khortytsia Natl. Acad. (Series: Pedagogy. Social Work). **1**(8), 39–48 (2023). https://doi.org/10.51706/2707-3076-2023-8-3
22. Shvardak, M.V.: Osvitni trendy v umovakh Novoi ukrainskoi shkoly [Educational trends in the conditions of the New Ukrainian School]. Naukovyi chasopys Natsionalnoho pedahohichnoho universytetu imeni M. P. Drahomanova. Seriia 5. Pedahohichni nauky: realii ta perspektyvy. **89**, 136–140 (2022). https://doi.org/10.31392/NPU-nc.series5.2022.89.28
23. Pobryzghaieva, V., Nalyvaiko, O.: Tsyfrovi instrumenty heimifikatsii navchalnoho protsesu uchniv pochatkovoi shkoly v umovakh dystantsiinoho navchannia [Digital tools for gamification of the learning process of primary school students in terms of distance learning]. Osvitolohichnyi dyskurs. **45**(2), 25–35 (2024). https://doi.org/10.28925/2312-5829.2024.2.3
24. Kartashova, L.A.: Tsyfrovi tekhnolohii dlia stvorennia navchalnoho seredovyshcha novoho pokolinnia yak ekosystemy [Digital technologies for creating a new generation learning environment as an ecosystem]. In: Proceeding of the All-Ukrainian Scientific and Practical Seminar "Digital Competence of a Modern Teacher of a New Ukrainian School: Safe Environment for Students and Teachers: Challenges and Practical Solutions", pp. 28–31. Institute for Digitalisation of Education of the NAES of Ukraine, Kyiv (2022) https://lib.iitta.gov.ua/id/eprint/731095
25. Kambouri, M., Wilson, T., Pieridou, M., Quinn, S.F., Liu, J.: Making partnerships work: proposing a model to support parent-practitioner partnerships in the early years. Early Childhood Educ. J. **50**, 639–661 (2022). https://doi.org/10.1007/s10643-021-01181-6
26. Witzel, B. et al.: Digital game-based spelling intervention for children with spelling deficits: a randomized controlled trial. Learn. Instr. **89**, 101842 (2024). https://doi.org/10.1016/j.learninstruc.2023.101842
27. Samaniego López, M.V., Orrego Riofrío, M.C., Barriga-Fray, S.F., Paz Viteri, B.S.: Technologies in inclusive education: solution or challenge? A systematic review. Educ. Sci. **15**(6), 715 (2025). https://doi.org/10.3390/educsci15060715
28. Bilous, I., Shcherbiak, I., Maslak, V., Alnsour, R., Jarmoch, E.Z., Hubková, S.: Game-based technologies as a tool for developing communicative competence in primary school students within a student-centered educational context. J. Educ. Cult. Soc. **16**(2), 585–603 (2025). https://doi.org/10.15503/jecs2025.3.585.606

Hierarchical Object Detection in Historical Handwritten Substitution Cipher Keys Using YOLOv11

Stanislav Marochok(✉) and Michal Balogh

Faculty of Electrical Engineering and Information Technology, Institute of Computer Science and Mathematics, Slovak Technical University in Bratislava, Bratislava, Slovakia
{stanislav.marochok,xbalogh}@stuba.sk

Abstract. Digitizing historical handwritten nomenclator cipher keys can be treated as a modular pipeline in which object detection extracts the visual units required for later reconstruction and interpretation of the cipher keys. The keys in our study follow a three-level structure of sections, plaintext–ciphertext pairs, and elements: sections may be horizontal or vertical, and elements span plaintext tokens from letters to n-grams, words, and short phrases, as well as ciphertext tokens including letters, digits, and non-alphabetic symbols or writer-specific glyphs. We compiled and annotated 76 cipher-key pages with train/validation/test splits that preserve variation in layout, handwriting style, age, degradation, and background color. We evaluate YOLOv11 on our dataset under low-data conditions across all three hierarchy levels. We compare a single detector trained jointly on all levels, separate detectors per level, and sequential fine-tuning that transfers from element detection to higher-level structures, while also varying model size. We also test initialization with and without handwriting-domain pretraining.

Keywords: historical documents · handwritten · cipher keys · object detection · YOLOv11

1 Introduction

Historical handwritten nomenclator cipher keys constitute primary sources for the study of historical cryptography and diplomatic communication. Many are preserved as digitised page images in archival collections and public portals, enabling large-scale analysis but posing challenges for automated processing [7,19]. Digitisation aims to transform such images into structured, machine-readable representations suitable for indexing, comparison, and application in decryption workflows.

We model digitisation as a modular pipeline in which object detection performs the initial structural decomposition of a page. The considered cipher keys follow a three-level hierarchy of *sections*, *plaintext–ciphertext pairs*, and *elements*.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 222–234, 2027.
https://doi.org/10.1007/978-3-032-32104-6_20

Sections are arranged horizontally or vertically and contain grouped pairs. Elements correspond to individual plaintext tokens (letters, n-grams, words, short phrases) and ciphertext tokens (letters, digits, non-alphabetic symbols, writer-specific glyphs).

Using YOLOv11 [16], we evaluate detection performance at all hierarchy levels under low-data conditions. The dataset consists of 76 annotated cipher-key pages from HCPortal, split into training, validation, and test subsets preserving variation in layout, handwriting, age, degradation, and background colour [7]. We compare three training regimes - multi-class, single-class, and progressive fine-tuning (element→pair→section; element→section) - across five model sizes and two initialization variants, including handwriting-domain pretraining on IAM [18].

2 Related Work

Research on historical encrypted manuscripts spans structural cryptanalysis, symbol-level recognition, and document layout analysis.

Empirical analysis of early modern cipher keys from the Hessisches Staatsarchiv Marburg demonstrates that many nomenclator systems exhibit systematic structural regularities affecting cryptanalytic difficulty [10]. Among 93 examined keys, numeric alphabets often allow separation into distinguishable number intervals, and homophonic tables display ordered rows, column-wise sequences, and fixed-length patterns. The historical development and typology of cipher systems, including their multi-layered graphical organization and subdivision into encryption subsystems, are further examined in [11], reinforcing the hierarchical structure of nomenclator keys.

A digitization workflow for encrypted manuscripts combining archival acquisition, symbol annotation, Mask R-CNN detection, and transcription post-processing is presented in [8]. To address dense symbol layouts, pages are divided into 128×128 blocks prior to inference, and reading order is reconstructed from bounding-box statistics. While achieving high accuracy for digit-only ciphertext, the approach operates at symbol level. The HHCS dataset provides curated cipher symbol annotations for machine learning tasks [9], strengthening data-driven research but remaining focused on isolated symbol detection.

Line segmentation has been addressed using transfer learning with YOLOv8 models fine-tuned on encrypted manuscripts [21], achieving an F1 score of 0.9423. More broadly, document layout analysis decomposes pages into regions and structural relations [14]. Recent approaches predict hierarchical relations among detected entities to reconstruct document structure [5,6]. Multi-task models jointly predict text regions and separators in historical handwriting [20,22], and instance-based methods treat lines or regions as objects [15].

Cipher keys differ from conventional documents by combining heterogeneous symbol systems, tabular subdivisions, and semantically linked but visually separated regions. Existing approaches focus on symbols or lines and do not explicitly model hierarchical layout components such as plaintext–ciphertext pairs

and sections. The present work addresses detection across these structural levels, treating cipher keys as structured graphical systems rather than flat symbol collections.

3 Dataset

The dataset consists of 76 digitised page images of historical handwritten nomenclator cipher keys collected primarily via the Portal of Historical Ciphers [7,12,13]. Each page is treated as a full document image and annotated at three hierarchical levels: sections, plaintext–ciphertext pairs, and elements. The dataset targets low-data evaluation at page level while exhibiting high object density within pages, particularly at the element level.

Pages were selected to ensure clear nomenclator structure and reliable bounding-box annotation. Included documents exhibit tabular organisation and repeated plaintext–ciphertext correspondences, with ciphertext represented by letters, digits, or non-alphabetic symbols. Severely degraded pages preventing consistent annotation were excluded. The resulting set spans heterogeneous handwriting styles, ink intensity, background colour, and degradation effects, including common acquisition artefacts.

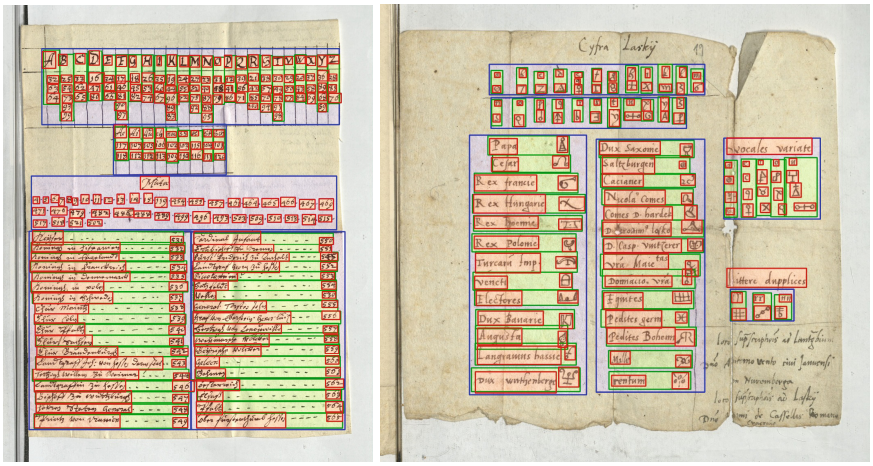
Annotations follow a strict hierarchy. **Sections** are the largest units and occur in horizontal or vertical orientation, containing grouped plaintext–ciphertext pairs. **Pairs** represent local correspondences within sections. **Elements** correspond to individual plaintext tokens (letters, n-grams, words, short phrases) and ciphertext tokens (letters, digits, symbols, writer-specific glyphs). All annotations are axis-aligned bounding boxes in pixel coordinates, with each element assigned to exactly one pair and each pair to one section. Figure 1 illustrates annotated examples.

The dataset is split at page level into training/validation/test subsets (53/15/8 pages), preserving diversity in section orientation, layout density, and visual appearance. Near-duplicate scans are confined to a single subset. The same split is used across all experiments.

Although the dataset is limited to 76 pages, it remains object-dense, particularly at the element level, providing a large number of training instances within each image. Nevertheless, the restricted number of documents limits coverage of layout variability across archival collections. The reported results therefore reflect performance under controlled low-data conditions, and generalization to structurally different cipher keys remains a subject of future work.

4 Methodology

The objective is to determine the most effective detector configuration for each hierarchy level of handwritten nomenclator cipher keys. Detection is evaluated independently for *elements*, *plaintext–ciphertext pairs*, and *sections*, using identical images, splits, and evaluation protocol; only the training regime, model size, and initialization vary. Results are ranked by $AP_{50:95}$.



HLA-HStAM, f. 4 d, Nr. 1218.[3] HLA-HStAM, f. 3, Nr. 3123.[4]

Fig. 1. Annotated cipher-key pages showing sections (blue), pairs (green), and elements (red). (Color figure online)

Given an image I, the task is to localize axis-aligned bounding boxes $B = \{b_i\}$ for one hierarchy level. Sections correspond to large horizontal or vertical layout regions, pairs represent local plaintext–ciphertext correspondences, and elements correspond to individual plaintext or ciphertext tokens. Although annotations are nested (elements $\subset$ pairs $\subset$ sections), each level is trained and evaluated independently.

All experiments use YOLOv11 [16] in five sizes (n, s, m, l, x) with input resolution 640 px. YOLOv11 is selected due to its strong performance in dense object detection scenarios and its ability to scale across model sizes while maintaining efficiency under low-data conditions. The prediction head is configured either for three classes (MC) or a single class (SC, PSC). Three training regimes are evaluated. **Multi-class (MC)** trains a single detector jointly on elements, pairs, and sections using shared parameters, with results indexed by evaluation level (MC_E, MC_P, MC_S). **Single-class (SC)** trains separate detectors for each hierarchy level, isolating optimization per target. **Progressive single-class (PSC)** performs staged fine-tuning with weight transfer, starting from an element-level model (SC_E). Evaluated trajectories are:

- PSC_P: SC_E → CK-P,
- PSC_S: SC_E → CK-S,
- PSC_P_S: SC_E → CK-P → CK-S.

PSC_E is identical to SC_E and is not treated separately.

The experimental design varies four factors: hierarchy level (E, P, S), training regime (MC, SC, PSC), model size (n–x), and initialization. Two initialization variants are used: BASE (standard YOLOv11 weights) and IAM, where the

model backbone is initialized using weights pretrained on the IAM handwriting dataset [18]. The detection head is randomly initialized, and all layers are fine-tuned during training.

Models are trained for up to 400 epochs (batch size 4, 640 px) using AdamW with cosine learning-rate decay and early stopping (patience 20); EMA checkpoints are enabled and augmentation is limited to mild geometric and photometric transforms. Detection budgets reflect object density: `max_det`=1300 (MC) and 1000/400/300 for elements/pairs/sections. PSC uses staged optimization with BASE learning rates of 1×10^{-4} (CK-E), 5×10^{-5} (CK-P), and 1×10^{-5} (CK-S); IAM-PSC uses 5×10^{-4} for CK-E followed by the same schedule. Stage durations are 35/25/20 epochs (CK-E/CK-P/CK-S), with partial backbone freezing in CK-P and early CK-S.

Evaluation follows the COCO bounding-box protocol [17]. Ground truth annotations are converted from `LabelMe` format. The primary metric is $\mathrm{AP}_{50:95}$ (IoU 0.50–0.95, step 0.05), with AP_{50}, AP_{75}, and $\mathrm{AR@}\{1, 10, \mathrm{maxDet}\}$ also reported. Inference uses `imgsz` = 640, `iou` = 0.70, and `conf` = 0.001. Level-specific budgets are $\mathrm{maxDet}_E = 1000$, $\mathrm{maxDet}_P = 400$, and $\mathrm{maxDet}_S = 300$; multi-class models use the maximum budget to enable per-class evaluation, whereas single-class models use the corresponding level budget. Metrics are computed directly from accumulated COCO tensors at the designated maxDet.

5 Results

This section reports detection performance at three hierarchical levels: elements, pairs, and sections. Results are presented using standard object detection metrics and are reported separately for BASE and IAM pretrained models. For each level, configurations are ranked by AP50-95.

5.1 Elements Detection

Table 1 reports element-level detection, where elements constitute the smallest structural units and provide the geometric foundation for higher hierarchy levels. The strongest overall configuration is the IAM-pretrained multi-class extra-large model MCE^x, achieving $\mathrm{AP}50:95 = 0.442743$, $\mathrm{AP}50 = 0.862885$, and $\mathrm{AR@maxDet} = 0.523412$, improving over its BASE counterpart (0.418865). In the BASE regime, the highest result is obtained by SCE^x with $\mathrm{AP}50:95 = 0.426273$, slightly surpassing BASE MCE^x; however, this advantage does not persist under IAM initialisation, where SCE^x decreases to 0.408331. Overall, IAM pretraining improves $\mathrm{AP}50:95$ in seven of ten configurations, with more pronounced gains for lower-capacity models (e.g., SCE^n: +0.054843; MCE^s: +0.029815). For higher-capacity models, the effect is configuration-dependent: MCE^x benefits from IAM, whereas SCE^x and MCE^l exhibit slight reductions. In several instances, AP50 and recall increase despite small decreases in $\mathrm{AP}_{50:95}$, suggesting improved coarse localisation accompanied by minor degradation under stricter IoU thresholds.

Capacity scaling trends differ between regimes. In BASE, performance increases with model size for both MC and SC, indicating effective utilisation of additional parameters. IAM reduces performance gaps between smaller and larger models, particularly improving nano and small variants. Differences between MC and SC remain modest: SC slightly outperforms MC at large capacity in BASE, while MC becomes superior under IAM. Across configurations, AP50 values remain high (above 0.75 except for nano variants), confirming stable coarse detection, whereas the primary limitation lies in strict localisation accuracy, reflected in the persistent gap between AP50 and AP50 : 95. Qualitative examples produced by the best-performing MCE^x (IAM) configuration are shown in Fig. 2, illustrating reliable element detection with residual boundary deviations at higher IoU thresholds.

Table 1. Element detection (BASE/IAM), sorted by AP50–95.

#	Config	AP50–95	AP50	AR@maxDet
1	MC_E^x	0.418865/**0.442743**	0.826388/**0.862885**	0.497512/**0.523412**
2	SC_E^x	**0.426273**/0.408331	**0.854733**/0.813628	**0.510218**/0.498712
3	SC_E^m	**0.419970**/0.415970	0.841385/**0.866617**	0.506575/**0.508974**
4	SC_E^s	0.407603/**0.414129**	0.844417/**0.847365**	0.503687/**0.514127**
5	SC_E^l	0.395398/**0.413040**	0.780095/**0.837133**	0.481653/**0.507508**
6	MC_E^l	**0.412057**/0.410873	**0.841552**/0.837003	0.500844/**0.508352**
7	MC_E^m	0.380390/**0.397263**	0.801307/**0.831514**	0.458774/**0.496357**
8	MC_E^s	0.356304/**0.386119**	0.790071/**0.813748**	0.446513/**0.478943**
9	SC_E^n	0.291828/**0.346671**	0.722084/**0.754670**	0.417948/**0.460462**
10	MC_E^n	0.303698/**0.324925**	0.751615/**0.754085**	0.419458/**0.434562**

5.2 Pairs Detection

Table 2 reports plaintext–ciphertext pair detection, capturing spatial grouping beyond individual tokens. The highest performance is achieved by the progressively trained extra-large model PSC_P^x under IAM initialisation, reaching AP50 : 95 = 0.707037, AP50 = 0.989535, and AR@maxDet = 0.756901. This corresponds to a substantial gain of +0.509534 AP over its BASE counterpart (0.197503), which represents the weakest configuration in the table. While IAM-pretrained PSC clearly dominates at the top, other strong IAM configurations cluster in the 0.53–0.58 AP range, including PSC_P^l = 0.579468 and MC_P^x = 0.576587. In contrast, BASE results are more tightly concentrated between 0.47 and 0.53 AP, indicating limited architectural differentiation in the absence of handwriting pretraining. Overall, IAM pretraining improves AP50 : 95 in 10 out of 15 configurations, often substantially (e.g., MC_P^n: +0.054408; SC_P^x: +0.064347;

(a) HLA–HHStaW, f. 130i, Nr. 1576 [1]

(b) HLA–HStAD, f. G 3, Nr. 21/4 [2]

(c) HLA–HStAM, f. 4d, Nr. 1218 [3]

Fig. 2. Qualitative element-level detection using the best-performing configuration (MC_E^x, IAM). Ground truth boxes are shown in green and predictions in red. (Color figure online)

PSC_P^m: +0.093451). These improvements are consistently reflected in both AP50 and recall, demonstrating that IAM enhances not only coarse localisation but also performance under stricter IoU thresholds.

The interaction between progressive training (PSC) and IAM initialisation is particularly pronounced. Under BASE, PSC does not consistently outperform MC or SC and exhibits a collapse in the PSC_P^x configuration. Under IAM, however, PSC becomes dominant: PSC_P^x clearly exceeds all alternatives, while PSC_P^l and PSC_P^m rank among the strongest models. Differences between MC and SC remain moderate across regimes; for instance, in BASE, MC_P^x (0.526049) and SC_P^x (0.473048) are comparable, and under IAM both improve with a reduced performance gap outside PSC variants. Model capacity scaling is more evident in IAM, where x and l variants generally outperform smaller models, although IAM reduces capacity gaps relative to BASE. Furthermore, for strong IAM configurations, the gap between AP50 and AP50 : 95 is smaller than in BASE, indicating stable coarse detection with residual errors concentrated primarily at stricter localisation thresholds. Qualitative pair-level detections obtained with the best-performing PSC_P^x (IAM) model are presented in Fig. 3, illustrating consistent plaintext–ciphertext grouping with only minor boundary deviations in densely written regions.

5.3 Sections Detection

Table 3 reports section-level detection, targeting large layout regions defined primarily by global arrangement rather than stroke-level clusters. The highest strict localisation performance is achieved in the BASE regime by MC_S^s

Table 2. Pair detection (BASE/IAM), sorted by AP50–95.

#	Config	AP50–95	AP50	AR@maxDet
1	PSC_P^x	0.197503/**0.707037**	0.559233/**0.989535**	0.365012/**0.756901**
2	PSC_P^l	0.523616/**0.579468**	0.892611/**0.914471**	0.623487/**0.651332**
3	MC_P^x	0.526049/**0.576587**	0.854994/**0.894489**	0.607627/**0.639225**
4	PSC_P^m	0.469221/**0.562672**	0.827748/**0.879177**	0.563196/**0.622760**
5	PSC_P^s	0.501274/**0.549789**	0.850322/**0.879805**	0.578935/**0.629782**
6	SC_P^x	0.473048/**0.537395**	0.823953/**0.887709**	0.559201/**0.619249**
7	MC_P^l	0.517217/**0.535005**	**0.869043**/0.860508	0.601090/**0.620218**
8	MC_P^s	**0.523032**/0.478424	**0.851431**/0.809891	**0.593947**/0.571792
9	MC_P^m	**0.516368**/0.510349	0.832591/**0.856546**	0.575424/**0.603390**
10	SC_P^l	0.484346/**0.512818**	0.820297/**0.879527**	0.565981/**0.617070**
11	SC_P^s	**0.502915**/0.475256	**0.846127**/0.829118	0.569613/**0.570460**
12	SC_P^m	**0.490953**/0.487074	0.820520/**0.829490**	0.560775/**0.572881**
13	PSC_P^n	0.466077/**0.471455**	0.813272/**0.866036**	0.542615/**0.573002**
14	MC_P^n	0.416858/**0.471266**	0.788888/**0.844092**	0.536320/**0.572881**
15	SC_P^n	**0.457546**/0.457186	0.794880/**0.827689**	0.539831/**0.559927**

with $AP50:95 = 0.675996$, $AP_{50} = 0.914457$, and AR@maxDet $= 0.735897$, exceeding all IAM configurations. Other strong BASE results include PSC_{PS}^n (0.670703) and SC_S^s (0.658436). The strongest IAM configuration is PSC_{PS}^x with $AP50:95 = 0.604855$, which remains below the BASE maximum. In contrast to element- and pair-level detection, IAM pretraining frequently reduces section-level performance. For example, MC_S^s decreases from 0.675996 to 0.288915, and SC_S^s from 0.658436 to 0.569511, with degradations affecting both $AP50:95$ and AP_{50} in several configurations. The unusually low BASE PSCxP performance indicates unstable optimization at large capacity without handwriting pretraining.

However, progressive training via pairs (PSC_{PS}) shows a different trend under IAM, where performance improves (e.g., PSC_{PS}^x: +0.158777), suggesting that staged adaptation partially mitigates the negative impact of handwriting pretraining at this hierarchy level. Section detection appears to depend primarily on large-scale layout cues rather than fine-grained feature adaptation, as reflected by the fact that the best-performing configuration is the small MC_S^s variant rather than a higher-capacity model. Capacity scaling is therefore less consistent than at lower hierarchy levels, and differences between MC and SC remain modest in BASE. Direct progressive training to sections (PSC_S) does not consistently outperform standard approaches. Across configurations, AP50 values remain high (often exceeding 0.85), indicating stable coarse region detection, while performance limitations are concentrated at stricter IoU thresholds, reflecting variability in precise boundary placement. Representative detec-

tions obtained with the best-performing MC_S^s (BASE) configuration are shown in Fig. 4, illustrating correct identification of large layout regions with typical boundary deviations along whitespace-defined borders.

6 Discussion

The results show that hierarchical detection of handwritten cipher keys behaves differently at each structural level, despite using a shared detection architecture. Element detection is primarily limited by strict localization accuracy: high AP50 values indicate reliable detection, while lower AP50:95 reflects sensitivity to boundary deviations in dense layouts with irregular spacing. Handwriting-domain initialization (IAM) improves stability, particularly for smaller models, confirming the importance of fine-grained stroke features.

(a) HLA–HStAM, Best. 4d, Nr. 1219 [3] (b) HLA–HHStaW, f. 130i, Nr. 1576 [1] (c) HLA–HStAM, f. 4d, Nr. 1218 [3]

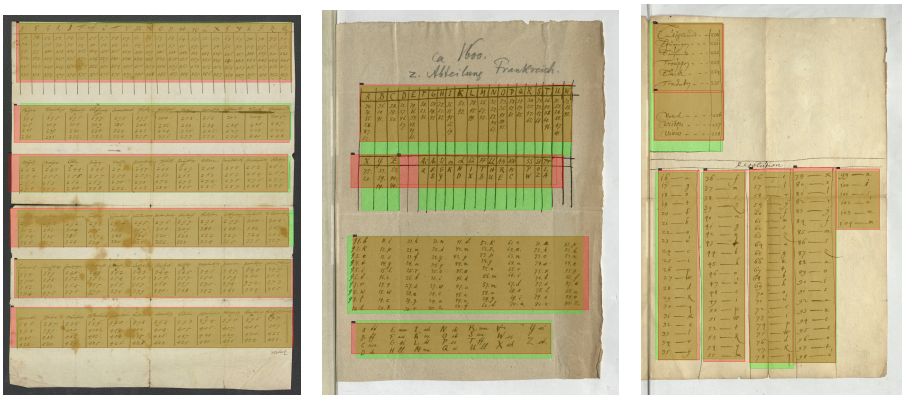
Fig. 3. Qualitative plaintext–ciphertext pair detection using the best-performing configuration (PSC_P^x, IAM). Ground truth is shown in green and predictions in red. Examples illustrate correct grouping of multi-token pairs, alignment consistency, and boundary deviations at dense or degraded regions. (Color figure online)

Pair detection achieves the highest performance and benefits most from progressive training. Transferring from element-level representations enables more reliable spatial grouping, indicating that mid-level structures are best learned through staged adaptation. The combination of progressive training and IAM initialization further improves performance, suggesting that curriculum-style optimization aligns well with the hierarchical structure. In contrast, section detection depends primarily on global layout rather than local visual features. IAM pretraining does not consistently improve results and may degrade strict localization accuracy, with errors often arising from ambiguous boundaries defined by whitespace or alignment rather than explicit separators.

Table 3. Section detection (BASE/IAM), sorted by AP50–95.

#	Config	AP50–95	AP50	AR@maxDet
1	MC_S^s	**0.675996**/0.288915	**0.914457**/0.662160	**0.735897**/0.512821
2	PSC_{PS}^n	**0.670703**/0.544959	**0.968821**/0.853626	**0.753846**/0.674359
3	SC_S^s	**0.658436**/0.569511	**0.889884**/0.848681	**0.771795**/0.684615
4	MC_S^n	**0.620787**/0.545854	**0.917656**/0.865405	**0.671795**/0.643590
5	PSC_S^n	**0.614429**/0.527749	**0.889806**/0.853973	**0.735897**/0.661538
6	SC_S^m	**0.607995**/0.434062	**0.870315**/0.826971	**0.738462**/0.643590
7	PSC_{PS}^x	0.446078/**0.604855**	0.768985/**0.924398**	0.641026/**0.730769**
8	SC_S^n	0.543512/**0.598958**	0.760765/**0.922010**	**0.715385**/0.707692
9	PSC_S^s	0.581540/**0.590169**	0.918307/**0.926784**	**0.694872**/0.689744
10	MC_S^l	**0.589901**/0.430766	**0.895862**/0.750275	**0.705128**/0.584615
11	PSC_{PS}^m	0.468284/**0.587664**	0.757994/**0.875035**	**0.705128**/0.682051
12	PSC_{PS}^l	**0.578751**/0.495405	**0.911799**/0.884279	**0.735897**/0.628205
13	MC_S^x	**0.560191**/0.440450	**0.845114**/0.761918	**0.694872**/0.602564
14	PSC_S^x	**0.543934**/0.374069	**0.922654**/0.645354	**0.669231**/0.628205
15	PSC_S^l	**0.535222**/0.426230	**0.917067**/0.804596	**0.658974**/0.628205
16	PSC_{PS}^s	**0.527271**/0.519343	**0.900332**/0.900303	0.628205/**0.669231**
17	MC_S^m	**0.513841**/0.362165	**0.818892**/0.679845	**0.651282**/0.569231
18	SC_S^l	**0.512770**/0.408913	**0.865868**/0.764824	**0.712821**/0.612821
19	SC_S^x	0.476240/**0.486252**	0.800067/**0.856266**	**0.643590**/0.628205
20	PSC_S^m	**0.265140**/0.231023	**0.509522**/0.482795	**0.600000**/0.530769

Across all levels, performance limitations are concentrated at stricter IoU thresholds, reflecting boundary uncertainty rather than detection failure, particularly in dense regions and section-level predictions. Although the study focuses on YOLOv11, the findings can be interpreted in relation to other detector families: region-based approaches such as Mask R-CNN typically provide stronger localization but require more data and computation, making them less suitable for the low-data setting considered here. The results therefore isolate the effect of hierarchical training strategy under a controlled architecture rather than performing exhaustive cross-model comparison. Overall, detection performance depends on aligning optimization strategy with structural abstraction, as a single approach is insufficient across hierarchy levels.



(a) HLA–HHStaW, f. 130i, Nr. 1576 [1]

(b) HLA–HStAM, Best. 4d, Nr. 1219 [3]

(c) HLA–HStAM, f. 4d, Nr. 1218 [3]

Fig. 4. Qualitative section-level detection using the best-performing configuration (MC^s_S, BASE). Ground truth sections are shown in green and predictions in red. Examples illustrate large layout regions with whitespace-defined boundaries and minor boundary deviations at strict IoU thresholds. (Color figure online)

7 Conclusion

This work shows that hierarchical object detection of historical handwritten substitution cipher keys is feasible under low-data conditions when training strategy is aligned with structural level. Element detection benefits primarily from handwriting-aware priors and sufficient model capacity. Pair detection benefits most from progressive adaptation anchored in element-level representations. Section detection performs best when optimized with layout-oriented bias rather than stroke-focused pretraining.

Strict localization accuracy reaches competitive levels across all three hierarchy levels, demonstrating that cipher keys can be decomposed into meaningful graphical units using modern detection architectures. The results support a modular digitization strategy in which element detection serves as the foundational learned component, while higher-level structural organization is derived through staged adaptation or geometry-based reasoning. Such alignment between optimization strategy and structural abstraction provides a stable basis for downstream reconstruction, transcription, and decryption workflows. The results suggest that hierarchical object detection in structured cipher documents benefits from level-specific optimization strategies rather than uniform multi-scale training.

Funding Information. This research was supported by the Ministry of Education, Science, Research, and Sport of the Slovak Republic through grant VEGA 2/0054/24.

References

1. HLA-HHStaW. Hessisches Landesarchiv — Hessisches Hauptstaatsarchiv Wiesbaden, fond Walramisches Hausarchiv (130 i), nr. 1576
2. HLA-HStAD. Hessisches Landesarchiv — Hessisches Staatsarchiv Darmstadt, fond Gesandtschaften (g 3), nr. 21/3
3. HLA-HStAM. Hessisches Landesarchiv — Hessisches Staatsarchiv Marburg, fond Politische Akten nach Philipp dem Großmütigen: Kanzlei- und Geheimeratskorrespondenz (4d), nr. 1218; nr. 1219
4. HLA-HStAM. Hessisches Landesarchiv — Hessisches Staatsarchiv Marburg, fond Politisches Archiv Landgraf Philipps des Großmütigen (3), nr. 3123
5. Detect-order-construct: a tree construction based approach for hierarchical document structure analysis. Pattern Recogn. (2024). https://doi.org/10.1016/j.patcog.2024.110836
6. UniHDSA: a unified relation prediction approach for hierarchical document structure analysis. Pattern Recogn. (2025). https://doi.org/10.1016/j.patcog.2025.111617
7. Antal, E.: Portal of Historical Ciphers (2019). https://hcportal.eu/. Accessed 27 Feb 2025
8. Antal, E., Marák, P.: Automated transcription of historical encrypted manuscripts. Tatra Mt. Math. Publ. **82**(2), 65–86 (2023)
9. Antal, E., Marák, P., Mikuš, F.: HHCS: a dataset of cipher symbol annotations from handwritten historical encrypted documents for machine learning tasks. IEEE Access **14**, 9226–9240 (2026)
10. Antal, E., Mírka, J.: Wrong design of cipher keys: analysis of historical cipher keys from the Hessisches Staatsarchiv Marburg used in the Thirty Years' War. In: HistoCrypt 2022: 5th International Conference on Historical Cryptology. pp. 1–11. Linköping University Electronic Press, Amsterdam (2022)
11. Antal, E., Mirka, J., Kováč, D.: Development of obfuscation techniques in Vienna during the early modern era. Cryptologia (2025). Accessed 01 May 2025
12. Antal, E., Zajac, P.: HCPortal overview. In: HistoCrypt 2020: 3rd International Conference on Historical Cryptology, pp. 18–20. Linköping University Electronic Press (2020)
13. Antal, E., Zajac, P.: HCPortal modules for teaching and promoting cryptology. In: HistoCrypt 2021: 4th International Conference on Historical Cryptology, pp. 1–11. Linköping University Electronic Press (2021)
14. Bosch Campos, V.: Advances in Document Layout Analysis. Ph.D. thesis, Universitat Politècnica de València (2020). https://doi.org/10.4995/Thesis.10251.138397
15. Islam, A., Anjum, T., Khan, N.: Line extraction in handwritten documents via instance segmentation. Int. J. Doc. Anal. Recogn. (IJDAR) **26**, 417–430 (2023). https://doi.org/10.1007/s10032-023-00438-7
16. Khanam, R., Hussain, M.: YOLOv11: An Overview of the Key Architectural Enhancements (2024). https://arxiv.org/abs/2410.17725
17. Lin, T.-Y., et al.: Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
18. Marti, U.V., Bunke, H.: The IAM-dataset: an English sentence database for off-line handwriting recognition. Int. J. Doc. Anal. Recogn. **5**(1), 39–46 (2002). https://doi.org/10.1007/s100320200071

19. Megyesi, B., et al.: Keys with nomenclatures in the early modern Europe. Cryptologia **48**(2), 97–139 (2024). https://doi.org/10.1080/01611194.2022.2113185
20. Quirós, L.: Multi-Task Handwritten Document Layout Analysis (2018). https://doi.org/10.48550/arXiv.1806.08852. arXiv preprint https://arxiv.org/abs/1806.08852
21. Šebeňa, M., Kačur, J.: Line detection in nomenclator-encrypted historical manuscripts. In: Proceedings ELMAR-2025: 67th International Symposium, pp. 307–310. Croatian Society Electronics in Marine, Zadar (2025)
22. Xu, Y., Yin, F., Zhang, Z., Liu, C.L.: Multi-task layout analysis for historical handwritten documents using fully convolutional networks. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18) (2018). https://doi.org/10.24963/ijcai.2018/147

Generative AI in Academia: Student Perspectives

David Martinčík, Olga Martinčíková Sojková(✉), Václav Procházka, and Jan Tlučhoř

Faculty of Economics, University of West Bohemia, Pilsen, Czech Republic
{martinci,sojkova,jtluchor}@fek.zcu.cz, vaproch@students.zcu.cz

Abstract. This study examines student attitudes, usage patterns, and institutional awareness regarding generative artificial intelligence in higher education, based on institutional guidance and student responses at two universities: the University of West Bohemia in Pilsen (Czechia) and Marquette University (United States). A structured survey measured the frequency and purposes of AI use, attitudes toward ethical and unethical use, perceived responsibility for preventing misuse, and perceived clarity of university rules.

Using descriptive statistics, semantic differential visualization, and latent profile analysis, the study explores differences between the two examined institutional contexts. UWB students reported higher acceptance of AI use in coursework but lower awareness of institutional communication. In contrast, students at Marquette University reported stronger ethical condemnation of misuse and greater perceived institutional clarity. Within the examined sample, institutional awareness was associated with more restrictive ethical attitudes and higher perceived institutional responsibility.

The findings suggest that student attitudes toward generative AI are multidimensional and may vary across institutional settings. The results suggest the potential importance of institutional communication and policy visibility for students' perceptions and ethical evaluations of AI use in higher education. Given the limited and institution-specific sample, the findings should be interpreted as exploratory.

Keywords: Artificial Intelligence · Comparative study · University student attitudes

1 Introduction

Generative AI has reshaped higher education with unusual speed. Universities were abruptly confronted with tools that can generate sophisticated essays, solve mathematical problems, and write code within seconds, shifting the debate from abstract forecasts to urgent questions about curriculum redesign, assessment integrity, and what learning means in an AI-enabled environment.

Universities' responses remain highly diverse. Some see generative AI as a direct challenge to traditional pedagogy, worrying that students will replace engagement and

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 235–247, 2027.
https://doi.org/10.1007/978-3-032-32104-6_21

critical thinking with algorithmic output. Others fear dependence will erode core skills and produce graduates without essential competencies. In contrast, many educators view AI as an opportunity and argue that curricula should teach students how to use these tools effectively for an AI-augmented workplace. Which of these narratives will dominate remains unclear, as long-term pedagogical effects are still emerging.

Yet most discussion focuses on institutional policy and faculty adaptation, often neglecting students' perspectives. This omission matters because much learning occurs in independent, low-visibility settings where instructors cannot reliably observe student practices. In these private contexts, students make their own ethical and practical decisions—whether to engage deeply with course content or delegate work to AI. Understanding students' motivations and usage patterns is therefore crucial for aligning pedagogical intentions with contemporary learning realities.

This study centers on the student perspective through a comparative survey at two geographically distinct institutions: the University of West Bohemia in Pilsen, Czech Republic, and Marquette University in Milwaukee, Wisconsin, USA. We examine how students use AI, how they perceive its broader impact on higher education, and how aware they are of institutional rules. Specifically, we investigate whether students understand their universities' policies on AI use and how they interpret their institution's overall stance toward technology.

2 Literature Review

The fast-growing literature on generative AI in higher education can be grouped into three main research streams: (i) pedagogical uses and instructional innovation, (ii) academic integrity and assessment challenges, and (iii) institutional governance and policy responses.

Pedagogical Applications
The pedagogical stream investigates how generative AI can support teaching and learning. Systematic reviews show growing use of large language models for writing scaffolds, brainstorming, formative feedback, and personalized pathways [1, 2]. Syntheses suggest that, under clear guidance, AI-supported environments can improve outcomes, especially in writing and problem-solving tasks [3]. However, benefits depend on transparent use, strong instructional design, and alignment with learning goals—positioning AI as an augmentation tool rather than a substitute for human thinking.

Academic Integrity
A second research stream focuses on academic integrity and shifting assessment practices. Because generative AI can rapidly produce essays, code, and analytical summaries, it raises questions about authorship, originality, and the validity of evaluation. Reviews emphasize both misuse risks and the limits of AI-detection tools [4]. Many standard assessment formats, designed for a pre-generative AI era, may no longer capture individual cognitive effort. As a result, scholars increasingly recommend process-based evaluation, oral exams, and authentic assessments that allow transparent AI use instead of blanket bans [5].

Institutional Governance and Policy Responses
The third stream examines institutional governance and guidelines. Universities have adopted varied approaches, from restrictive rules to structured integration frameworks [6]. Early analyses describe responses shaped by urgency and uncertainty, often casting generative AI as a disruptive threat [7]. More recent work takes a more balanced view, focusing on how institutions reconcile innovation with academic standards and student preparation for AI-augmented workplaces [8].

Across all three streams, bibliometric evidence indicates exponential growth in publications since ChatGPT's public release in 2022 [1]. Yet much of this literature remains institution-centered, emphasizing governance and pedagogical design. Even when student perceptions are included, students are often treated as targets of regulation rather than autonomous decision-makers in independent study contexts [6]. This points to a gap: how students interpret institutional policies, justify AI use, and assess its long-term implications for their learning.

2.1 Student Use, Perceptions, and Agency in AI-Augmented Learning Environments

Early research on generative AI in higher education was institution-focused mainly, but recent work increasingly examines student behavior and decision-making.

Studies show students use generative AI for tasks such as idea generation, summarizing academic texts, drafting assignments, programming support, and exam preparation [6, 7]. Use varies by discipline and study level: writing-intensive programs more often report AI-assisted drafting, while technical fields emphasize debugging and code optimization—suggesting that disciplinary norms and assessment structures shape adoption, not just individual preference.

Student attitudes are similarly mixed. Many view AI as a productivity tool that reduces cognitive load and increases efficiency [7], while others are ambivalent, weighing benefits against potential long-term costs to skills and intellectual autonomy [8]. Concerns about dependency and reduced critical thinking appear in student reports and faculty discourse [6], indicating that AI use is often accompanied by reflection rather than purely opportunistic behavior.

Growing literature also highlights the influence of perceived institutional norms. Students frequently report uncertainty about what counts as acceptable AI use, especially when policies are vague or inconsistently communicated [6]. In these cases, behavior tends to follow peer norms and informal interpretations more than formal documents, underscoring the importance of transparent governance.

Despite progress, cross-institutional comparisons remain scarce: most studies focus on single universities or national contexts [6, 7]. As a result, we know too little about how institutional culture, academic traditions, and regulatory frameworks shape student reasoning and adoption patterns.

2.2 Institutional Policy Awareness and Normative Interpretation

The rapid diffusion of generative AI has pushed universities to formalize rules on acceptable use, often covering authorship transparency, citation, assessment redesign, and the

boundary between assistance and misconduct. Yet having guidelines does not guarantee that students interpret or internalize them consistently.

Studies of institutional responses show wide variation. Some universities adopt restrictive approaches centered on prohibition and detection, while others favor guided integration supported by AI-literacy initiatives [6].

Empirical research suggests students often face ambiguity about expectations [6, 7]. Even when policies are public, awareness differs, and inconsistent messaging across courses and instructors fragments norms. As rules appear to vary by subject, acceptable use becomes context-dependent, and students may follow peer practices and informal cues more than official documents.

The challenge is not only policy design but also normative interpretation. Students do not always treat AI assistance as equivalent to traditional plagiarism; many distinguish full task delegation from limited support such as ideation or language refinement [4].

Assessment scholars argue that unclear expectations make compliance difficult to judge and enforce [5]. Detection tools cannot resolve normative uncertainty, especially when AI output can be heavily edited. Accordingly, researchers increasingly recommend proactive education—clear guidance, explicit curricular discussion of AI, and shared definitions of permissible practice—as foundations of sustainable governance [8].

Despite growing attention to policy frameworks, relatively little work compares how students perceive and internalize these rules across cultural contexts. Most studies remain single-institution and rarely examine how policy awareness relates to attitudes and self-reported behavior [6].

3 Methodology

The objective of this study is to compare student AI usage patterns, attitudes toward its educational role, and awareness of institutional policies across two universities operating in distinct academic environments.

The research is structured as a comparative case study with a cross-sectional design, capturing self-reported behavior and attitudes at a single point in time and enabling statistical comparisons between institutions. The design is exploratory rather than causal, aiming to identify structured differences and internal associations among key variables.

The analysis focuses on three dimensions derived from the literature review: frequency and forms of AI use; attitudinal evaluation, including ethical reasoning; and perceived institutional clarity and adaptation. Descriptive statistics and inferential tests are employed to assess differences between institutions and relationships among variables.

3.1 Research Context and Institutional Setting

The first examined institution is a public university in Central Europe operating within a state-funded higher education system. The second is a private Jesuit university in the United States functioning within a tuition-based governance model. Both institutions enroll approximately 12,000 students and maintain comparable faculty structures. In the Quacquarelli-Symonds (QS) World University Rankings, both universities fall into the overall top 1201–1400 category.

The institutions differ in their funding structures, governance traditions, and normative framings of academic integrity. Both universities have publicly communicated positions regarding the use of generative artificial intelligence in educational settings.

The case selection was intentional and aimed at examining how institutional context may influence student AI usage patterns, ethical evaluation, and policy awareness.

3.2 Participants and Sampling

The data was sampled between 25th of March and 7th of April 2025. The overall sample size is 194 respondents consisting of 160 respondents from the University of West Bohemia and the remaining 34 being enrolled at Marquette University. This imbalance of sample size is accounted for in the statistical analysis.

The questionnaire collected information on participants' current study level. Table 1 presents the distribution of respondents across these categories.

Table 1. Level of study distribution within sample.

Characteristic	University of West Bohemia (n = 160)	Marquette University (n = 34)	Whole sample (n = 194)
Level of study			
Undergraduate	91.9% (147)	91.2% (31)	91.8% (178)
Graduate	5.6% (9)	8.8% (3)	6.2% (12)
Postgraduate	2.5% (4)	–	2.1% (4)

The questionnaire was distributed via convenience sampling, with no reward for completing it. Distribution occurred primarily during scheduled lectures, where instructors displayed a QR code linking to the online survey. Additional distribution occurred using posters, e-mails, and mass communication platforms such as Discord. Only students at the examined universities were allowed to fill out the survey.

3.3 Questionnaire Design

The questionnaire was administered using Google Forms and structured into multiple sections corresponding to the main analytical dimensions of the study. The structure separates ethical attitudes and institutional perceptions. This organization allows independent analysis of individual dimensions as well as examination of relationships between them.

The first section examined ethical attitudes and perceived responsibility related to AI use in education. To ensure consistent interpretation, respondents were provided with an explicit definition of unethical AI use, referring to situations in which AI replaces genuine learning, such as submitting AI-generated work without understanding or using AI dishonestly during assessment (Table 2).

The next section assessed student perceptions of institutional adaptation, communication, and enforcement capacity related to generative AI (Table 3).

Table 2. Questionnaire section: Ethical attitudes and responsibility

Question	Response format
Q_01: Using AI by students in university courses should be allowed.	Scale 0 (Disagree) – 10 (Agree)
Q_02: I condemn unethical use of AI by students.	Scale 0 (Disagree) – 10 (Agree)
Q_03: I sometimes use AI unethically.	Scale 0 (Disagree) – 10 (Agree)
Q_04: It is the university's responsibility to prevent unethical use of AI.	Scale 0 (Disagree) – 10 (Agree)
Q_05: It is the student's responsibility to avoid unethical use of AI.	Scale 0 (Disagree) – 10 (Agree)

Table 3. Questionnaire section: Perceived institutional adaptation

Question	Response format
Q_06: The university has made its stance on AI clear to me.	Scale 0 (Disagree) – 10 (Agree)
Q_07: My courses have been appropriately adapted to the emergence of AI.	Scale 0 (Disagree) – 10 (Agree)
Q_08: The university strives to educate students on the use of AI tools.	Scale 0 (Disagree) – 10 (Agree)
Q_09: I attended courses where I learned less because AI made completion easier.	Scale 0 (Disagree) – 10 (Agree)
Q_10: Professors are capable of detecting unethical AI use.	Scale 0 (Disagree) – 10 (Agree)

The final section measured awareness of official institutional communication regarding AI policies. This variable allows distinction between general institutional perceptions and actual exposure to formal guidance (Table 4).

Table 4. Questionnaire section: Exposure to institutional AI guidelines

Question	Response format
Have you seen official university materials addressing AI use?	Yes / No

All attitudinal questions used eleven-point Likert-type scales ranging from 0 to 10. This scale provides sufficient resolution while maintaining interpretability and comparability across variables.

3.4 Ethical Considerations

Survey participation was voluntary and uncompensated. The questionnaire was anonymous, and no personally identifiable data (e.g., names, student IDs, email addresses, IP addresses) were collected, ensuring responses could not be traced to individuals.

Before starting, respondents were informed of the study's purpose and the anonymous data collection process; continuing to the questionnaire constituted informed consent for academic use of their responses. No sensitive personal data was gathered, and participation involved no foreseeable academic, psychological, or personal risk. Given the survey's anonymous, non-invasive design, formal ethical approval was not required.

3.5 Statistical Analysis

The statistical analysis was designed to address two primary research questions derived from the theoretical framework and questionnaire structure:

- RQ1: Do students at the two institutions differ in their attitudes toward the use of generative artificial intelligence in education and in their ethical evaluation of its use?
- RQ2: Is awareness of institutional communication regarding AI associated with differences in student attitudes and ethical evaluation?

To answer these questions, the analysis combined descriptive visualization and inferential statistical testing. All analyses were conducted using standard statistical procedures suitable for Likert-type data and categorical variables.

Visualizations and Group Comparison
The initial stage of analysis focused on descriptive comparison of responses between institutions. Likert-scale distributions were visualized [9] to examine the relative frequency of response categories across student groups. These visualizations allowed direct comparison of attitudinal patterns and provided an overview of differences in ethical evaluation, perceived responsibility, and institutional perception.

To facilitate comparison of central tendencies across multiple variables, semantic differential plots were constructed showing mean values and confidence intervals for each attitudinal item. Due to the smaller sample size at Marquette University, confidence intervals were estimated using bootstrap resampling methods [10, 11], which provide robust interval estimation without assuming normality of the underlying distribution [12].

Latent Profile Analysis
To examine whether awareness of institutional AI guidelines was associated with distinct response patterns, latent profiles were identified within the University of West Bohemia sample. Latent profile analysis [13] is a clustering approach that partitions respondents into groups based on similarity of response patterns, allowing identification of internally consistent attitudinal profiles [14].

The association between latent profiles and awareness of institutional AI communication was evaluated using contingency tables and the Pearson chi-square test of independence [15]. The chi-square test assesses whether the observed distribution of cases

across categories differs significantly from the distribution expected under statistical independence.

4 Results

Likert-scale items are summarized in Fig. 1 using diverging stacked bar charts. Overall, UWB students express a more positive attitude toward using AI in classes. They also show lower approval of unethical AI use, yet report engaging in it more often. On responsibility for preventing unethical use, UWB students place somewhat more responsibility on the university than MU students do.

In the section on the university's approach to AI, MU students more strongly report being well informed about institutional rules and believe the university has adapted courses effectively. UWB students, in contrast, agree more that the university educates students about AI use. Both groups report similar views on whether courses are too easy to complete with AI, and their median responses are nearly identical regarding instructors' ability to detect unethical AI use.

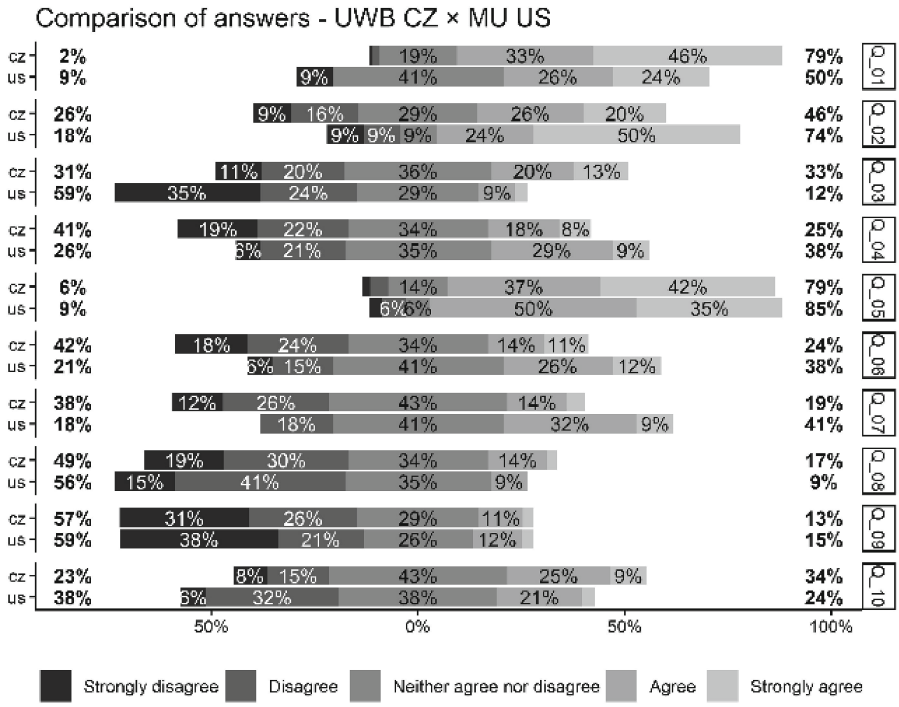


Fig. 1. Comparison of Likert scale answers distribution for both universities

To further examine differences in central tendencies across institutions, mean response values were compared using semantic differential plots with bootstrapped confidence intervals. This graph can be seen in Fig. 2.

The semantic differential visualization confirms the patterns observed in the distribution analysis. Students at the University of West Bohemia showed higher acceptance of AI use in education, while Marquette students demonstrated stronger ethical reservations and greater confidence in institutional adaptation and oversight.

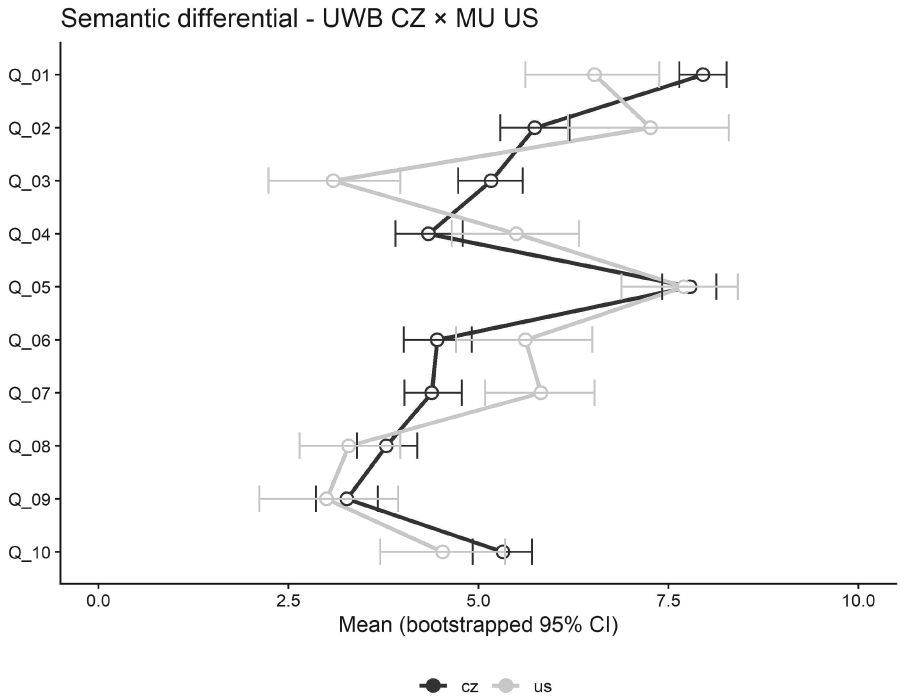


Fig. 2. Bootstrapped semantic differential graph on Likert scale questions

4.1 Association Between Institutional Awareness and Student Attitudes

The second research question examined whether awareness of official institutional communication regarding AI was associated with differences in student attitudes. This analysis focused on respondents from the University of West Bohemia, where variation in institutional awareness was observed. The comparison of mean responses between students aware and unaware of institutional guidance can be seen in Fig. 3.

Students who reported awareness of official institutional guidance exhibited systematically different response patterns. These respondents expressed higher levels of agreement with statements supporting ethical responsibility and institutional adaptation. They also showed stronger confidence in the university's ability to address AI-related challenges.

In contrast, students unaware of institutional guidance demonstrated more neutral or uncertain responses across multiple dimensions. This pattern suggests that institutional

communication is associated with measurable differences in student perceptions and normative evaluation.

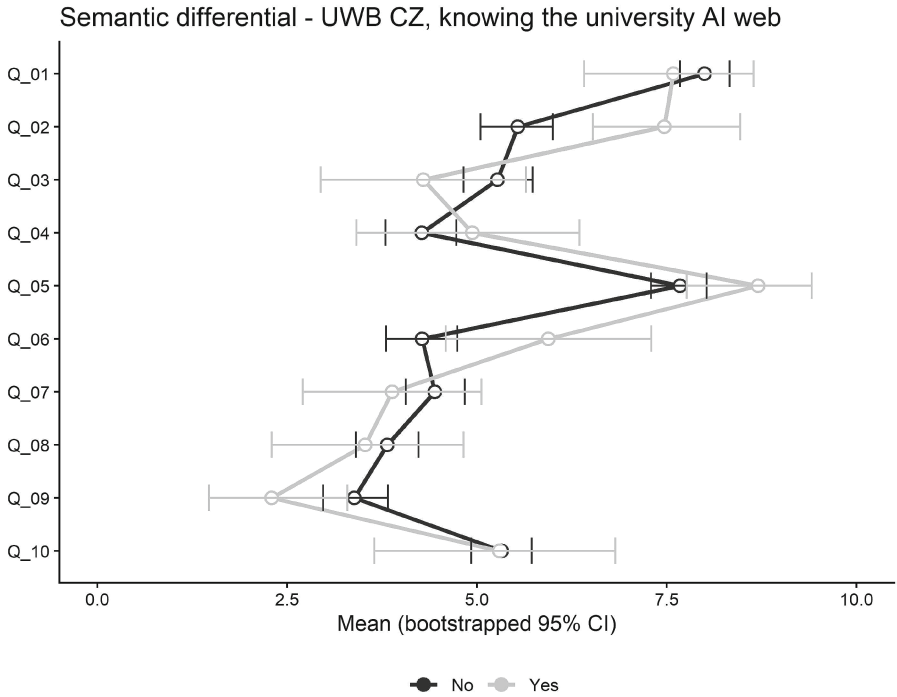


Fig. 3. Semantic differential in UWB students differentiated by their exposure to official AI guidelines.

Next, we conducted latent profile analysis on the UWB students' data which separated the observations into more natural groups of students by identifying students with similar responses. The result of the latent profile analysis can be seen visualized in Fig. 4.

We found that class 2 has more positive attitude towards allowing AI to be used in classes, is less likely to condemn and more likely to engage in unethical use and feels lesser responsibility both from the students and the institutions while feeling less aware of the official UWB stance on AI.

After examining these results, we noticed a visual similarity between the semantic differential graph dividing students based on their exposure to the official UWB AI website and the latent profiles. And thus, we constructed a contingency table in order to see if the results of these two analyses correlate. The contingency table can be seen in Table 5.

Pearson's Chi-squared test with Yates' continuity correction confirmed a relationship between the latent classes and guideline exposure ($p = 0.01876$). We then measured the size of this association using Cramer's V which confirmed a weak association ($V = 0.206$). The weakness of association is most likely caused by the imbalance of group sizes with students aware of the UWB AI web making up only 10.6% of the sample.

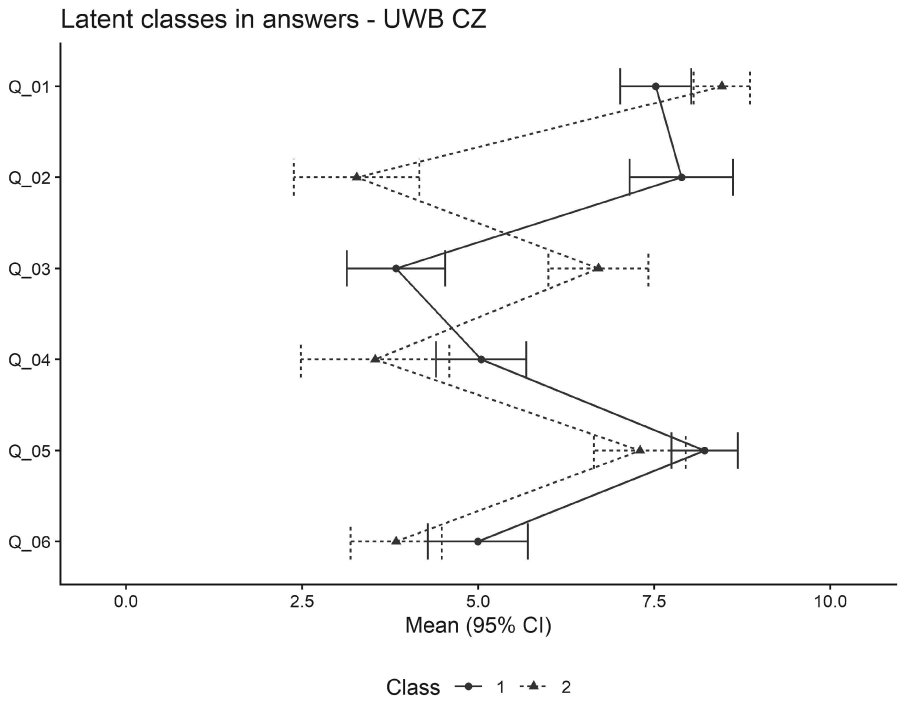


Fig. 4. Latent profile analysis results on UWB students.

Table 5. Contingency table of latent classes and exposure to official UWB guidelines

LPA class / web exposure	No	Yes
1	70	14
2	73	3

This analysis suggests that the latent factor differentiating between the classes correlates with exposure to the official university's AI website and thus we could theorize that exposure to official guidelines makes students more likely to condemn and less likely to engage in the unethical usage of AI. However, we can only support the claim that correlation exists between these factors as we don't have any proof of a causal relationship.

5 Discussion and Conclusion

This study compared student attitudes toward generative AI at two universities operating in different institutional contexts. The findings suggest that student perceptions are multidimensional, structured, and related to the institutional environment in which they are formed.

Differences between the two examined institutions were observed. Students at the University of West Bohemia reported greater acceptance of AI use, but lower perceived clarity of institutional rules and weaker course adaptation. Students at Marquette University expressed stronger ethical reservations and greater confidence in institutional communication and adaptation. These patterns suggest that institutional context may be related not only to reported behavior but also to how students normatively evaluate AI use.

Institutional awareness also appeared to differentiate attitudes. Students familiar with official guidance more strongly condemned misuse and expressed greater confidence in institutional responsibility. Latent profile analysis similarly linked policy awareness with more restrictive ethical stances, indicating a possible relationship between institutional communication and students' normative frameworks, even in the absence of strict enforcement.

Overall, student attitudes cannot be reduced to simple approval or rejection; rather, students appear to weigh institutional legitimacy, ethical acceptability, and academic consequences simultaneously.

Key limitations include the restricted comparison to two universities, the strong imbalance in sample composition, a cross-sectional design that limits causal claims, potential bias in self-reports, and convenience sampling that may reduce representativeness. The study should therefore be interpreted as exploratory rather than confirmatory. Future research should broaden comparative coverage, include more balanced institutional samples, and track how student attitudes shift as institutional policies mature.

Acknowledgement. This paper was supported by the project SGS-2024-030, "Research on quantitative approaches for economic problems solving" of the Faculty of Economics of the University of West Bohemia in Pilsen.

References

1. Rivadeneira, D.X.R., Toledo, J.A.J.: Applications of generative artificial intelligence in higher education: a systematic review of the literature. J. Posthumanism. **5**(11), 521–535 (2025). https://doi.org/10.63332/joph.v5i11.3770
2. Chanpradit, T.: Generative artificial intelligence in academic writing in higher education: a systematic review. Edelweiss Appl. Sci. Technol. **9**(4), 889–906 (2025). https://doi.org/10.55214/25768484.v9i4.6128
3. Chen, S., Cheung, A.C.K.: Effect of generative artificial intelligence on university students' learning outcomes: a systematic review and meta-analysis. Educ. Res. Rev. **49**, 100737 (2025). https://doi.org/10.1016/j.edurev.2025.100737
4. Bittle, K., El-Gayar, O.: Generative AI and academic integrity in higher education: a systematic review and research agenda. Information 16(4), 296 (2025). doi:https://doi.org/10.3390/info16040296
5. Xia, Q., Weng, X., Ouyang, F., Lin, T.J., Chiu, T.K.F.: A scoping review on how generative artificial intelligence transforms assessment in higher education. Int. J. Educ. Technol. High. Educ. (2024). https://doi.org/10.1186/s41239-024-00468-z

6. Johnston, H., Wells, R.F., Shanks, E.M., Boey, T., Parsons, B.N.: Student perspectives on the use of generative artificial intelligence technologies in higher education. Int. J. Educ. Integr. **20**, 2 (2024). https://doi.org/10.1007/s40979-024-00149-4
7. Barrett, A., Pack, A.: Not quite eye to AI: student and teacher perspectives on the use of generative artificial intelligence in the writing process. Int. J. Educ. Technol. High. Educ. **20**, 59 (2023). https://doi.org/10.1186/s41239-023-00427-0
8. Bobula, M.: Generative artificial intelligence in higher education: a comprehensive review of challenges, opportunities, and implications. J. Learn. Dev. Higher Educ. **30** (2024). https://doi.org/10.47408/jldhe.vi30.1137
9. Heiberger, R.M., Robbins, N.B.: Design of diverging stacked bar charts for Likert scales and other applications. J. Stat. Softw. **57**(5), 1–32 (2014). https://doi.org/10.18637/jss.v057.i05
10. Canty, A., Ripley, B.: Boot: Bootstrap Functions. https://CRAN.R-project.org/package=boot (2025)
11. Davison, A.C., Hinkley, D.V.: Bootstrap Methods and their Applications. Cambridge University Press, Cambridge (1997). https://doi.org/10.1017/CBO9780511802843
12. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman & Hall, New York, NY (1993). https://doi.org/10.1007/978-1-4899-4541-9
13. Rosenberg, J.M., Beymer, P.N., Anderson, D.J., Van Lissa, C.J., Schmidt, J.A.: tidyLPA: an R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. J. Open Source Softw. **3**(30), 978 (2018). https://doi.org/10.21105/joss.00978
14. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer, New York, NY (2009). https://doi.org/10.1007/978-0-387-84858-7
15. Agresti, A.: Categorical Data Analysis, 3rd edn. Wiley, Hoboken, NJ (2013). https://doi.org/10.1002/9781119405269

A High-Fidelity VR Framework for Immersive Performance Analytics and Seamless User Feedback in Construction 4.0 Environments

Mariana Moreiras[1](✉), Leonel Deusdado[1], and Luís Afonso[2]

[1] CITed: Transdisciplinary Research Center in Education and Development, IPB, 5300 Bragança, Portugal
a59462@alunos.ipb.pt, leodeus@ipb.pt

[2] MoreColab: Edifício do Brigantia Ecopark, Avª Cidade de Leon 506, 5300-358, Bragança, Portugal
lafonso@morecolab.pt

Abstract. Digital transformation in Construction 4.0 is increasingly dependent on high-fidelity Virtual Reality (VR) to simulate high-risk operational environments. However, a significant gap exists in capturing granular behavioral data without compromising the user's sense of presence. This paper proposes a data-driven framework for real-time performance analytics and immersive feedback within a standalone VR ecosystem. Developed using Unity 3D and Meta Quest 3, the system implements an asynchronous telemetry pipeline with Firebase to transform complex immersive interactions into objective Key Performance Indicators (KPIs). The methodology focuses on a diegetic interaction design, utilizing World-Space UIs for data collection and feedback, ensuring sensorimotor contingency remains intact. Future validation will assess whether the integration of real-time objective metrics with immersive subjective assessment enables a longitudinal tracking approach for operational readiness. This research contributes to the field of Immersive Analytics by providing a scalable solution for data persistence and performance monitoring in standalone VR simulations.

Keywords: Virtual Reality (VR) · Immersive Analytics · Human-Computer Interaction (HCI) · Training Platform · Construction 4.0 · Key Performance Indicators (KPI)

1 Introduction

The construction industry is undergoing a digital revolution under the Construction 4.0 paradigm, where the fusion of the Internet of Things (IoT), robotics, and Virtual Reality (VR) creates interconnected cyber-physical ecosystems. Within this framework, VR has transitioned from a simple visualization tool to a high-fidelity environment for sensorimotor training in high-risk scenarios [1]. By providing safe, fully controlled simulations of complex construction sites, VR enhances knowledge retention and mitigates operational risks through embodied learning.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 248–259, 2027.
https://doi.org/10.1007/978-3-032-32104-6_22

However, a critical bottleneck in current immersive training solutions is the "immersion-evaluation paradox": while systems excel at providing high presence, they often lack non-intrusive mechanisms for systematic performance tracking. Most existing platforms prioritize visual experience but fail to implement real-time telemetry for capturing objective behavioral data (e.g., precision, reaction time, efficiency) and subjective psychological metrics (e.g., cognitive load, confidence) without disrupting the user's sense of presence. This lack of structured data persistence hinders the development of adaptive learning systems and rigorous performance analytics.

To address this gap, this work proposes an integrated architectural framework for real-time performance monitoring and feedback within a standalone VR environment. The novelty of this work lies not in individual components, but in the integrated architecture that enables seamless coupling of real-time telemetry, diegetic interaction, and immersive evaluation without inducing breaks-in-presence. By leveraging an asynchronous data pipeline between the immersive interface and a cloud-based backend, the system enables the collection of multi-dimensional metrics through diegetic World-Space UIs. This approach ensures that performance evaluation becomes an intrinsic part of the simulation, contributing to the evolution of intelligent, data-driven, and highly immersive training ecosystems in Construction 4.0.

2 Related Works

The digitalization of the construction sector has shifted the training paradigm from traditional methods to high-fidelity immersive environments. Recent literature emphasizes that the effectiveness of Virtual Reality (VR) is not merely a product of visual realism but depends on complex interaction design and robust data tracking. Dincelli [2] identifies "presence by design" as a fundamental requirement for knowledge transfer, arguing that the interface's architecture must minimize the cognitive load to maintain psychological immersion. This is corroborated by Dickinson et al. [3], who established the importance of diegetic tool management—integrating interface elements into the virtual narrative—to prevent "breaks in presence" (BIP) that typically occur with traditional 2D overlays.

Despite the maturity of simulation technologies, a significant gap remains in the systematic assessment of user performance. Khorrami Shad et al. [4] highlight that while Augmented Reality (AR) and VR are relevant tools for risk simulation, there is a persistent lack of automated metric mapping in Construction 4.0. While some approaches, such as those by Ding et al. [5] and Cordeiro et al. [6], explore safety training for laboratory and height-related risks, their evaluation modules often rely on post-simulation data or manual observation. This limitation is addressed in other fields, such as nursing education, where Morishita et al. [7] demonstrated that real-time feedback loops significantly improve procedural proficiency. However, translating these feedback mechanisms into a standalone, scalable VR architecture for industrial environments remains a challenge.

Methodological constraints also extend to how subjective data is collected within immersive sessions. Gorup and Lesar [8] recently proposed a framework for immersive questionnaires, emphasizing that collecting user perceptions without removing the headset is vital for perception research. Similarly, Tao et al. [9] used AI-supported VR to

monitor self-regulated learning patterns, suggesting that behavioral analytics can reveal deeper insights into trainee readiness than binary task-success metrics. Ahmed et al. [10] and Wang [11] further argue that the integration of gamified elements and systematic evaluation in classroom settings enhances engagement, yet they point to the scarcity of persistent backend solutions for longitudinal tracking.

Finally, the evolution of VR training is moving toward multi-sensory and data-driven ecosystems. Early works by Clifford et al. [12] and comparative studies by Yigitbas et al. [13] laid the groundwork for situational awareness in VR. More recently, Bedard et al. [14] emphasized the need for integrating "hard" and "soft" approaches in Human-Computer Interaction (HCI) to support decision-making in public infrastructure. Our study positions itself at the intersection of these trends, proposing an integrated framework that combines diegetic interaction, as suggested by Dickinson, with a real-time cloud-based telemetry system, filling the data-persistence gap identified by the most recent literature.

A consolidation of this literature (Table 1) and the thematic analysis conducted by Zhang et al. [15] reveal a clear predominance of solutions focused on the simulation component. As illustrated in Fig. 1, approximately 55% of studies are dedicated to Safety Training and Education, highlighting the power of VR for immersive risk simulation. However, this strong focus on visualization contrasts with a critical weakness identified by Khorrami Shad et al. [4]: the lack of integrated, structured mechanisms for objective performance evaluation in Construction 4.0 environments.

Overall, the literature, now encompassing cutting-edge studies from 2024 to 2026, shows that while simulation fidelity has reached high levels, a significant "immersion-evaluation gap" remains. This gap can be categorized into three technological barriers:

(i) Absence of Scalable Data Architecture: As noted by Ahmed et al. [10] and Wang [11], current systems lack persistent, cloud-based backends for longitudinal tracking of user progress.
(ii) Disruption of Presence (BIP): Recent research by Dincelli [2] and Gorup and Lesar [8] emphasizes that traditional evaluation methods cause "Breaks-in-Presence," negatively impacting the educational outcome.
(iii) Shallow Metric Mapping: While Morishita et al. [7] and Ding et al. [5] explore real-time feedback, there is still a scarcity of systems that transform complex immersive interactions into objective operational readiness KPIs.

Unlike existing systems that rely on post-hoc analysis or non-immersive evaluation methods, the proposed framework embeds both quantitative and qualitative assessment directly within the VR environment through diegetic interfaces, enabling continuous, non-intrusive performance monitoring.

This study positions itself in this research frontier, proposing a high-fidelity VR framework that integrates an automated assessment module with a Firebase-driven telemetry pipeline. By utilizing World-Space UIs for diegetic feedback, the system ensures that performance evaluation becomes an intrinsic part of the training process without compromising psychological immersion.

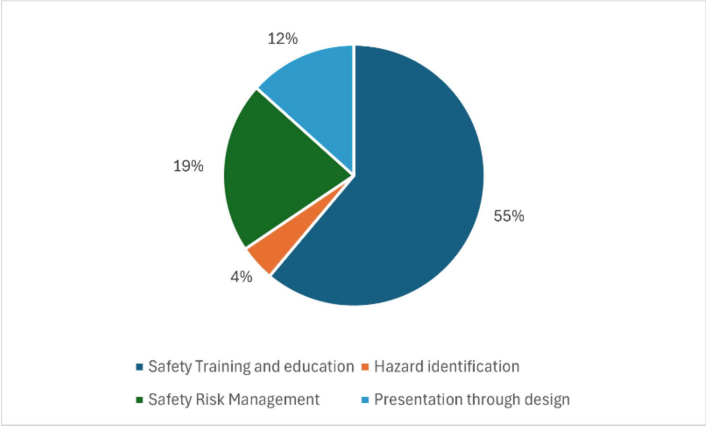


Fig. 1. Thematic focus of Virtual Reality research in construction safety

Table 1. Comparative summary of the studies analysed

Gap Category	Reference Articles	Detailed Description of the Gap	Solution Proposed by this project
Lack of Data Persistence	Ahmed et al. [10], Wang [11], Abotaleb et al. [16]	Absence of scalable cloud databases for longitudinal tracking across multiple sessions.	Firebase integration for real-time telemetry and persistent historical data flow.
Interruption of Presence (BIP)	Dincelli [2], Gorup & Lesar [8], Getuli et al. [17]	Evaluation methods require headset removal or non-diegetic overlays, disrupting the sense of presence.	Diegetic World-Space UI for immersive questionnaires, maintaining immersion autonomously.
Limited Behavioral Metrics	Morishita et al. [7], Ding et al. [5], Scorgie et al. [1]	Metrics are limited to binary success or manual observation, missing granular interaction logs.	Event-based C# scripts for automatic collection of real-time KPIs (time, errors, efficiency).
Procedural Vs. Technical Gap	Khorrami Shad et al. [4], Jakobo & Taifa [18]	Prioritization of hazard awareness over the assessment of technical proficiency in complex tasks.	Remote Robotics Control integration (via virtual camera) to assess technical competence.

3 Methodology

The research design follows an applied, iterative framework that bridges systematic literature analysis with the engineering of an integrated Immersive Analytics ecosystem.

3.1 Research Questions and Theoretical Framing

To address the current technological gaps in Construction 4.0 training, the following questions were formulated:

- Q1: How can high-fidelity VR simulations improve cognitive and sensorimotor retention in high-risk training?
- Q2: What multidimensional metrics are required to quantify operational readiness in immersive environments?
- Q3: How do different modalities of real-time feedback impact the user's sense of presence?
- Q4: What architectural limitations prevent the scalability of data persistence in standalone VR?
- Q5: How can diegetic interaction design integrate objective and subjective evaluation without causing Breaks-in-Presence (BIP)?

3.2 VR System Architecture and Development

The development of the assessment and feedback module was based on a user-centered design, prioritizing psychological fidelity and system performance. The system was built using Unity 3D (C#) for real-time rendering and Blender for the optimization of 3D assets to ensure high frame rates on standalone hardware.

The hardware interaction layer utilizes the Meta Quest 3, leveraging its inside-out tracking capabilities for precise sensorimotor input. The framework is architected into a three-tier modular subsystem:

(i) Monitoring Layer: Captures raw interaction events (triggers, collisions, spatial coordinates) and user-robot interaction logs.
(ii) Processing Layer: Implements an event-based logic to transform raw telemetry into structured Key Performance Indicators (KPIs).
(iii) Persistence Layer: Executes an asynchronous data pipeline to a Firebase Real-time Database, ensuring data integrity and individual performance tracking without impacting the simulation's latency.

3.3 Diegetic Interaction and Telemetry Flow

A core innovation of this methodology is the use of Diegetic Interaction Design. The user interacts with a World-Space Spatial UI (virtual tablet) that serves as a bridge between the virtual and digital twins of the task. This interface provides:

- Real-time Telemetry Visualization: Live feedback of KPIs and a remote camera feed from the robot, simulating complex teleoperation scenarios.
- Control Interface: A hybrid input system (buttons and virtual joystick) that generates Remote Procedure Calls (RPC) for robot kinematics.

All sensorimotor data, task metrics, and responses to Immersive Qualitative Questionnaires are streamed to the cloud. This ensures a closed-loop system where execution, evaluation, and analysis coexist within the immersive experience.

3.4 Evaluation Framework: Objective and Subjective Integration

The assessment module is designed to solve the "immersion-evaluation gap" by combining:

- Automatic Quantitative Analytics: Sub-second tracking of execution time, error frequency, interaction fluidity, and safety protocol compliance.
- Immersive Subjective Assessment: Embedded Likert-scale questionnaires designed as 3D objects to prevent the withdrawal of the HMD (Head-Mounted Display), thus preserving the user's Sense of Presence.

By aligning these dimensions, the proposed framework contributes to the evolution of Construction 4.0 training from passive visualization to an intelligent, data-driven, and highly immersive learning ecosystem.

4 System Architecture and Validation Framework

4.1 Immersive Infrastructure

The virtual reality training ecosystem was architected upon the base infrastructure of the ROBETARME project [19]. The primary engineering goal was to extend the original simulation into a data-driven environment by integrating a High-Frequency Telemetry Module and an asynchronous feedback loop for real-time performance monitoring within Construction 4.0 scenarios.

4.2 Telemetry Implementation and Unity Integration

The module leverages the Unity 3D engine and C# for low-latency event processing. The interface employs a hybrid UI strategy, combining standard Canvas with Diegetic World-Space UIs. This approach anchors performance dashboards directly into the virtual geometry, allowing for bidirectional communication with Firebase. All sub-second interaction data—including spatial coordinates and trigger events—are streamed to the cloud to ensure reliable data persistence.

4.3 Cyber-Physical Robot Kinematics and Visual Fidelity

The industrial robot's behavior is governed by custom kinematics scripts, enabling dynamic interaction. To enhance cognitive fidelity, we implemented a remote monitoring system via a virtual tablet (Fig. 2a). This tablet renders a real-time feed from a camera attached to the robot's end-effector, forcing users to rely on mediated observation—a critical skill in remote teleoperation and safety management.

Fig. 2. Robot, interactive tablet and questionnaire

4.4 Event-Driven Interaction Model

The system architecture follows an Event-Driven Design (EDD). Utilizing Unity Events and Triggers, the framework captures granular user actions through Meta Quest 3 controllers. This ensures that every safety protocol violation or procedural success is time-stamped and logged, facilitating a detailed Behavioral Traceability between the user's sensorimotor input and the recorded KPIs.

4.5 Immersion-Preserving Assessment

To mitigate Break-in-Presence (BIP), qualitative assessment is conducted via Immersive Questionnaires (Fig. 2b). These are rendered as interactive 3D objects within the user's reach, associated with a Firebase Authentication layer. This allows for personalized, longitudinal profiling without requiring the removal of the Head-Mounted Display (HMD).

4.6 Validation Criteria and Behavioral Analytics

The validation of the framework relies on the transformation of raw telemetry into structured Operational Readiness KPIs.

4.7 Quantitative Performance Analytics

The system automatically computes the following metrics, synchronized with the cloud backend:

- Task execution time.

  The total execution time ($T_{i,j}$) for an individual user, j, on a main task, i, is the sum of the times of all their completed K subtasks ($t_{i,j,k}$):

$$T_{i,j} = \sum_{k=1}^{K} t_{i,j,k} \tag{1}$$

- Number and type of errors made, including collisions, procedural failures, and incorrect decisions.

  For an individual user j in task i, the total number of errors ($Errors_{total,i,j}$) is calculated by summing the individual errors ($e_{i,j,k}$) across all K subtasks:

$$Errors_{total,i,j} = \sum_{k=1}^{K} e_{i,j,k} \tag{2}$$

- Overall task efficiency, calculated by the ratio of execution time to performance quality.

  Time-based efficiency considers the total success of the task and the total time spent.For individual user j:

$$E_{i,j} = \frac{N_{i,j}}{T_{i,j}} \tag{3}$$

  Where $N_(i,j)$ is the result of the main task (1 if completed successfully overall, 0 if failed) and $T_{i,j}$ is the total execution time (sum of subtasks).
- Safety protocol compliance rate.

  The compliance rate measures the percentage of subtasks completed successfully. For user j, this is calculated by the sum of subtasks classified as successful ($\text{Successful}_{i,j,k=1}$) divided by the total number of subtasks K:

$$SCR_{i,j} = \sum_{K}^{K} k = 1\text{Successful}_{i,j,k} \times 100\% \tag{4}$$

- Performance evolution over multiple training sessions.

  Performance improvement, whether in terms of reduced execution time (T) or number of errors ($Errors_{total}$), will be modelled using Wright's learning curve formula. This formula describes a geometric relationship where the average time or number of errors decreases at a constant rate whenever the output (or number of training sessions) doubles. The general formula is: $Y = A \cdot X^B$. Where Y is the observed performance metric (e.g., cumulative average time per session, or cumulative average number of errors per session); A is the initial value of the metric in the first session ($X = 1$); X is the cumulative number of training sessions completed by the user; B is the slope coefficient of the learning curve, which determines the rate of improvement. The value of B is typically negative, indicating a reduction in time/errors as X increases.

  The coefficient B can be calculated from the learning rate (LR):

$$B = \frac{\log(LR)}{\log(2)} \tag{5}$$

  The LR is the percentage of the original time required to complete the task whenever the number of repetitions doubles (e.g., a learning rate of 80% means that the time

is 80% of the previous one when the sessions double). The validation of learning will be demonstrated by statistical analysis (e.g., repeated measures ANOVA or paired t-tests) which will confirm whether there is a statistically significant reduction in $T_{i,j}$ and $Errors_{total,i,j}$ between the first session (Session 1) and the last session (Session N) for each user j.

These results are converted into KPIs presented in trend graphs, allowing the reduction in time and errors throughout the sessions to be visualized, validating the user's progress.

The metrics described are presented to the user in an integrated manner in the virtual environment. Figure 3 illustrates the visualization of these performance KPIs on the interactive tablet, allowing the user to analyse their performance immediately and intuitively.

Fig. 3. Presentation of performance results in the virtual environments

4.8 Qualitative UX and Immersive Assessment

The qualitative component is captured through three-dimensional interfaces integrated directly into the virtual environment. This assessment aims to evaluate the user experience (UX) and the pedagogical effectiveness of the simulation through three core dimensions:

- Platform Usability and User Satisfaction: Evaluates the ergonomic quality of the interaction and the efficiency of the diegetic interface (virtual tablet). This corresponds to the system's ability to facilitate task execution without technical frustration.
- Perceived Learning and Operational Confidence: Assesses the user's subjective perception of skill acquisition and the increase in confidence regarding the safety procedures performed. This measures the psychological transition from trainee to a digitally-ready operator.

Sense of Presence and Representational Fidelity: Analyzes the degree of immersion and the realism of the virtual scenario. This metric determines how effectively the simulation mimics the environmental pressures and visual cues of a real Construction 4.0 site.

5 Discussion and Contributions

The primary contribution of this research is the engineering of a unified telemetry framework that resolves the "immersion-evaluation paradox" identified in the literature (Sect. 2). While previous studies often sacrificed data granularity for presence, or vice versa, this work proposes and illustrates that asynchronous data persistence combined with diegetic interaction provides a scalable solution for Construction 4.0 training.

The proposed module (Sect. 4) leverages a Firebase-driven backend, overcoming the data fragmentation and lack of longitudinal tracking identified in contemporary studies such as Ahmed et al. [10] and Wang [11]. Furthermore, by integrating automatic metrics with World-Space qualitative interfaces (Sect. 4.6), this framework addresses the multi-dimensional assessment needs highlighted by Cordeiro et al. [6], but with the technical advantage of a standalone, autonomous architecture (Meta Quest 3), eliminating the need for intrusive external sensors or tethered setups.

5.1 Synthesis of Research Findings

This section provides conclusive answers to the research questions formulated in Sect. 3.1:

- Q1 (VR Efficacy): Literature and development confirm that high-fidelity VR is an effective tool for sensorimotor training. However, the paradigm is shifting from simple visualization to cyber-physical simulations where technical proficiency is as critical as hazard awareness.
- Q2 (Evaluation Metrics): Beyond the binary "success/fail" metrics seen in early works [17], this research establishes that performance must be quantified through Interaction Fluidity and Error Density (Sect. 4.6.1), providing a deeper understanding of the user's operational readiness.
- Q3 (Feedback Modalities): The implementation of Diegetic Real-time Feedback (via virtual tablet) proves to be a relevant factor for pedagogical retention. It mitigates the absence of immediate feedback identified in [7], ensuring that the trainee corrects posture and procedural errors without exiting the immersive state.
- Q4 (Technical Limitations): The investigation identified that the main barriers in current Construction 4.0 VR are the lack of Asynchronous Telemetry and the high frequency of Breaks-in-Presence (BIP). This work directly addresses these by moving the evaluation into the simulation's narrative space.
- Q5 (System Integration): The module validates the feasibility of a Hybrid Analytics approach. By utilizing an Event-Driven Architecture (EDA) in Unity for objective data and Spatial UIs for subjective feedback, we achieve a closed-loop system where data collection is non-intrusive and persistent.

5.2 Contribution to Construction 4.0 and HCI

In summary, the developed system fills an identified gap in digital training by providing a data-driven, secure, and scalable platform. The integration of Wright's Learning Curve (Sect. 4.6) for predictive performance modeling represents a novel application of industrial engineering principles within a VR ecosystem. The expected results indicate that this architectural approach not only enhances knowledge acquisition but also ensures the longitudinal persistence of safety protocols, contributing to a more rigorous and technologically advanced construction workforce.

6 Conclusion

The core contribution of this framework is its ability to bridge the "immersion-evaluation gap." Unlike existing solutions that suffer from Breaks-in-Presence (BIP) or fragmented data tracking, this module is designed to provide a closed-loop evaluation cycle. Through diegetic interaction design, complex sensorimotor interactions are transformed into objective operational readiness KPIs, presented to the user via a World-Space Dashboard. This ensures that the learning process remains data-driven without compromising the psychological fidelity of the simulation.

Although validated in Construction 4.0 scenarios, the proposed framework is domain-agnostic and transferable to other safety-critical VR training contexts such as healthcare, aviation, and emergency response.

7 Future Works

Considering the proposed framework for immersive performance analytics and the current validation stage, several future lines of work are envisioned. First, conducting longitudinal experimental studies with larger and more diverse user groups is essential to empirically validate the correlation between telemetry-based KPIs and actual operational readiness. The integration of artificial intelligence algorithms represents a critical next step, enabling automatic performance analysis, error detection, and the generation of personalized training recommendations based on historical data. Furthermore, the expansion of simulation scenarios to include multi-user collaborative tasks and interaction with various types of Construction 4.0 robots will enhance the system's versatility. From a technical perspective, the development of a complementary web platform for remote monitoring and the improvement of haptic feedback loops are planned to increase the ecological validity of the training. Finally, synchronization with real-time Digital Twins will allow the framework to inform safety planning and predictive maintenance in physical construction environments.

References

1. Scorgie, S., et al.: Virtual reality for safety training in construction. J. Comput. Civ. Eng. **38**(2) (2024)
2. Dincelli, E.: Presence by design: a framework for immersive VR training. Comput. Hum. Behav. **162** (2026)
3. Dickinson, J., et al.: Diegetic tool management in VR: preventing breaks in presence. Virtual Reality. **25**, 112–125 (2021)
4. Khorrami Shad, H., et al.: State-of-the-art of AR/VR in construction 4.0. Adv. Eng. Inform. (2024)
5. Ding, S., et al.: Augmented reality for enhancing educational experience in laboratory safety training. J. Loss Prev. Process Ind. **100**, 105848 (2026)
6. Cordeiro, A., et al.: Immersive technologies for height safety training. Autom. Constr. **155** (2025)
7. Morishita, Y., et al.: Impact of real-time feedback in VR simulators. Med. Educ. Online. (2025)
8. Gorup, G., Lesar, Ž.: Immersive questionnaire framework for perception research. Virtual Reality. **30** (2026)
9. Tao, L., et al.: Exploring students' self-regulated learning behavioural patterns and perceptions in an English speaking task within a generative AI-supported immersive VR environment. Comput. Educ. **225** (2026)
10. Ahmed, T., et al.: Efficacy of VR-based industrial safety training. Saf. Sci. **180** (2026)
11. Wang, H.: Gamified balance training in VR environments. IEEE Trans. Neural Syst. (2025)
12. Clifford, R.M.S., et al.: The effects of multi-sensory aerial firefighting training in virtual reality on situational awareness, workload, and presence. In: Proceeding of the IEEE VRW (2020)
13. Yigitbas, E., et al.: Comparative evaluation of AR-based, VR-based, and traditional basic life support training. In: Proceeding of the IEEE SEGAH (2022)
14. Bedard, M., et al.: HCI for public infrastructure decision-making. J. Manag. Eng. (2025)
15. Zhang, X., et al.: VR construction safety training and emotions. Autom. Constr. **148** (2024)
16. Abotaleb, I.S., Hosny, O.A.: VR integration in construction management. IEEE Access. **11** (2023)
17. Getuli, V., et al.: VR-based evaluation of construction safety training. J. Inf. Technol. Constr. **29** (2024)
18. Jakobo, E.P., Taifa, I.W.R.: Industry 4.0 in road construction. Eng. Rep. **7**(10) (2025)
19. ROBETARME Project: ROBETARME: Human-robot collaborative solutions for inspection and predictive maintenance in civil engineering. [Online]. https://www.robetarme-project.eu/ (2024). Accessed 07Jan 2026

Developing Spatial Imagination and STEM Skills Using Unmanned Technologies

Radek Nemec(✉)

University of Hradec Kralove, Hradec Králové, Czech Republic
radek.nemec@uhk.cz

Abstract. This narrative literature review presents a comprehensive synthesis of current knowledge in the field of educational robotics, focusing on unmanned aerial vehicles (UAV) and their impact on the cognitive architecture of the human brain. Based on a defined search strategy and inclusion criteria across relevant academic databases, this review examines the documented critical lack of spatial literacy in the population within the context of the emerging fourth and fifth industrial revolutions. Furthermore, it synthesizes existing literature proposing the implementation of drones as a key tool for neurocognitive development. Rather than providing new empirical data, this paper systematically reviews the mechanisms of manual piloting, FPV navigation, and drone programming, theoretically aligning these activities with Uttal's taxonomy of spatial abilities. Finally, it consolidates previous research on pilots' psychophysiological responses, including the phenomenon of cognitive dissonance during control inversion, to discuss broader implications for STEM education.

Keywords: Unmanned aerial vehicles · Drones · Spatial imagination · STEM · Education

1 Introduction

In the current period, often referred to as the Fourth Industrial Revolution (Industry 4.0), and with the imminent advent of Industry 5.0, the global labor market and, with it, education systems are undergoing unprecedented transformation. Automation, robotization, and digitization are changing the fundamental requirements for workforce competencies [1]. The World Economic Forum estimates that in the near future, up to a third of key job skills will be those that are not considered crucial today, with a significant proportion of employees requiring retraining [1]. In this context, the importance of STEM education (Science, Technology, Engineering, Mathematics) comes to the fore, but traditional pedagogical approaches often fail to maintain student interest and effectively develop critical cognitive abilities, especially spatial intelligence [2]. A practical way to sustain learners' engagement in demanding STEM topics is the use of gamified digital learning tools [3].

Spatial imagination, or the ability to generate, store, recall, and transform structured visual images, is one of the strongest predictors of success in technical fields. Studies

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 260–271, 2027.
https://doi.org/10.1007/978-3-032-32104-6_23

consistently show that students with developed spatial abilities exhibit higher retention rates in engineering programs and achieve better academic results [4]. Despite this, spatial education is often neglected in school curricula or limited to static tasks in geometry textbooks.

Unmanned Aerial Vehicles, popularly known as drones, are filling this void as a revolutionary educational tool. They represent a unique synthesis of the physical and digital worlds, allowing students to interact with three-dimensional space in ways that were previously unimaginable [2]. Unlike passively viewing 3D models on a computer screen, drones require active participation, whether in the form of manual piloting, which demands psychomotor coordination and rapid spatial orientation, or in the form of programming, which trains algorithmic thinking and trajectory visualization [5].

This report presents a comprehensive analysis of the mechanisms by which drones stimulate cognitive development. It draws on an extensive body of academic studies and technical reports to provide detailed insight into the neurocognitive processes associated with UAV control. The goal is not only to describe the benefits, but also to identify specific teaching strategies, cognitive barriers (such as spatial disorientation), and technological innovations (including brain-computer interfaces) that will define the future of spatial education.

2 Theoretical Framework: Anatomy of Spatial Cognition

In order to competently assess the impact of drones on brain development and skills, we must first break down the concept of spatial intelligence itself and place it within the relevant theoretical models of cognitive psychology and neuroscience.

2.1 Taxonomy of Spatial Abilities According to Uttal

In the professional literature, the typology proposed by Uttal et al. [6] is most commonly used to classify spatial skills, which is key in the context of drone research for understanding what specific sub-skills are being trained. This taxonomy distinguishes abilities based on two dimensions: intrinsic vs. extrinsic and static vs. dynamic.

1. Intrinsic-Dynamic: This category includes the ability to mentally transform, rotate, or deform individual objects. A typical example is mental rotation, i.e., the ability to imagine how an object looks from a different angle. In the context of drones, this ability is critical for manual piloting, where the pilot must constantly recalculate the drone's orientation relative to their own position [6, 7].
2. Extrinsic-Dynamic: This involves understanding the relationships between different objects in motion or the observer's navigation through space. This component, often referred to as perspective taking or spatial orientation, is essential for understanding flight paths, multi-drone formations, and the drone's relationship to terrain obstacles [8].
3. Intrinsic-Static: The ability to perceive and analyze the structure of an object without transforming it (e.g., perceiving geometric shapes). Students use this ability when constructing drones or analyzing their components [1].

4. Extrinsic-Static: Understanding the position of objects in space, e.g., reading a map. This ability is fundamental when planning autonomous missions in photogrammetry [9].

Drones are unique in that, as an educational tool, they cover all four quadrants of this taxonomy, placing extreme emphasis on the dynamic components that are least represented in traditional teaching.

2.2 Neurobiological Foundations and Hippocampal Navigation

Although most educational studies focus on behavioral outcomes, it is necessary to mention the neurobiological basis of spatial orientation. Spatial navigation is controlled by a complex network of brain structures, with the hippocampus and parahippocampal cortex playing a dominant role. These areas are responsible for the creation of so-called cognitive maps – mental representations of the environment [10].

Research suggests that active navigation (such as that performed by a drone pilot) stimulates neuroplasticity in these areas more effectively than passive observation. Studies confirm that different reference frames (egocentric vs. allocentric, see Sect. 4.1) activate different neural circuits in the parietal cortex [11]. Drone piloting can thus be understood as intensive neurocognitive training that strengthens the structural integrity of the areas of the brain responsible for spatial memory and orientation.

2.3 Embodied Cognition Theory and Motor Simulation

An important theoretical pillar for the use of manually controlled drones is embodied cognition theory. This theory postulates that cognitive processes are not isolated in the brain but are deeply rooted in the body's interactions with the environment ([7]. Thinking is shaped by the motor system.

Studies by Wiedenbauer and Jansen-Osmann [7], cited in the context of drone research, have shown that manual rotation training (physically rotating an object) leads to improvements in mental rotation tasks, even when visual stimulation remains the same [7]. This suggests that the motor activity associated with controlling a drone (manipulating the transmitter's levers) is not just a means of controlling the machine, but directly constitutes the process of spatial thinking. When a student moves the lever to the left to make the drone fly to the left, their brain encodes the spatial relationship through motor action. The drone becomes an extension of the pilot's body ("tool embodiment"), which allows spatial problems (e.g., flying through a narrow slit) to be solved more intuitively than with a purely abstract solution.

3 Psychophysiology of Manual Piloting: Cognitive Dissonance and Adaptation

Manual drone control, especially in Line of Sight (LOS) mode, is one of the most cognitively demanding tasks for beginners. It requires simultaneous processing of visual data, maintaining orientation, motor execution, and constant prediction of the machine's physical behavior [12].

3.1 The Phenomenon of Control Reversal and Dynamic Mental Rotation

A key mechanism that makes manual flight such an effective tool for developing spatial intelligence is the need to solve the so-called control reversal problem [1, 13].

When the drone flies away from the pilot (its orientation is the same as the pilot's), the controls are congruent: moving the stick to the right causes the drone to move to the right from the pilot's perspective. The cognitive load is relatively low at this point. A dramatic change occurs when the drone turns to face the pilot (so-called "nose-in" orientation). In this configuration, there is a discrepancy between the pilot's egocentric frame of reference and the drone's egocentric frame.

- Cognitive conflict: The pilot wants the drone to move to the right (from their perspective). If they move the stick to the right, the drone will move to the right from their perspective, but to the left for the pilot.
- Real-time mental rotation: To prevent a crash, the pilot must perform a mental rotation. They must imagine themselves in the drone's cockpit or mentally rotate the control vector by 180°. This operation must take place in a fraction of a second [6, 14].

This process is not static as in the Shepard-Metzler test (where the subject has unlimited time to decide whether two objects are identical). When flying a drone, mental rotation is under extreme time pressure and involves a high degree of risk (collision). Repeated exposure to this situation forces the brain to automate the process of coordinate transformation, resulting in massive training of spatial plasticity [7]. Experienced pilots no longer perform conscious rotation, but have developed intuitive mapping ("move the stick to the lower wing to level off"), which demonstrates a deep internalization of spatial relationships.

3.2 Line of Sight (LOS) Flying as Cognitive Training

Line of Sight (LOS) flying is often considered more difficult than flying with FPV goggles precisely because of the need to constantly recalculate orientation [15]. Discussions in pilot communities and academic studies confirm that while FPV provides immersion, LOS builds fundamental situational awareness (SA) [12].

- Parallax and distance estimation: In LOS flight, the pilot must estimate the distance of the drone from obstacles based on the size of the drone's image on the retina and binocular disparity. As distance increases, these cues become less reliable, forcing the pilot to rely on contextual cues and knowledge of the environment (allocentric orientation) [10].
- Training transfer: Studies suggest that skills acquired in LOS flying (particularly mental rotation) are transferable to other domains requiring spatial orientation, such as surgery, crane operation, or sports [7].

3.3 Spatial Disorientation in Aviation and Its Parallels in Drones

The phenomenon of spatial disorientation (SD), which causes 5–10% of accidents in general aviation (90% of which are fatal), is also relevant for drone operators [10,

16]. SD occurs when there is a mismatch between the visual, vestibular (balance), and proprioceptive (muscle) systems.

Although the drone pilot sits on the ground and their vestibular system is not affected by overload as in a fighter pilot, visual disorientation occurs.

- Loss of context: If the drone is flying at a great distance and the sky is homogeneous (cloudless), the pilot loses reference points for determining orientation (which side is forward?). This leads to panic and incorrect control inputs [8].
- Cognitive tunneling: Under high stress, the pilot may focus so intensely on maintaining altitude that they lose awareness of horizontal position and wind drift [10].

Research into SD management strategies for geologists using drones has shown that the key is to maintain a fixed body orientation (cardinal orientation) and use prominent landmarks in the terrain to quickly recalibrate the mental map [8].

3.4 Biometric Analysis of Pilots: EEG and Eye Tracking

Modern research uses advanced biometric methods to objectify cognitive load during piloting. A study by Dalilian and Nembhard [1] combined EEG (electroencephalography) and eye-tracking to analyze the process of pilot skill acquisition [1].

- Visual scanning: Novice pilots exhibit chaotic eye movement patterns with long fixations on irrelevant elements, while experts have efficient, "economical" scanning strategies focused on key telemetry data and drone position.
- Cognitive load (EEG): Analysis of brain waves (particularly in the theta and alpha bands) shows that cognitive load is highest in the early stages of training and when dealing with crisis situations (e.g., loss of signal). With increasing practice, there is a decrease in activity in the prefrontal cortex (planning) and a shift of control to the motor centers, indicating automation.
- Fatigue detection: Studies using convolutional neural networks (CNN) to analyze pilots' EEG data have shown that with time spent in the simulator, there is a measurable decrease in attention and an increase in error rates, which has implications for operational safety and training lesson design [17].

4 Immersive technology and First-Person View (FPV): Alteration of perception

First-Person View (FPV) technology, where the pilot watches the drone's camera feed in real time through video goggles, fundamentally changes the nature of spatial interaction. The pilot "becomes" the drone, which eliminates the problem of control inversion but brings new cognitive challenges and benefits [18].

4.1 Egocentric vs. Allocentric Reference Frames

In FPV flying, the egocentric reference frame (relative to the drone's "body") is dominant. The pilot perceives the world from the aircraft's perspective. This facilitates immediate maneuvering but can lead to a loss of global situational awareness (allocentric orientation). The pilot may fly perfectly through a building window but lose track of where the building is located relative to the takeoff point [11, 16].

Research shows that the ability to switch flexibly between these frames is a sign of advanced spatial intelligence. Virtual reality (VR) experiments suggest that contextual cues (e.g., landmarks on the horizon) help maintain an allocentric map even during egocentric navigation [11]. FPV flying thus trains a specific type of "local" spatial visualization and rapid response to visual stimuli.

4.2 Flow and Depth Perception: A Case Study of Matte Painting

A unique study on the impact of FPV on art education was conducted by researchers in a matte painting course (creating digital backgrounds for film). Students used FPV drones to explore environments that they then digitally reproduced [18].

The results showed that:

- Improved depth perception: FPV flight allowed students to perceive the 3D structure of the landscape (parallax, object occlusion) better than when simply looking at photographs.
- Flow state: Students using FPV showed significantly higher levels of "flow" (a state of optimal experience and concentration). This immersive experience led to higher motivation ("technology incentive") and a deeper understanding of spatial relationships in the landscape [18].
- Transfer to digital creation: The flight experience was directly reflected in the quality of their 3D models, which were more realistic and spatially coherent.

4.3 Visual Attention and Prediction: Humans vs. Artificial Intelligence

An interesting insight into human spatial cognition is provided by a comparison of human pilots with autonomous AI systems in drone racing. A study published in 2025 shows that human pilots still outperform neural networks in many respects, thanks to their ability to focus their visual attention selectively [19].

Human pilots do not scan the entire image evenly, but fix their gaze on key points (apexes of turns, obstacles). This ability to predict trajectory and ignore visual noise is key to fast flight. Researchers used human pilot eye movement data to train an "attention-prediction" model for an autonomous drone, which led to an increase in track completion success from 61% (pure RGB inputs) to 88% [19]. This confirms that effective spatial navigation is not just about processing the entire image, but about strategically selecting relevant spatial information.

5 Programmable Drones: Algorithmic Space Construction

While manual piloting is the domain of reflexes and intuition, drone programming (Coding Drones) shifts the focus to analytical thinking and executive functions. This approach is consistent with the concept of Computational Thinking, but enriches it with a physical spatial dimension [5].

5.1 From Manipulation to Abstraction: Computational Thinking in 3D

When programming a drone, students do not control the flight in real time, but must plan it in advance (Pre-flight Planning). This requires a high degree of spatial visualization and sequencing [1]. Comparable demands on spatial visualization are also when students learn 3D modeling for 3D printing [20]. The student must:

1. Create a mental model of the environment and the desired trajectory.
2. Decompose this model into discrete commands (take off, move forward 100 cm, rotate 90° to the right).
3. Simulate the execution of these commands in their mind (mental simulation) to reveal potential collisions [5].

This process is cognitively different from manual flight. While manual piloting burdens working memory with the need for quick corrections, programming burdens the ability to keep a sequence of steps and their spatial consequences in memory.

5.2 Geometric Interpretation of Code and Coordinate Systems

Programmable drones (e.g., DJI Tello using DroneBlocks or Python) are an ideal tool for teaching 3D coordinate systems (X, Y, Z axes) and vector geometry [21].

- Relative vs. absolute coordinates: Most simple drones use relative commands (move 50 cm forward from the current position). This forces students to constantly update the mental origin of the coordinates. More advanced systems allow navigation to absolute points, which requires an understanding of the global coordinate system [22].
- Rotation and orientation (Yaw): The command rotate_clockwise(90) changes the orientation of the drone, which in turn changes the meaning of the forward() command. This is a practical demonstration of local coordinate systems and transformation matrices, which are the basis of robotics and computer graphics.
- Mathematical applications: Drones can be used to visualize mathematical functions. For example, flying along a sine wave or circle trajectory requires the application of trigonometric functions and parametric equations in the real world [21].

5.3 Sequencing and Debugging Errors as Spatial Tasks

Debugging code for drones has specific characteristics. An error in the code is not only manifested by an error message "Syntax Error", but also by unexpected physical behavior in space (the drone crashes into a wall or flies in the opposite direction) [5].

This immediate and tangible feedback forces the student to analyze the discrepancy between the intended trajectory (mental model) and the actual trajectory (reality). They must identify which command in the sequence caused the spatial deviation. This process reinforces causal thinking and spatial analysis skills.

5.4 Empirical Evidence: Study of the Impact on Elementary School Students

Key evidence of the effectiveness of this approach comes from a study conducted in Taiwan that examined the impact of a six-week drone programming course on third-grade students [5]. The researchers used a pre-test and post-test design to measure spatial abilities. Such repeated measurements can be effectively performed, for example, through an LMS (e.g., Moodle) using an external tool [23].

Results: A significant increase ($p < 0.01$) was observed in spatial visualization and sequencing tests. The average difference in spatial visualization scores was 5.4 points, indicating a strong effect of the intervention [5]. The study confirmed that drone programming is an effective method for developing these skills at an early age (Table 1).

Table 1. Comparison of the cognitive demands of manual and programmed flight

Cognitive dimension	Manual Flight (LOS/FPV)	Programmed Flight (Coding)
Type of processing	Reactive, Real-time, Implicit	Analytical, Planned, Explicit
Cognitive load	High instantaneous load (maintaining stability)	High load during planning, low during flight
Mental rotation	Critical (solving control inversion)	Medium (prediction of orientation after rotation)
Reference frame	Constant switching Egocentric/Allocentric	Predominantly Allocentric (view from above)
Error rate	Risk of crash due to slow reaction	Risk of logical error in command sequence
Primary skill developed	Psychomotor skills, Spatial orientation, SA	Computational thinking, Geometric visualization

6 Applied Photogrammetry and Geospatial Sciences

A specific area where drones are becoming an indispensable tool is the collection of data for the creation of 3D models of terrain and objects. This involves combining manual piloting with subsequent data processing, which develops a comprehensive understanding of spatial relationships on a macro scale [9].

6.1 Reconstructing Reality: From 2D Images to 3D Models

The process of photogrammetry requires students to understand the principles of how to reconstruct 3D space from flat 2D photographs.

- Overlap and angles: Students learn that for a high-quality model, it is necessary to ensure sufficient image overlap (e.g., 75%) and correct shooting angles. They must visualize the "stereo" effect on a huge scale [24].
- Geospatial thinking: Working with orthophotomaps and digital surface models (DSMs) develops the ability to read terrain, understand contour lines, and measure volumes (e.g., the volume of material extracted from a quarry) [9]. Research at Clemson University has shown that virtual modules for teaching UAVs in forestry effectively prepare students to work with spatial data, even if they do not have physical access to a drone [9].

6.2 Cognitive Challenges in Field Data Collection

Fieldwork presents specific challenges. Geologists and researchers often face the problem of losing their bearings when monitoring a drone and moving around the terrain at the same time. A study by Cheng and Newcombe [8] in the context of UAV geology points out that the need to simultaneously monitor the drone, check telemetry, and perceive one's own position in the terrain can overload cognitive capacity and lead to errors in judgment [8]. Training with drones thus teaches students strategies for divided attention (multitasking) and effective management of cognitive resources.

7 Advanced Interfaces and the Future: Brain-Computer Interfaces (BCI)

Brain-computer interfaces represent a whole new frontier in human-drone interaction, allowing UAVs to be controlled using only thoughts. Although this is an experimental technology, its implications for cognitive training are fascinating [11, 13].

7.1 Mind Control: Motor Imagery and Spatial Navigation

BCI systems use EEG caps to sense the brain's electrical activity. Motor Imagery (MI) – imagining movement – is the technique most commonly used to control drones.

- Mental commands: The user must intensely imagine, for example, the movement of their left hand to make the drone turn left, or the movement of both hands to make it climb. This is not about "thinking of the word left," but about activating the motor cortex without performing a physical movement [11].
- Spatial concentration: This method requires an extreme degree of concentration and the ability to visualize space vividly. The user must maintain a mental image of the movement for several seconds. BCI training thus functions as an intensive "gym" for specific neural networks in the motor and parietal cortex [13].

7.2 Hybrid Neural Networks and Intention Decoding

Current research focuses on the use of hybrid neural networks (e.g., CNN-LSTM) for more accurate decoding of these signals and their conversion into smooth flight commands [11, 17]. The combination of BCI with gyroscopic sensors that detect head movements (Head Movement) allows for more intuitive control of the drone's camera, further strengthening the connection between the pilot and the machine and creating a sense of "telepresence."

8 Curricular Integration and Pedagogical Strategies

The effective use of drones in teaching requires thoughtful integration into the curriculum. Simply "flying for fun" does not bring lasting cognitive benefits.

8.1 Differences in Impact on Engineering and Architecture

Studies comparing engineering and architecture students show interesting differences.

- Architecture: Students benefit greatly from spatial skills training, which is reflected in better grades in design studios. Drones help them understand the context of a building and spatial relationships in the landscape [4].
- Engineering: For engineering students, training led to improvements in spatial tests, but this did not always translate directly into results in basic courses (calculus, physics). This suggests that spatial intelligence is necessary for engineers, but other factors also play a role in academic success in the early years [4].
- Gender aspects: Research suggests that although men often score higher on mental rotation tests, targeted training (including drones) can erase these differences. Projects such as "She Maps" target girls and demonstrate that drones can increase their confidence in technical tasks [25].

8.2 Teacher Readiness and the TPACK Framework

A key barrier is teacher readiness. The TPACK (Technological Pedagogical Content Knowledge) framework is used to assess teachers' ability to integrate drones into teaching. Studies show that teachers often need intensive training not only in operating drones, but above all in the methodology of using them to explain abstract STEM concepts [26–31]. Without methodological guidance, there is a risk that drones will become nothing more than a disruptive element.

9 Conclusion and Recommendations for Educational Practice

A synthesis of the analyzed studies [2, 4–6] confirms that drones are an extremely effective tool for developing spatial imagination. This is not a monolithic intervention; different modes of use develop different cognitive components:

1. Manual flight (LOS): This is irreplaceable for the development of rapid mental rotation, coping with cognitive dissonance (inversion), and psychomotor coordination. It is recommended as an introductory activity for understanding the physics of flight and building stress resilience.
2. Drone programming: Ideal for developing spatial visualization, algorithmic thinking, and understanding geometric principles in 3D. It should follow manual flight or run in parallel.
3. FPV and Photogrammetry: Develops depth perception and the ability to transform between 2D and 3D representations, which is key for fields such as architecture and geography.

Student engagement is crucial for successful education [32]. To maximize educational impact, it is essential that the curriculum includes a reflective component—students should not just fly, but analyze their decisions, plan missions, and evaluate mistakes. In an era where the boundaries between the physical and digital worlds are blurring, drones provide the ideal training ground for the 21st-century mind.

Acknowledgements. The article was funded by the Faculty of Education, University of Hradec Králové.

References

1. Dalilian, F., Nembhard, D.: Biometrically Measured Affect for Screen-Based Drone Pilot Skill Acquisition. University of Iowa, Iowa City (2023)
2. Edulakanti, M., Ganguly, R.: Drone-based learning: a review of the educational potential of UAVs. J. Educ. Technol. Syst. (2023)
3. Voborník, P.: New modern application for Gamified memorizing paired facts by more fun form of memory game with the possibility of own content development. In: DIVAI 2020 – The 13th International Scientific Conference on Distance Learning in Applied Informatics, pp. 217–231. Wolters Kluwer, Štúrovo (2020)
4. Gómez-Tone, H.C., et al.: Spatial Ability and Academic Performance in First-Year Engineering and Architecture Students. Education Sciences (2024)
5. Chou, P.: Smart technology for sustainable curriculum: using drone to support young students' learning. Sustainability. **10**(10), 3819 (2018)
6. Uttal, D.H., et al.: The malleability of spatial skills: a meta-analysis of training studies. Psychol. Bull. **139**(2), 352–402 (2013)
7. Wiedenbauer, G., Jansen-Osmann, P.: Manual training of mental rotation in children. Learn. Instr. **18**(1), 30–41 (2008)
8. Newcombe, N.S., Shipley, T.F.: Thinking about spatial thinking: New typology, new assessments. In: Studying Visual and Spatial Reasoning for Design Creativity, pp. 179–192. Springer Netherlands, Dordrecht (2014)
9. He, Y., et al.: Developing a virtual UAV education module for natural resource students. J. For. Educ. (2022)
10. Epstein, R.A., Patai, E.Z., Julian, J.B., Spiers, H.J.: The cognitive map in humans: spatial navigation and beyond. Nat. Neurosci. **20**(11), 1504–1513 (2017)
11. Gramann, K., Onton, J., Riccobon, D., Mueller, H.J., Bardins, S., Makeig, S.: Human brain dynamics accompanying use of egocentric and allocentric reference frames during navigation. J. Cogn. Neurosci. **22**(12), 2836–2849 (2010)

12. Brouwers, S., Wiggins, M.W., Helton, W., O'Hare, D., Griffin, B.: Cue utilization and cognitive load in novel task performance. Front. Psychol. **7**, 435 (2016)
13. Goga, N.: Drone control based on mental commands and facial expressions. In: Proc. 12th Int. Conf. Electron., Comput. Artif. Intell. (ECAI), pp. 1–4 (2020)
14. Shepard, R.N., Metzler, J.: Mental rotation of three-dimensional objects. Science. **171**(3972), 701–703 (1971)
15. Cho, K., Cho, M., Jeon, J.: Fly a drone safely: evaluation of an embodied egocentric drone controller interface. Interact. Comput. **29**(3), 345–354 (2017)
16. FAA: Normalizing Unmanned Aircraft Systems Beyond Visual Line of Sight Operations. Federal Register (2025)
17. Wang, Y., et al.: Cognitive load recognition utilizing EEG signals in simulated flight missions. Front. Neurosci. (2024)
18. Fan, J., et al.: The use of first-person view (FPV) drones in matte painting courses. PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC9326556/, last accessed 2025/01/04
19. Xiao, J.: Vision-Based Learning for Drones: A Survey. ResearchGate (2025)
20. Němec, R., Voborník, P.: Using 3D modeling software for 3D printers at primary (lower secondary) schools in the Czech Republic and use it for create simple model. Int. J. Educ. Inf. Technol. **12**, 80–85 (2018)
21. Pepkolaj, L.: Adopting Drone Technology in Mathematical Education. Online (2021)
22. Qiu, H., et al.: AVR: augmented vehicular reality. In: Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '18), pp. 81–95. ACM, Munich (2018)
23. Voborník, P.: Universal testing environment as an external tool of moodle. In: 10th International Scientific Conference on Distance Learning in Applied Informatics (DIVAI 2014), pp. 215–225. Wolters Kluwer, Prague (2014)
24. Tucci, G., et al.: Guidelines for UAV Mapping Flights. Remote Sensing (2023)
25. She Maps: Drones in the Classroom: Teaching STEM by Stealth. https://shemaps.com/blog/drones-in-the-classroom-teaching-stem-by-stealth/, last accessed 2025/01/04
26. Ng, W., Cheng, G.: Integrating drone technology in STEM education: a case study to assess teachers' readiness and training needs. Issues Inform. Sci. Inf. Technol. **16**, 61–70 (2019)
27. Němec, R., Berková, A., Hubálovský, Š.: The use of method of multidisciplinary approach, modeling and simulation in learning of gear ratios. Int. J. Math. Comput. Simul. **10**, 46–52 (2016) ISSN 1998-0159.
28. Coufal, P.: Project-based STEM learning using educational robotics as the development of student problem-solving competence. Mathematics. **10**, 4618 (2022)
29. Němec, R., Šrámek, F., Berková, A.: Mapping the multi-instrumental approaches to teaching at primary (lower secondary) schools. In: 20th International Conference on Circuits, Systems, Communications and Computers (CSCC 2016): EDP Sciences – Web of Conferences (2016)
30. Jahodová Berková, A.: Effect of the use of computer-aided assessment system in the teaching of mathematical analysis with regard to students' approaches to learning. J. Effic. Responsib. Educ. Sci. **10**(3), 71–75 (2017) ISSN: 2336-2375
31. Turčáni, M., Balogh, Z., Kohútek, M.: Evaluating computer science students reading comprehension of educational multimedia-enhanced text using scalable eye-tracking methodology. Smart Learn. Environ. **11**, 29 (2024). https://doi.org/10.1186/s40561-024-00318-5
32. Milkova, E.: Multimedia applications - effective support of education. In: DIVAI 2012: 9th International Scientific Conference on Distance Learning in Applied Informatics, pp. 13–21

Kandó Cloud Campus: A Project-Based, Distributed-Resource Higher-Education Ecosystem

István Németh, Gergely Sebestyén, Dávid Sik, and György Molnár(✉)

Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Tavaszmező u. 17, 1084 Budapest, Hungary
{nemeth.istvan,sebestyen.gergely,sik.david}@kvk.uni-obuda.hu, molnar.gyorgy@uni-obuda.hu

Abstract. The Kandó Cloud Campus initiative proposes a nationwide, network-based model for project-driven higher education and knowledge transfer that tightly couples universities, regional economic actors, and industrial partners. The concept responds to rapid shifts in labour-market skill demand and to the growth of decentralized, micro-credential based lifelong learning pathways [1, 2, 4]. Its central technical element is a knowledge-graph powered collective "second brain" that stores modular knowledge units ("knowledge molecules") and their relations, enabling data-driven personalization and cooperation across institutions. Learning is organized around industry-relevant project portfolios that map to curricular outcomes via a transparent credit-allocation matrix; assessment is performed jointly by academic mentors and external experts, similarly to thesis supervision. The platform further introduces learner and project digital twins: with user consent, multimodal learning-analytics signals can be used to train AI tutors that recommend optimal individual learning paths and support mentoring rather than replacing educators. We describe a layered reference architecture (security/validation, data, operations, services, interfaces), the regional governance model (coordination forum and advisory board), and an implementation roadmap from pilot sites to scalable operation. The approach aims to increase responsiveness, reduce time-to-competence, and strengthen regional innovation ecosystems by embedding education, research, and industrial problem solving into a shared cloud campus.

Keywords: project-based learning · distributed cooperative learning environments · knowledge graphs · digital twins · micro-credentials · lifelong learning · regional innovation ecosystems

1 Introduction

Higher education is entering a phase where curricula, skills, and occupational profiles evolve faster than traditional programme cycles. The World Economic Forum projects large-scale reskilling needs and significant changes in the skills required to perform existing jobs in the coming years [1]. At the same time, employers increasingly seek demonstrable competences and problem-solving experience, not only formal degrees. These

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 272–284, 2027.
https://doi.org/10.1007/978-3-032-32104-6_24

pressures coincide with technological enablers—AI, pervasive connectivity, virtual collaboration tools—that make the creation, distribution, and validation of knowledge more decentralized than before [2].

In parallel, policy and practice are moving towards stackable learning and micro-credentials, allowing learners to build recognized skill bundles through shorter, targeted learning units across a lifetime [2, 4, 5]. For engineering education, this shift also implies stronger coupling between universities and regional innovation ecosystems: education, applied research, and industrial problem solving must interact as a single value chain. Evidence from European research universities indicates that universities generate substantial economic value through both education and research-to-industry transfer [3].

This paper presents an English, shortened version of the "Kandó Felhőkampusz" concept originally developed for the Kandó Kálmán Faculty of Electrical Engineering (Óbuda University). The proposal is a project-based, distributed-resource "cloud campus" ecosystem in which (i) learning is organized around real projects defined with industry and societal partners, (ii) multiple universities and sites can jointly operate competence clusters, and (iii) a knowledge-graph driven collective "second brain" supports personalization, cooperation, and continuous renewal of learning resources.

The core claim is that a modern engineering faculty can no longer optimize its educational output by tuning internal processes alone. Instead, it must co-design curricula with the innovation ecosystem that will absorb graduates, and it must provide an operational interface to that ecosystem: a structured pipeline for projects, mentoring, assessment, and knowledge capture. The Cloud Campus is proposed as that interface.

2 Related Work - Economic and Policy Context - Vision

Universities contribute to economic development not only through graduates but also through knowledge creation and transfer. A detailed analysis of 23 European research universities reported a gross value added of 99.8 billion EUR and 1.3 million jobs created in the EU in the 2014–2016 period [3]. Extrapolations suggested that the wider European university sector contributed on the order of 400 billion EUR (about 2.7% of total gross value added) and roughly 5.1 million jobs (about 2.2% of EU employment) [3].

Beyond scale, the same analysis emphasized the high return of research and innovation transfer: in 2016 the investigated universities generated nearly 5 EUR of gross value added per 1 EUR invested, and around 33 billion EUR of value added via licensing, spin-offs, and research services [3]. For an engineering university, these numbers imply that educational modernization and industrial integration are not peripheral activities; they are central to regional competitiveness.

At the policy level, several trends reinforce the need for a flexible educational delivery model. First, micro-credentials are increasingly recognized as a mechanism to document learning outcomes in smaller, stackable units, with explicit relevance to employability and lifelong learning [4]. Second, strategic frameworks such as OECD's Education 2030 emphasize adaptive skills and learning-to-learn capabilities, not only domain facts [2]. Third, UNESCO's calls for a renewed social contract for education highlight inclusion, collaboration, and system-level responsiveness [5]. The Cloud Campus concept is positioned to operationalize these directions in an engineering context.

In the proposed vision, a university is not primarily a collection of isolated degree programmes. Instead, it acts as a networked knowledge hub that is embedded in its region and interoperable with other hubs. Its main product is verified competence: the ability to apply knowledge in complex, collaborative contexts. Achieving this requires that learning outcomes be continuously aligned with real problem portfolios emerging from industry, public services, and research challenges.

The Kandó Cloud Campus therefore builds on three design principles that reframe higher-education delivery:

- Project primacy: curricula are expressed through project portfolios that operationalize learning outcomes as deliverables and measurable competences [8].
- Distributed resources: teaching capacity, laboratories, datasets, mentors, and industrial problems are shared across sites and partner institutions to reduce fragmentation and increase responsiveness.
- Personalized yet cooperative learning: each learner follows an individualized path through a common knowledge graph while working in teams on projects, combining personalization with social learning.

3 Methodology for Building a Project-Driven Distributed Ecosystem and for Network-Centric Learning - the Model

Traditional higher education is optimized for relatively stable bodies of knowledge and for scheduled delivery formats. However, the pace of technological change shifts the bottleneck from teaching capacity to curricular agility. As a result, educational "manufacturing technology" has evolved through several paradigms often described as Learning 1.0 (teacher-centric transmission), Learning 2.0 (interactive and outcomes-oriented), and Learning 3.0 (digitally supported, data-informed, and increasingly personalized). Modern systems can already support a form of mass customization: group education where individual progress is monitored and supported [6].

The next step is deeper integration: educational units become tightly coupled to the economic and societal processes they serve. In an engineering ecosystem, this means that projects, internships, prototyping cycles, and applied research create a continuous feedback loop between curricula and practice. Project-based learning literature provides strong foundations for this approach, emphasizing authentic tasks, learner autonomy, and reflective practice [7, 8].

Connectivist approaches add an operational perspective: knowledge is not a static set of topics but a network of nodes (concepts, resources, people) and connections that can be strengthened or rewired [9]. In a Cloud Campus, the knowledge graph makes this idea explicit: it is both a representation of the domain and a scaffold for learning pathways. Finally, open educational resources (OER) and modular authoring allow rapid integration of new content, reducing dependency on slow textbook cycles [11].

Implementing project-driven education at scale requires coordinated development of governance, partner engagement, and learning infrastructure. The methodology below is intended to be repeatable across regions and competence clusters.

1. Identify regions and local sites where the university's educational and research profile can be embedded with measurable impact (economic structure, demographics, industrial base, and complementary educational providers).
2. Map potential industrial and societal partners and their medium-term problem portfolios, including opportunities for joint labs, open innovation projects, and co-supervised student work.
3. Define competence clusters (programme groups) that reflect regional needs; within each cluster, design degree and micro-credential pathways with maximal permeability for learners.
4. Establish a Regional Coordination Forum that includes university leadership, regional/municipal stakeholders, partner universities, and companies; the forum sets goals, approves annual budgets, and monitors outcomes.
5. Create a Professional Advisory Board (domain experts from academia and industry) that validates project portfolios, learning outcomes, and assessment rubrics and supports quality assurance.

Two cross-cutting tasks accompany every step. First, a capability inventory is needed: what laboratories, software, datasets, and mentorship capacities are available across nodes, and what gaps exist. Second, a change-management plan is required, because staff roles shift from primarily lecturing to mentoring, project leadership, and knowledge-base development.

At a given training site, competence clusters replace isolated programmes. A cluster groups programmes and micro-credentials that share project portfolios and laboratory resources (e.g., energy systems, embedded AI, cyber-physical systems). Within a cluster, curricula are designed for permeability: learners can pivot between closely related qualifications based on accumulated micro-credits, rather than restarting from scratch.

Learner progression is organized by projects rather than timetabled classes. A learner enters a project team, receives a role, and is guided by an academic mentor and, where relevant, an external industrial mentor. The team produces deliverables (design documents, code, prototypes, test reports) that are assessed using rubrics aligned to learning outcomes. This is similar to capstone evaluation, but applied continuously across the whole programme. Industrial partners benefit because projects are drawn from real needs and can be scoped to feasibility. Universities benefit because projects act as vehicles for integrating research results into teaching, and for creating applied research pipelines.

4 Curriculum as a Project Portfolio and Credit-Allocation Matrix

In a project-portfolio model, a learner does not primarily attend scheduled lectures; instead, the learner works in rotating teams on projects led by academic mentors and external experts. Each project is associated with a set of partial credits (micro-credits) across multiple curricular elements. A credit-allocation matrix specifies how project deliverables contribute to learning outcomes and formal credit.

Assessment resembles the evaluation of a thesis or capstone: academic and industrial supervisors jointly evaluate the deliverables, teamwork process, and demonstrated competences. Successful completion awards the corresponding micro-credits. Over

time, accumulated micro-credits satisfy degree requirements or can be stacked into micro-credentials aligned with labour-market needs [4].

The model also enables institutional permeability. Where partner universities agree on interoperability rules, learners may obtain micro-credits for projects executed in another institution's environment, and final degree issuance can reflect cross-institution contributions. Such permeability supports distributed competence clusters and helps regions access specialized expertise without duplicating full programmes.

To illustrate the mechanism, consider a learner admitted to an engineering programme with a predefined set of curricular components (the standard curriculum). Instead of taking each component as a standalone course, the learner selects a set of projects in each period. Project completion yields micro-credits for multiple curriculum components based on the allocation matrix. Over several periods, the learner accumulates a unique credit vector. This vector may satisfy the originally targeted qualification, or it may satisfy a closely related qualification—in which case the learner can switch pathways with minimal friction. Such flexibility is valuable when early project choices reveal stronger fit for a related discipline, or when labour-market demand shifts.

In multi-institution settings, allocation matrices can be institution-specific while remaining interoperable via agreed mapping rules. A learner might start at one institution but complete the final qualification at another, based on project availability, mobility, or regional relocation. This extends traditional exchange programmes by allowing continuous credit accumulation through shared project marketplaces.

5 Reference Architecture of the Kandó Cloud Campus

The Cloud Campus is conceived as a layered system. The layers separate trust and compliance concerns from learning services, enabling modular development and gradual adoption. While specific implementations may vary, the following functional layers are proposed.

- Security and validation layer: identity management, access control, consent handling, audit logs, data integrity checks, and policy enforcement.
- Data layer: (i) the collective knowledge graph ("second brain") and (ii) a learner-sensor databank storing consented multimodal signals for learning analytics.
- Operations layer: collaboration spaces where learners, mentors, and trained digital twins interact; APIs that expose knowledge-graph queries and project-workflow data.
- Service layer: AI-supported tutoring, recommendation, search, content authoring, project management, assessment support, and reporting.
- Interface layer: web/mobile clients, virtual-lab frontends, and partner integration endpoints (LMS, HR systems, credential registries).

Virtual laboratories and distributed infrastructure are integrated in the service layer. For engineering education, this includes remote-access measurement platforms, simulation environments, digital twin testbeds, and secure sandboxes for embedded and cyber-physical experiments. Such capabilities allow projects to be executed even when specialized hardware is located at another node in the network. The architecture intentionally separates "knowledge representation" from "learning delivery". This allows the

same knowledge graph to support multiple delivery modes: self-paced study, mentor-led project clinics, intensive micro-credential modules, and industry upskilling programmes. It also supports reuse of knowledge molecules across projects, reducing duplication and improving consistency.

5.1 Knowledge-Graph Based Collective Second Brain

The knowledge graph stores the learning and project knowledge required across competence clusters. Instead of monolithic lecture notes, the graph consists of modular knowledge units—"knowledge molecules"—that can be reused and recombined. Each molecule includes (a) explanatory content, (b) prerequisites, (c) links to projects and deliverables, (d) assessment hints, and (e) references to open educational resources where applicable [11]. Relations between molecules capture conceptual dependencies, alternative learning routes, and cross-domain bridges. In graph terms, learning is a controlled traversal from prerequisite nodes to target competence nodes, with projects acting as integrative checkpoints that require multi-node application. The ontology and representation choices follow established principles for knowledge sharing and reuse [14] and can leverage foundational work on knowledge representation [13]. Content creation is distributed: educators and researchers contribute molecules in their domain, and updates can enter the graph as soon as new research results or industrial practices appear—without waiting for a full textbook revision. This shortens the time between knowledge creation and curricular availability and supports continuous modernization of engineering education.

The knowledge graph can also support peer-based knowledge building. For example, learners can contribute annotations, alternative explanations, and links to resources; latent semantic analysis and related methods have been used to support such knowledge-building processes and can inform similarity search and recommendation within the graph [12]. However, contributions must be curated to preserve correctness, especially in safety-critical engineering domains.

5.2 Learner and Project Digital Twins, Learning Analytics, and Mentoring

A second innovation is the use of digital twins—AI agents trained alongside learners to support personalization and mentoring. With explicit consent, the platform can collect multimodal signals during learning activities (e.g., interaction traces, handwriting dynamics, eye-gaze proxies, and optional physiological indicators). These signals are analysed with machine-learning methods to infer learning state, cognitive load, and engagement trends [10, 15].

Digital twins are not intended to replace educators. Their primary role is to (i) recommend individualized routes through the knowledge graph, (ii) provide conversational access to graph content, (iii) detect potential knowledge gaps early, and (iv) escalate issues to the responsible educator by opening a one-to-one consultation channel within the collaboration environment. Educators remain the owners of learning goals, assessment decisions, and project leadership.

At the cohort level, aggregated analytics can help identify "critical nodes" in the knowledge graph where many learners struggle, motivating targeted improvements of

learning molecules or additional mentoring interventions. Because such analytics is sensitive, governance must enforce purpose limitation, transparency, and measurable benefit, and must provide opt-out options without academic penalty.

In addition to learner digital twins, project digital twins can capture the evolving state of a project: requirements, design decisions, tests, and risk registers. This supports traceability and enables learners to practice professional engineering documentation. It also supports portfolio evidence: graduates can present verified artefacts demonstrating competence, which complements (but does not replace) degrees and certificates (Fig. 1).

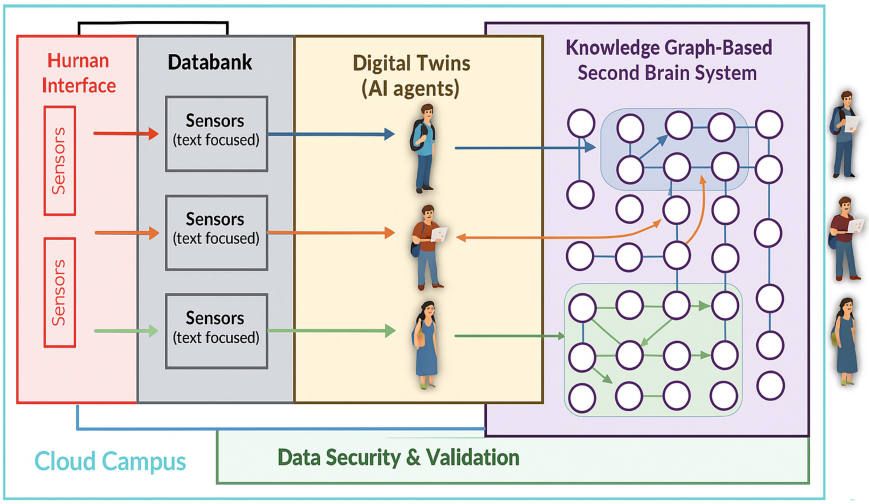


Fig. 1 Architecture of a cloud campus supporting a project-based cooperative teaching/learning environment (own diagram)

5.3 Data Governance, Trust, and Ethics

Trust is the enabling condition of a Cloud Campus. Because the platform integrates educational records, project evidence, and potentially sensitive learning-analytics signals, governance must be designed together with the technical architecture. The security and validation layer therefore is not a deployment detail but a first-class design artifact: it determines which actors can access which data under which purpose, and how this access is audited.

A minimal governance baseline should include: (i) explicit consent management for any non-essential data collection, (ii) role-based access control separating students, mentors, programme leads, administrators, and external experts, (iii) immutable audit logs for access and changes, (iv) retention schedules and deletion workflows, and (v) transparent user-facing explanations of what is collected and why. These requirements

become stricter if behavioural or physiological signals are used for digital-twin training. In such cases, the platform must allow participation in the educational core without participating in analytics, and must ensure that opting out does not disadvantage the learner.

The concept also implies a shift in what counts as educational record. In addition to grades, the system may store project artefacts, test logs, design rationales, and peer feedback. These can be valuable competence evidence, but they raise questions of intellectual property, confidentiality, and industrial sensitivity. Accordingly, each project should be associated with a data-classification profile and an agreed publication policy: which outputs remain internal, which can be shared within the university network, and which can become public portfolio items. For industrial projects, this can mirror standard non-disclosure practices while still allowing assessment transparency.

Algorithmic accountability is another requirement. Recommendation and tutoring functions must be evaluated for bias and failure modes. In engineering education, incorrect guidance can have downstream safety implications if it leads to persistent misconceptions. Therefore, AI outputs should be treated as decision support, not as authority: mentors remain responsible for final academic decisions and for interpreting analytics in context. Where possible, the system should provide explanations: why a learning route was recommended, which prerequisite gaps were detected, and which evidence supports a suggested intervention.

Finally, governance must support quality without turning the system into bureaucratic overhead. One strategy is to treat governance artifacts as reusable templates: standard consent forms, standard project classification profiles, standard assessment rubrics, and standard review workflows for knowledge molecules. This "governance-as-modular-infrastructure" approach matches the general philosophy of the Cloud Campus: modularity and reuse at the level of knowledge, projects, and operations.

6 Roles of Educators and the Human-in-the-Loop Model - Regional Embedding, Governance, and Sustainability

A frequent concern is whether an AI-supported Cloud Campus reduces the need for educators. The proposed model explicitly assumes the opposite: educators remain indispensable, but their work profile shifts. Instead of focusing on frontal delivery, educators become (i) project leaders and mentors, (ii) designers and maintainers of knowledge molecules, and (iii) stewards of assessment quality.

Mentoring is continuous. Educators monitor progress indirectly through project performance—whether the learner has the knowledge needed for the assigned role—and directly through learning analytics and digital-twin reports, while preserving privacy and consent. When gaps are detected, the educator can intervene with targeted support: short clinics, focused explanations, or assignment restructuring. This enables high-impact teaching effort to be concentrated where it matters most.

Educators also ensure that research results enter education rapidly. Because knowledge is integrated as modular molecules, new methods and findings can be published into the graph as soon as they are validated, rather than waiting for a new full course

material revision. This reduces the gap between research and education and helps keep project portfolios technologically current [11].

Distributed cloud-campus operation requires governance structures that align educational quality with regional impact. The proposed Regional Coordination Forum provides a formal mechanism for shared responsibility. Its tasks include: (i) defining target competence clusters and priority domains, (ii) approving annual operational budgets contributed by stakeholders, and (iii) reviewing annual performance reports.

- University leadership and faculty representatives (education, research, and infrastructure).
- Regional and municipal stakeholders (economic development, public services).
- Industrial partners contributing projects, mentors, datasets, or equipment.
- Partner universities offering complementary competence clusters or shared laboratories.

A Professional Advisory Board complements the forum by providing domain-level validation of project portfolios, learning outcomes, and assessment rubrics. Together, these bodies ensure that the university is not left alone with the cost of maintaining a modernized, project-driven offering, while also preserving academic standards and transparency.

Operationally, the forum produces an annual plan: prioritized project themes, target learner cohorts, resource commitments, and budget allocations. At year end, the operating units produce a report covering educational outcomes and partner value. This creates a feedback loop for continuous improvement and for adjusting the project marketplace to regional changes.

A sustainability opportunity is a service model where alumni and professionals retain access to parts of the cloud campus for upskilling, reskilling, and career development—an approach consistent with lifelong learning frameworks [2]. For employers, the system can support talent pipelines via competence evidence and optional "virtual job interview" project trials executed in the platform's virtual environment.

A further sustainability mechanism is subscription-style access to certain services (e.g., curated upskilling pathways, project clinics, credential issuance) for professionals and company cohorts. Such services can help finance continuous content maintenance and infrastructure modernization while extending the university's societal mission.

7 Service Portfolio and Use Cases

The Cloud Campus should be understood as a service portfolio rather than a single application. Different user groups—students, educators, alumni, industrial mentors, and employers—consume different slices of the portfolio. A modular service view helps implementation planning because services can be deployed incrementally and evaluated separately.

Core educational services (baseline for the pilot phase) include project team workspaces, knowledge-graph search and navigation, versioned resource authoring, and structured submission and assessment workflows. For project-driven education, these

services are the equivalent of the "operating system" that coordinates teams, evidence, and feedback.

Advanced services (typical for expansion and analytics phases) include AI-assisted tutoring, learning-route recommendation, automated formative quizzes linked to knowledge molecules, and mentor dashboards summarizing team progress and risk signals. In engineering domains with strong laboratory components, virtual-lab services are equally important: remote access to instruments, simulation environments, and digital twin testbeds that allow controlled experimentation from any node.

From the perspective of regional partners, the key use cases are:

- Industry-defined project marketplaces where partners publish problem statements with constraints, required competences, and expected deliverables; the university matches these to learner teams and mentor capacity.
- Co-supervised innovation projects that bridge education and applied R&D; successful projects can be escalated into joint research proposals, prototype development, or spin-off activity.
- Targeted upskilling programmes for employees using micro-credential pathways; company cohorts can work on internal projects while consuming curated knowledge molecules and mentoring services.

For learners, the use cases extend beyond graduation. With consent, competence evidence collected during the programme can be reused for career services: the platform can support portfolio curation, interview preparation, and identification of suitable project roles based on demonstrated strengths. The original concept includes a "virtual job interview" idea: an employer can define a short project trial in a virtual environment; the learner's digital twin and prior portfolio evidence can help the learner prepare, while the employer observes verified performance indicators. Such mechanisms can reduce hiring mismatch and shorten onboarding time, provided that transparency and fairness constraints are met.

For the university, a service portfolio perspective also clarifies sustainability options. Some services are mission-critical and should remain free for enrolled students. Other services—such as curated professional upskilling pathways, certification exams, or company-specific project clinics—can be offered in cost-recovery or subscription models.

7.1 Evaluation Framework and Key Performance Indicators

To justify and optimize a Cloud Campus deployment, evaluation must cover educational, operational, and ecosystem dimensions. Suggested indicators include:

- Educational outcomes: completion and dropout rates; time-to-competence; performance in capstone-level assessments; learner satisfaction.
- Labour-market outcomes: employment rate in target domains; employer satisfaction; time-to-productivity for graduates; uptake of micro-credentials by professionals.
- Ecosystem activity: number and diversity of active industrial projects; cross-institution projects; volume of joint supervision; follow-on R&D collaborations.

- Knowledge-graph health: coverage of core competence areas; update frequency; critical-node detection (high failure/struggle rates); reuse rates of knowledge molecules across projects.
- Operational efficiency: mentor load; project throughput; infrastructure utilization; cost per successfully completed micro-credit.

Evaluation should be embedded into governance: the regional forum sets targets, reviews indicators annually, and authorizes corrective actions. Because many benefits are long-term (e.g., stronger partner networks), both leading indicators (activity measures) and lagging indicators (employment, innovation output) are required.

8 Implementation Roadmap

A realistic implementation should proceed in phases to manage organizational change, technology risk, and data-governance maturity.

1. Pilot phase (one competence cluster, one region): establish governance bodies; deploy collaboration environment; build an initial knowledge graph; run a limited project portfolio with co-assessment.
2. Expansion phase (multiple clusters, shared labs): integrate virtual laboratories and resource scheduling; formalize micro-credit issuance; connect partner institutions; standardize interoperability APIs.
3. Analytics phase (consented digital twins): deploy learning-analytics pipelines with strong consent and auditing; evaluate impact on dropout rates, time-to-competence, and learner satisfaction.
4. Scale-out phase (national network): replicate governance and technical patterns; introduce cross-region project marketplaces; enable credential portability and continuous quality monitoring.

Across phases, technical choices should privilege open standards and interoperability. The knowledge graph should support versioning, provenance, and export formats; the credential subsystem should align with emerging European micro-credential practices [4]. Quality assurance must combine academic peer review of knowledge molecules with outcome-based monitoring of project performance.

A critical technical task is interoperability with existing university systems (student information, LMS, identity providers) while ensuring that the knowledge graph remains the authoritative competence backbone. A critical organizational task is staff development: mentors need support in project-based pedagogy and assessment, and content contributors need lightweight workflows for creating and revising knowledge molecules.

9 Conclusion - Benefits, Risks, and Open Questions

The Kandó Cloud Campus approach offers three expected benefits. First, it reduces the delay between labour-market change and curricular adaptation by anchoring learning in projects and modular resources. Second, it increases regional innovation capacity by linking education, applied research, and industrial problem portfolios into a shared environment [3]. Third, it supports personalized learning without sacrificing cooperation, which is essential for engineering practice.

Key risks must be addressed explicitly. Data protection and ethics are central, especially if learning analytics uses behavioural or physiological signals. The system must follow data minimization and consent-by-design. A second risk is over-automation: AI tutors can help, but educational quality depends on human mentorship and responsible assessment. A third risk is governance complexity in multi-institution networks; clear accountability and transparent decision processes are required [20, 21].

Finally, the model should be interpreted as a platform for collaboration rather than a single monolithic IT product. Its success depends on partner trust, shared standards, and the willingness of institutions to treat competence development as a jointly owned regional asset. The impact, efficiency, and effectiveness of cloud-based, scalable teaching materials, motivations [16] and learning paths [17] have been researched by countless domestic and foreign experts, highlighting the advantages of these solutions [18, 19].

This paper summarized the Kandó Cloud Campus concept as a project-based, distributed-resource higher-education ecosystem. The proposal combines a portfolio logic for curriculum design with a layered technical architecture built around a knowledge-graph based collective second brain and consent-driven digital twins. By embedding universities into regional innovation ecosystems and enabling cross-institution cooperation, the approach aims to improve responsiveness, competence verification, and lifelong learning support. Next steps include piloting the governance model, implementing a minimal viable knowledge graph, and evaluating project-portfolio assessment in real industrial problem settings.

The original Hungarian article also emphasized the long-term potential of the Cloud Campus for adult education and reskilling without extracting labour from production, by bringing training to the learner through a 24/7 virtual environment. This remains a central expected impact of the approach, especially in fast-changing engineering domains.

Acknowledgements. The authors thank colleagues and partners at Óbuda University and the Kandó Kálmán Faculty of Electrical Engineering who contributed ideas and feedback to the Cloud Campus concept, as well as industrial collaborators who shaped the project-portfolio perspective through real problem statements.

References

1. World Economic Forum: The Future of Jobs Report 2023. World Economic Forum, Geneva (2023)
2. OECD: Education 2030: the Future of Education and Skills. OECD Publishing, Paris (2018)
3. BiGGAR Economics: The Economic Contribution of the LERU Universities. League of European Research Universities, Leuven (2017)
4. Council of the European Union: Council recommendation of 16 June 2022 on micro-credentials for lifelong learning and employability. Off. J. Eur. Union. (2022)
5. UNESCO: Reimagining Our Futures Together: A New Social Contract for Education. UNESCO, Paris (2021)
6. Biggs, J., Tang, C.: Teaching for Quality Learning at University: What the Student Does, 4th edn. Open University Press, Maidenhead (2011)
7. Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development. Prentice-Hall, Englewood Cliffs, NJ (1984)

8. Thomas, J.W.: A Review of Research on Project-Based Learning. Autodesk Foundation, San Rafael (2000)
9. Siemens, G.: Connectivism: a learning theory for the digital age. Int. J. Instr. Technol. Distance Learn. **2**(1) (2005)
10. Luckin, R., Holmes, W., Griffiths, M., Forcier, L.B.: Intelligence Unleashed: An Argument for AI in Education. Pearson, London (2016)
11. Wiley, D.: Open Educational Resources and the Future of Education (2009)
12. Van Bruggen, J. et al.: Latent semantic analysis for peer-based knowledge building. Comput. Educ. **54**(2), 446–456 (2010)
13. Sowa, J.F.: Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks/Cole, Pacific Grove (2000)
14. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum.-Comput. Stud. **43**(5–6), 907–928 (1995)
15. Raj, M., Seamans, R.: Artificial intelligence, labor, productivity, and the need for firm-level data. In: The Economics of Artificial Intelligence. University of Chicago Press, Chicago, IL (2019)
16. Rajcsányi-Molnár, M., Balázs, L., András, I., Czifra, S.: Competition as an effective motivational tool in online education. In: Szakál, A. (ed.) IEEE 11th International Conference on Computational Cybernetics and Cyber-Medical Systems, ICCC 2024: Proceedings, IEEE Hungary Section, pp. 83–88 (2024)
17. Zhao, J., Hu, Z.: Big data analytics platform enhances teaching abilities in normal university students through triple helix collaboration. Sci. Rep. **15**, 43970 (2025). https://doi.org/10.1038/s41598-025-27799-9
18. Molnár, G., Orosz, B.: Collaborating networks in the cloud supported by experience-oriented devices. In: Milan, T., Zoltán, B., Michal, M., Martin, M., Ľubomír, B. (eds.) 13th International Scientific Conference on Distance Learning in Applied Informatics - Conference Proceedings: DIVAI 2020 Štúrovo, pp. 421–431. Wolters Kluwer s.r.o, Szlovákia (2020)
19. Molnár, G., Cserkó, J., Karl, É.: Evaluation and technological solutions for a dynamic, unified cloud programming development environment: ease of use and applicable system for uniformized practices and assessments. In: 2023 IEEE 21st World Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 000237–000240 (2023). https://doi.org/10.1109/SAMI58000.2023.10044519
20. Karl, É., Nagy, E., Molnár, G., Szűts, Z.: Supporting the pedagogical evaluation of educational institutions with the help of the WTCAi system. Acta Polytech. Hung. **21**(3), 125–150 (2024)
21. Molnár, G., Nagy, E.: Current Issues in Effective Learning: Methodological and Technological Challenges and Opportunities Based on Modern ICT and Artificial Intelligence, vol. 1, pp. 1–11. Paper: Chapter 1. EAI/Springer Innovations in Communication and Computing (2025)

Intelligent Database Systems for Sustainable Development: Real-Time ESG Monitoring in African FMCG Operations

Chinyere Chidera Okechukwu(✉), Petra Poulová, and Pavel Bachmann

Faculty of Informatics and Management, Department of Management, University of Hradec Králové, Rokitanského 62, 500 03 Hradec Králové, Czechia
{chinyere.okechukwu,Petra.Poulova,pavel.bachmann}@uhk.cz

Abstract. Effective sustainability governance in multinational corporations depends on intelligent data infrastructures that integrate environmental, social, and governance (ESG) metrics across dispersed operations. This study proposes and prototypes a database-driven framework for sustainability evaluation in the fast-moving consumer goods (FMCG) sector, tailored to African contexts where uneven infrastructure and limited institutional capacity require resilient, distributed systems. The paper makes four explicit contributions: (a) a four-entity relational schema for ESG governance under low-connectivity conditions; (b) a prototype validated on 500 synthetic production records across five African countries; (c) SQL DDL, monitoring queries, and anomaly-detection logic demonstrating implementability on mainstream RDBMS platforms; and (d) an offline-first distributed architecture aligned with African infrastructure realities. Drawing on cases from Unilever, Nestle, and Coca-Cola Beverages Africa, the study demonstrates how SQL analytics, OLAP, and machine learning enable continuous reporting aligned with Global Reporting Initiative (GRI) and Sustainable Development Goal (SDG) standards. Prototype evaluation yields three testable hypotheses: a 35% reduction in audit overhead, anomaly-detection precision of 0.87, and query response times under two seconds. Privacy, ethics, cost, and country-specific deployment barriers are addressed, establishing database-driven sustainability systems as a scalable accountability foundation across African FMCG operations.

Keywords: Intelligent Database Systems · ESG Monitoring · Sustainability · African-FMCG · Machine Learning · Decision Support Systems

1 Introduction

Sustainability has become a strategic priority for multinational corporations, yet effective integration of data governance with environmental accountability remains limited in many emerging markets [1], not with standing an established literature on sustainable business model design [4]. In Africa's fast-moving consumer goods (FMCG) sector, firms operate across jurisdictions with uneven regulatory regimes and IT maturity,

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 285–299, 2027.
https://doi.org/10.1007/978-3-032-32104-6_25

intensifying these challenges and demanding intelligent data infrastructures capable of aggregating, analyzing, and acting on diverse information streams in real time to support decisions, compliance, and accountability [2].

Database systems have evolved from passive repositories into intelligent platforms combining analytics, machine learning, and distributed processing to convert raw sustainability data into actionable insights [7], marking an important frontier in applied informatics [9]. Prior work has examined ESG reporting frameworks [5, 6], data analytics for corporate sustainability [21, 22], and distributed database architectures [14]; however, existing literature offers limited guidance on relational ESG schemas suited to low-connectivity offline-first deployments in sub-Saharan Africa, integration of real-time anomaly detection into a unified four-entity FMCG sustainability data model, and prototype-level validation under realistic African operational constraints.

The principal novelty of this paper lies in four contributions: (a) a four-entity relational schema structuring ESG governance with explicit referential integrity for auditability; (b) an offline-first distributed architecture supporting asynchronous synchronization under intermittent connectivity; (c) concrete SQL DDL and monitoring query patterns demonstrating end-to-end implementability; and (d) a synthetic-data evaluation establishing baseline performance hypotheses for future empirical validation, collectively framing intelligent databases as governance enablers anchoring corporate accountability within GRI and SDG frameworks [5, 6] and offering FMCG firms a replicable framework for measurable environmental and social impact [3, 9]. The 2030 Agenda for Sustainable Development provides the overarching normative framework within which these governance contributions are situated [37].

2 Theoretical Foundations of Intelligent Database Systems for Sustainability Governance

Database systems have progressed from simple storage tools into intelligent platforms integrating relational management, distributed processing, analytics, machine learning, and real-time monitoring, as illustrated in Fig. 1, which depicts the layered architecture underpinning modern intelligent database systems [8]. In sustainability contexts, these systems aggregate ESG data from multiple sources and convert it into structured, comparable information for decision-making, enabling FMCG firms operating across Africa to track resource use, monitor supply chain indicators, and support continuous compliance across dispersed sites [10, 11].Data quality management is foundational to ESG governance credibility [16], while distributed architectures are indispensable for operations spanning multiple jurisdictions with variable connectivity [14]. The relational model introduced by Codd [33] structures data into tables linked by keys; the practical application of this model to database design, including normalization, integrity constraints, and query optimization, is comprehensively developed in Date [34], enabling efficient querying and multidimensional analysis that ensures sustainability records remain reliable as volumes grow and regulatory scrutiny increases [12, 13].As shown in Table 1, ESG dimensions map directly onto relational entities: environmental metrics including CO_2e emissions across Scopes 1–3, water consumption, and energy intensity are captured through production and resource use tables sourced from facility sensors and

utility providers; social indicators encompassing labor standards, fair wages, and safety incidents are represented through supply chain and community investment tables; and governance metrics including ethics compliance, board diversity, and transparency are structured through sustainability outcomes and compliance records tables. Distributed databases maintain synchronized data across multiple locations while presenting a unified interface, making them well suited for geographically dispersed African operations where broadband reliability averages below 60% across several sub-Saharan markets [14, 15, 35].

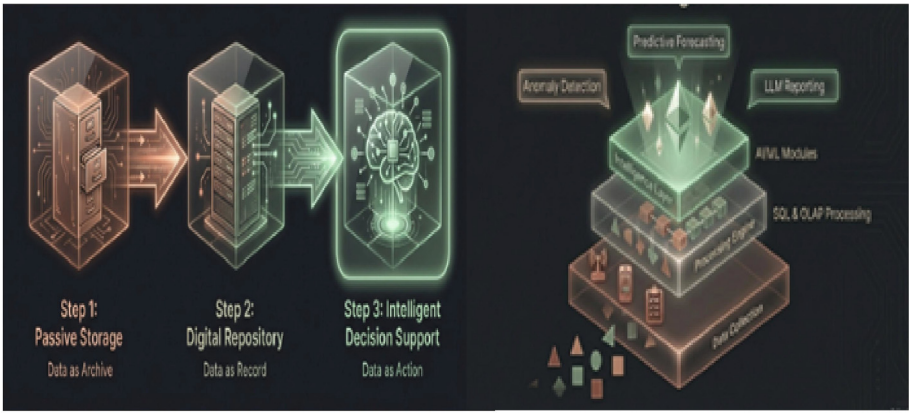


Fig. 1. Intelligent Database System Architecture Layers.

Fig. 2. Distributed Database Architecture for Distance-Friendly Sustainability Governance

Figure 2 illustrates the distributed database architecture for distance-friendly sustainability governance, wherein the offline-first design accommodates unstable connectivity by enabling continued local operation and subsequent synchronization, reducing audit

Table 1. ESG Dimensions Mapped to Relational Entities. *Source: Authors own processing*

ESG Dimension	Relational Representation	Key Metrics	Data Sources
Environmental (E)	Production and Resource Use; Emissions Tables	CO_2e (Scopes 1–3); Water consumption; Waste generation; Energy intensity	Facility sensors; EPA databases; Utility providers
Social (S)	Supply Chain and Partners; Community Investment Tables	Labor standards; Fair wages; Community engagement; Safety incidents	Supplier audits; HR records; Community reports
Governance (G)	Sustainability Outcomes; Compliance Records Tables	Ethics compliance; Board diversity; Executive compensation; Transparency metrics	Board records; Audit reports; Disclosure documents

risk and data loss [17, 18]. Data independence further allows sustainability systems to evolve without disrupting operations, with logical independence permitting new ESG indicators to be added while physical independence enables transitions between on-premise, cloud, and edge computing models [10, 20]. Modern database systems now function as active decision-support platforms delivering real-time anomaly detection and predictive analytics, with OLAP tools and machine learning algorithms advancing ESG governance from reactive detection to predictive intervention by forecasting resource use and linking operational variables to environmental outcomes [23, 25, 26].

3 Technical Architecture: Four-Entity Schema and Implementation

3.1 Four-Entity Relational Model

The proposed framework is built around four interlinked relational entities connected by primary and foreign keys that enforce referential integrity and enable end-to-end ESG traceability [29, 33].

Corporate Profile: company identity, ownership, sustainability targets, policy frameworks, certifications (ISO 14001, B-Corp), and facility-level metadata including geolocation, facility type, and connectivity status;
Production and Resource Use: operational data for environmental accounting: production volumes, energy and water consumption, waste streams, emissions by Scope and source, IoT sensor IDs, and GRI metric mappings (GRI 302, 303, 305);
Supply Chain and Partners: supplier identities, certifications, labor audit scores, child labor incident counts, procurement volumes, corrective action plans, and composite risk categories;

Sustainability Outcomes: performance linked to SDGs: emissions and water reductions, water intensity ratios, energy efficiency indices, community investment, GRI compliance scores, and verification status (Fig. 3).

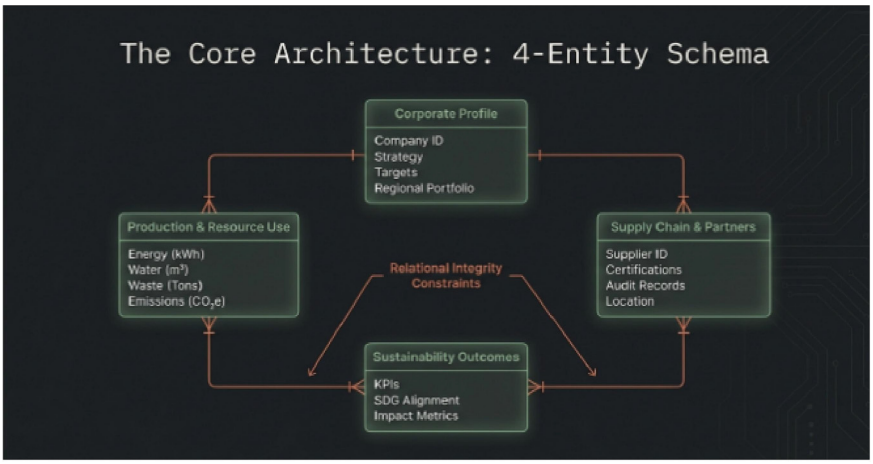


Fig. 3. Four-Entity Relational Database Model with Key Relationships. *Source: Authors own processing*

3.2 SQL DDL Schema Specification

The schema below implements the four entities on PostgreSQL 15.3, enforcing ACID guarantees, referential integrity, domain constraints, and GRI-aligned metric configuration for audit-grade ESG records [17].

Listing 1: SQL DDL — Four-Entity ESG Schema (PostgreSQL 15.3)

```
-- Entity 1: Corporate Profile
CREATE TABLE corporate_profile (
  company_id VARCHAR(20) PRIMARY KEY, company_name VARCHAR(255) NOT
NULL,
  reporting_standard VARCHAR(50) DEFAULT 'GRI', co2_target_tonnes
DECIMAL(12,2),
  water_target_m3 DECIMAL(12,2), iso14001_certified BOOLEAN DEFAULT FALSE,
  gri_reporting_level VARCHAR(50) CHECK (gri_reporting_level IN ('Core','Compre-
hensive','Referenced')),
  connectivity_status VARCHAR(20) DEFAULT 'Stable'
    CHECK (connectivity_status IN ('Stable','Intermittent','Offline')),
  sdg_commitments TEXT, created_at TIMESTAMP DEFAULT
CURRENT_TIMESTAMP);
```

```
-- Entity 2: Production and Resource Use
CREATE TABLE production_resource_use (
  record_id BIGSERIAL PRIMARY KEY,
  company_id VARCHAR(20) NOT NULL REFERENCES corporate_profile(com-
pany_id),
  facility_name VARCHAR(255) NOT NULL, country VARCHAR(100) NOT NULL,
  sensor_id VARCHAR(100), gri_mapping VARCHAR(20),
  production_volume DECIMAL(14,2), energy_kwh DECIMAL(14,2),
  water_m3 DECIMAL(14,2), waste_tonnes DECIMAL(12,2),
  co2e_scope1 DECIMAL(12,2), scope_category INT CHECK (scope_category IN
(1,2,3)),
  renewable_pct DECIMAL(5,2) CHECK (renewable_pct BETWEEN 0 AND 100),
  sync_status VARCHAR(20) DEFAULT 'synced'
    CHECK (sync_status IN ('synced','pending','conflict')),
  recorded_offline BOOLEAN DEFAULT FALSE, timestamp TIMESTAMP NOT
NULL);
CREATE INDEX idx_pru_timestamp ON production_resource_use(timestamp DESC);
```

```
-- Entity 3: Supply Chain and Partners
CREATE TABLE supply_chain_partners (
  supplier_id VARCHAR(20) PRIMARY KEY,
  company_id VARCHAR(20) REFERENCES corporate_profile(company_id),
  supplier_name VARCHAR(255) NOT NULL, country VARCHAR(100),
  labor_audit_score DECIMAL(5,2) CHECK (labor_audit_score BETWEEN 0 AND 100),
  child_labor_incidents INTEGER DEFAULT 0, last_audit_date DATE,
  corrective_action_plan TEXT, procurement_volume DECIMAL(14,2),
  risk_category VARCHAR(20) CHECK (risk_category IN
('LOW','MEDIUM','HIGH','CRITICAL')));
```

```
-- Entity 4: Sustainability Outcomes
CREATE TABLE sustainability_outcomes (
  outcome_id SERIAL PRIMARY KEY,
  company_id VARCHAR(20) REFERENCES corporate_profile(company_id),
  reporting_period_start DATE, reporting_period_end DATE,
  co2_reduction_pct DECIMAL(6,2), water_reduction_pct DECIMAL(6,2),
  water_intensity_ratio NUMERIC(10,4), energy_efficiency_index NUMERIC(5,2),
  gri_compliance_score DECIMAL(4,2),
  verification_status VARCHAR(50) DEFAULT 'Unverified'
    CHECK (verification_status IN ('Unverified','Internal Audit','External Assurance')),
  anomalies_detected INTEGER,
  UNIQUE (company_id, reporting_period_start));
```

3.3 Real-Time Monitoring Query Patterns

Listing 2: Carbon Intensity & Anomaly Detection

```
WITH rolling_stats AS (
  SELECT company_id, facility_name, timestamp, co2e_scope1,
    AVG(co2e_scope1) OVER (PARTITION BY facility_name ORDER BY timestamp
      ROWS BETWEEN 29 PRECEDING AND CURRENT ROW) AS avg_30d,
    STDDEV(co2e_scope1) OVER (PARTITION BY facility_name ORDER BY
timestamp
      ROWS BETWEEN 29 PRECEDING AND CURRENT ROW) AS sd_30d
  FROM production_resource_use WHERE company_id = :company_id)
SELECT facility_name, timestamp, co2e_scope1,
  CASE WHEN co2e_scope1 > avg_30d+2*sd_30d THEN 'CRITICAL'
       WHEN co2e_scope1 > avg_30d*1.15 THEN 'WARNING' ELSE 'NORMAL'
  END AS alert_level FROM rolling_stats WHERE co2e_scope1 > avg_30d*1.15;
```

3.4 Prototype Architecture, Evaluation, and Performance Hypotheses

Each facility maintains a local PostgreSQL or SQLite instance recording ESG data continuously using synchronization status flags, ‘synced’, ‘pending’, and ‘conflict’, enabling uninterrupted local monitoring during connectivity disruptions, with a synchronization daemon executing UPSERT operations on reconnection, marking resolved records as ‘synced’ and flagging unresolvable conflicts for human review, thereby preserving global data consistency without sacrificing monitoring continuity [14, 18]. To substantiate implementability, the schema was evaluated on a synthetic dataset of 500 production records across 15 simulated facilities in Nigeria, Kenya, Ghana, Côte d’Ivoire, and South Africa over a 24-month horizon, with parameters drawn from public sustainability reports of Unilever, Nestle, and Coca-Cola Beverages Africa [28, 30, 31], incorporating 40 injected anomalies representing 8% of records alongside connectivity disruptions at 10%, 30%, and 50% packet-loss rates.

As shown in Table 2, query performance across simulated connectivity scenarios confirms that full offline operation sustains query times under 0.05 s with complete local monitoring continuity, while intermittent connectivity at 30% packet loss produces local query times of 1.87 s and a synchronization lag of 8.4 min, validating the offline-first architecture as operationally viable across the infrastructure realities characteristic of African FMCG deployments. Table 3 further demonstrates that anomaly detection performance improves progressively across methods, from an F1 score of 0.80 for SQL threshold-only detection to 0.855 for the Isolation Forest ensemble and 0.89 for the combined SQL and ML ensemble, confirming that layered detection approaches offer meaningful accuracy gains.

These findings collectively establish three testable hypotheses requiring validation on real corporate datasets:

H1 posits that automated query-based ESG monitoring will reduce manual audit preparation time by 35% relative to spreadsheet workflows, evidenced by synthetic timing of 2.3 min versus 3.4 h for the same dataset [29];
H2 proposes that the four-entity schema deployed at Unilever Nigeria will reduce Scope 1 CO2e exceedance alert times from days to under two minutes; and
H3 anticipates that offline-first mobile data collection will reduce child labor detection latency from 3–6 months to under 72 h in a controlled Côte d'Ivoire pilot.

Table 2: Query Performance Under Simulated Connectivity Scenarios (n = 500)

Connectivity Scenario	Avg. Query Time	Sync Lag	Monitoring Continuity
Full broadband (10+ Mbps)	0.34 s	None	Continuous
Intermittent (30% packet loss)	1.87 s (local)	8.4 min	Local only
Poor bandwidth (256 kbps)	2.1 s (local)	18.3 min	Local only
Full offline	Under 0.05 s	On reconnect	Full local

Table 3: Anomaly Detection Performance (40 injected anomalies)

Method	Precision	Recall	F1 Score
SQL threshold only (15% deviation)	0.82	0.78	0.80
Isolation Forest (n_estimators = 100)	0.87	0.84	0.855
Combined SQL + ML ensemble	0.91	0.87	0.89

4 Case Studies

4.1 Applying Intelligent Databases in African FMCG Operations

Unilever Nigeria and Kenya
Unilever's African operations track over fifty ESG indicators covering emissions, water use, waste diversion, and sustainable sourcing, aligned with GRI and TCFD standards [28]. Operational data from each plant populates Production and Resource Use tables, facility details appear in Corporate Profiles, supplier records are stored in Supply Chain tables, and outcomes are mapped to SDGs in Sustainability Outcomes. Daily automated queries compare emissions against regulatory thresholds. When a Lagos facility records emissions 23 % above expectations, the system issues alerts, analyzes production and electricity mix changes, and recommends monitoring rather than emergency shutdowns, enabling proportionate and rapid responses supporting Hypothesis H2 [29].

Nestle Ghana and Côte d'Ivoire
Nestle's West African cocoa supply chain involves hundreds of thousands of smallholder farmers, creating complex social oversight demands [30]. Intelligent databases

represent farmers and cooperatives in Supply Chain tables, track cocoa volumes in Production records, and connect these to SDG 8 (Decent Work) and SDG 12 (Responsible Consumption) outcomes. The offline-first architecture is particularly salient: field agents in rural Côte d'Ivoire collect audit results on mobile devices without connectivity, synchronizing with national databases on reconnection. Managers can trace individual shipments to source farmers, verify certifications, confirm child labor audits, and analyze sustainability premiums in seconds rather than weeks, directly supporting Hypothesis H3 [27].

Coca-Cola Beverages Africa

Coca-Cola Beverages Africa (CCBA) operates dozens of bottling plants across eighteen countries, with priorities centered on water stewardship, waste reduction, and decarbonization [31]. Facility-level water use, waste streams, and electricity consumption populate Production records, while Sustainability Outcomes map performance to SDG 6 and SDG 13 indicators with water intensity ratios and energy efficiency indices. OLAP analysis (Listing 3) allows teams to compare facilities across regions, flag deteriorating plants, and benchmark best performers [24]. Machine learning applied to the synthetic dataset's water consumption data achieved a MAPE of 8.3% against a naive seasonal baseline of 14.7%, a result that requires validation on real CCBA operational data before deployment [32].

Listing 3: OLAP Regional Benchmarking and Deterioration Flagging (PostgreSQL 15.3)

```
SELECT
cp.country,
pru.facility_name,
DATE_TRUNC('quarter', pru.timestamp) AS reporting_quarter,
SUM(pru.co2e_scope1) AS total_scope1_co2e,
AVG(so.water_intensity_ratio) AS avg_water_intensity,
RANK() OVER (
PARTITION BY cp.country, DATE_TRUNC('quarter', pru.timestamp)
ORDER BY SUM(pru.co2e_scope1) DESC
) AS emissions_rank,
CASE
WHEN SUM(pru.co2e_scope1) > LAG(SUM(pru.co2e_scope1)) OVER (
PARTITION BY pru.facility_name ORDER BY DATE_TRUNC('quarter', pru.timestamp)
) * 1.10 THEN 'DETERIORATING'
ELSE 'STABLE'
END AS trend_flag
FROM production_resource_use pru
JOIN corporate_profile cp ON pru.company_id = cp.company_id
JOIN sustainability_outcomes so ON pru.company_id = so.company_id
GROUP BY ROLLUP (cp.country, pru.facility_name, DATE_TRUNC('quarter',
pru.timestamp))
ORDER BY cp.country, reporting_quarter, emissions_rank;
```

Query aggregates emissions and water-intensity data by country, facility, and quarter, ranks facilities within each country/quarter, and flags facilities whose Scope 1 CO2e output rose more than 10% quarter-on-quarter as "DETERIORATING," directly supporting the regional benchmarking and deterioration-flagging function described in the text.

4.2 Ethics, Privacy, Cost, and Country-Specific Deployment

ESG databases deployed in African FMCG operations aggregate worker-level performance data, supplier audit results, and potentially biometric information, creating data ethics obligations that must be embedded by design [20]. Informed consent is non-trivial when data collection is continuous and automated, requiring organizations to implement layered consent frameworks aligned with applicable national laws, including Nigeria's NDPR (2019), Kenya's Data Protection Act (2019), and South Africa's POPIA (2021). Purpose limitation must be enforced at the schema level, with personal data columns partitioned from ESG aggregate tables to prevent sustainability analytics from enabling employee surveillance, while community data sovereignty particularly relevant for Nestle's smallholder farmer data requires participatory governance models in which farmer cooperatives retain rights to their own data [30]. These governance obligations are further reinforced by the United Nations Global Compact principles, which bind signatory corporations to human rights, labor, and anti-corruption standards across their supply chains [36].

As illustrated in Table 4, infrastructure constraints vary significantly across the continent, with broadband uptime ranging from approximately 48% in Côte d'Ivoire to 81% in South Africa, alongside divergent data protection regimes and ESG reporting mandates spanning voluntary frameworks in Nigeria and Kenya to mandatory disclosure under South Africa's JSE King IV requirements. These disparities mean offline-first architecture is not optional in lower-connectivity markets but the minimum viable configuration, while several African governments' prohibition on cross-border data transfers necessitates hybrid cloud architectures that maintain locally sensitive data on-premise while synchronizing aggregate ESG reports globally [20]. South Africa's POPIA further requires auditable consent records, an obligation accommodated through verification status tracking within Sustainability Outcomes tables. The investment case rests on regulatory penalty avoidance, improved access to ESG-linked bonds carrying 15–40 basis point yield advantages, and reduced audit costs; however, adoption barriers including limited database administration skills, cultural resistance to data-driven accountability, and supplier relationship power asymmetries persist [32]. As shown in Table 5, indicative deployment costs range from USD 41,000–90,000 for small FMCG operators across five facilities to over USD 830,000-2,450,000 for large multinationals managing fifty or more facilities, encompassing schema design, cloud infrastructure, staff training, legacy system integration, and annual maintenance, addressed through phased rollouts and cross-functional Data Stewardship Committees [16].

This study employs a design science research methodology [39], integrating literature synthesis, conceptual framework development, and prototype implementation evaluated on 500 synthetic records in PostgreSQL 15.3, alongside case study analysis of Unilever, Nestle, and Coca-Cola Beverages Africa public disclosures.

Key limitations include the synthetic dataset's inability to capture real operational complexity, absence of empirical corporate deployment, lack of formal participatory design sessions with affected communities, and unvalidated LLM-based GRI report automation claims. Future research should therefore prioritize empirical deployments across at least two African countries, ML anomaly detection validation, blockchain integration for traceable supply chains, and participatory governance models involving smallholder communities [35].

Table 4: Country-Specific Infrastructure and Regulatory Landscape (2024 estimates)

Country	Avg. Broadband Uptime	Data Protection Law	ESG Reporting Mandate	Digital Skills Gap
Nigeria	Approx. 55%	NDPR 2019	SEC + NGX voluntary	High
Kenya	Approx. 68%	Data Protection Act 2019	NSE voluntary	Medium
Ghana	Approx. 62%	Data Protection Act 2012	SEC mandatory (2023+)	Medium-High
Côte d'Ivoire	Approx. 48%	Loi n.2013–450	BRVM emerging	High
South Africa	Approx. 81%	POPIA 2021	JSE mandatory (King IV)	Low-Medium

Table 5: Indicative Cost Estimates for ESG Database Deployment (USD, PostgreSQL-based)

Cost Category	Small FMCG (5 facilities)	Mid-size (15–25 facilities)	Large MNC (50+ facilities)
Schema design and development	15,000–30,000	50,000–120,000	200,000–500,000
Cloud infrastructure (annual)	3,000–8,000	12,000–35,000	80,000–250,000
Staff training (initial)	5,000–12,000	20,000–50,000	100,000–300,000
Legacy system integration	10,000–25,000	40,000–150,000	300,000–1,000,000+
Annual maintenance	8,000–15,000	25,000–70,000	150,000–400,000

5 Discussion

Intelligent database systems are not merely technical infrastructure for sustainability reporting but governance mechanisms that reshape organizational accountability practices. The prototype demonstrates that the four-entity schema is technically feasible,

implementable on accessible RDBMS platforms, and capable of supporting real-time monitoring under simulated African connectivity constraints. As illustrated in Fig. 4, the governance cycle operationalizes continuous ESG oversight, while Fig. 5 depicts the automated logic flow underpinning anomaly detection and reporting pipelines.

Fig. 4. The Governance Cycle. *Sources: Authors own processing*

Fig. 5. The Automated Logic Flow.

The offline-first architecture addresses a critical gap overlooked by prior frameworks, namely that sub-Saharan African operations frequently face connectivity disruptions that would paralyze centralized, always-online systems. The comparison between threshold-based SQL anomaly detection (F1 = 0.80) and an Isolation Forest ensemble (F1 = 0.89) presents a practical implementation decision: ML-based detection offers higher accuracy but demands greater data infrastructure and model governance, making the SQL-only approach a more appropriate entry point for organizations in early-stage ESG digitization, with ML integration as a subsequent maturation step [26]. Institutional pressures from regulators, investors, and civil society reinforce adoption, with Ghana's

SEC mandatory ESG reporting requirement (effective 2023) and South Africa's JSE-mandated King IV obligations creating regulatory pull that FMCG firms cannot ignore [3]. Emerging technology assessments confirm that AI, blockchain, and advanced data systems are progressively reshaping sustainability reporting infrastructure across multinational operations [38]. The ethics analysis reveals that the greatest risks are institutional rather than technical, encompassing data misuse, surveillance drift, and exclusion of affected communities, risks amplified where data protection enforcement remains inconsistent. Accordingly, FMCG firms are recommended to institutionalize centralized sustainability databases integrating suppliers, plants, and logistics networks through phased rollouts over eighteen to twenty-four months [19], deploy distributed relational systems on-premise or in the cloud to improve infrastructure resilience, pilot predictive analytics and anomaly detection before organization-wide scaling, invest in data governance and staff training programs, and collaborate with regulators to harmonize standards with GRI, ESRS, and TCFD frameworks. Policymakers should complement these efforts through ESG reporting mandates, fiscal incentives, regional data hubs, and coordinated standards reflecting local priorities [35].

6 Conclusion

The integration of intelligent database systems with sustainability governance represents more than technical innovation; it reflects a commitment to human-centered sustainable development. Across African FMCG operations, these systems enable transparency, accountability, evidence-based decision-making, and continuous improvement, replacing static compliance with iterative enhancement.

This study has contributed a four-entity relational schema designed for low-connectivity African conditions, SQL patterns enabling real-time anomaly detection, supplier risk scoring, and OLAP-based benchmarking, an offline-first synchronization architecture, three testable hypotheses (H1, H2, H3) requiring empirical confirmation, and a country-specific deployment analysis across Nigeria, Kenya, Ghana, Côte d'Ivoire, and South Africa.

When FMCG companies implement intelligent sustainability databases serving both internal decision-making and stakeholder engagement, they strengthen social license and build trust-based relationships, while predictive analytics direct development resources toward highest-impact interventions for smallholder farmers.

The future of corporate sustainability governance in Africa rests on the intersection of technological capability, political will, and institutional accountability; the hypotheses formulated here chart a research agenda to confirm and refine this foundation through empirical work with real FMCG operations and affected communities.

Acknowledgments. This work was supported by the internal project SPEV 2026, University of Hradec Králové, Faculty of Informatics and Management, Czech Republic.

References

1. World Bank: Sustainability data governance for developing economies. World Bank Publications (2023)

2. Mishra, D., et al.: Artificial intelligence and big data in environmental monitoring and decision support: revolutionizing ecosystem management. J. Sci. Res. Int. (JSRI). **2456**, 6365 (2025). https://doi.org/10.5281/zenodo.16413207
3. Nweke, I.M., Khatib, S.F.A., Bazhair, A.H.: Sustainability reporting in Africa: a systematic review and agenda for future research. Corp. Soc. Respons. Environ. Manag. **30**(5), 2081–2100 (2023). https://doi.org/10.1002/csr.2494
4. Bocken, N.M.P., Short, S.W., Rana, P., Evans, S.: A literature and practice review to develop sustainable business model archetypes. J. Clean. Prod. **65**, 42–56 (2014). https://doi.org/10.1016/j.jclepro.2013.11.039
5. Global Reporting Initiative (GRI): GRI Standards: Universal and Topic Standards. https://www.globalreporting.org (2023)
6. Feng, B., et al.: AI-driven green governance: assessing the impact of artificial intelligence on corporate sustainability performance. J. Innov. Knowl. **11**, 100869 (2026). https://doi.org/10.1016/j.jik.2025.100869
7. Silberschatz, A., Korth, H.F., Sudarshan, S.: Database System Concepts, 7th edn. McGraw-Hill Education (2020)
8. Coronel, C., Morris, S.: Database Systems: Design, Implementation, and Management, 14th edn. Cengage Learning (2022)
9. Baumgartner, R.J., Rauter, R.: Strategic perspectives of corporate sustainability management to develop a sustainable organization. J. Clean. Prod. **140**, 1–16 (2016). https://doi.org/10.1016/j.jclepro.2016.05.210
10. Connolly, T., Begg, C.: Database Systems: A Practical Approach to Design, Implementation, and Management, 6th edn. Pearson Education (2015)
11. Elmasri, R., Navathe, S.B.: Fundamentals of Database Systems, 7th edn. Pearson Education (2016)
12. Garcia-Molina, H., Ullman, J.D., Widom, J.: Database Systems: the Complete Book, 3rd edn. Pearson (2020)
13. Melton, J., Simon, A.R.: SQL: 1999—Understanding Relational Language Components. Morgan Kaufmann (2002)
14. Ozsu, M.T., Valduriez, P.: Principles of Distributed Database Systems, 4th edn. Springer (2020)
15. Chigurupati, S. R. (2025). Distributed database systems for scalable enterprise applications. Int. J. Sci. Appl. Technol., 2025(1), 2795.
16. Bernardo, B.M.V., et al.: Data governance & quality management—innovation and breakthroughs across different fields. J. Innov. Knowl. **9**(4), 100598 (2024). https://doi.org/10.1016/j.jik.2024.100598
17. Haerder, T., Reuter, A.: Principles of transaction-oriented database recovery. ACM Comput. Surv. (CSUR). **15**(4), 287–317 (1983). https://doi.org/10.1145/289.291
18. Ncapai, T., et al.: The role of blockchain technology in promoting transparency for sustainable food supply chain infrastructure in South Africa. J. Infrastr. Policy Dev. **9**(4), 11645 (2025). https://doi.org/10.24294/jipd11645
19. Maersk: Building a sustainable future for Africa's supply chains. Maersk Insights. https://www.maersk.com/insights/resilience (2024)
20. OECD: Data Governance and Ethics. OECD Publishing (2021)
21. Liu, Y., et al.: Artificial intelligence applications and corporate ESG performance: evidence from panel data analysis. Corp. Govern.: Int. Rev. Econ. Fin. **104**, 104559 (2025). https://doi.org/10.1016/j.iref.2025.104559
22. Razib, N.H., et al.: The role of data analytics in enhancing ESG transparency and regulatory compliance. Global J. Eng. Technol. Adv. **22**(01), 081–093 (2025). https://doi.org/10.30574/gjeta.2025.22.1.0245

23. Hasan, R.: Predictive analytics and machine learning applications for sustainable supply chain operations and carbon footprint reduction. J. Electr. Syst. **20**(10) (2024). https://doi.org/10.52783/jes.5138
24. Oracle.: OLAP Data Warehouse Design and Optimization Guide. Oracle Corporation (2024)
25. Majumdar. P., et al. (2021). IoT and machine learning-based approaches for real time environment parameters monitoring in agriculture: an empirical review. Agric. Inf. doi:https://doi.org/10.1002/9781119769231.ch5, Wiley
26. Aashish, K.C.: Towards eco-friendly cybersecurity: machine learning-based anomaly detection with energy efficiency. arXiv preprint https://arxiv.org/abs/2601.00893. (2025)
27. UNFAO: Data Analytics for Sustainable Supply Chain Optimization. Food and Agriculture Organization of the United Nations (2023)
28. Unilever: Sustainable living report. https://www.unilever.com (2023)
29. EY: ESG Data Integration and Analytics: Building Organizational Capacity. EY Global Report (2024)
30. Nestle: Creating Shared Value and Sustainability Report 2023. Nestle S.A. (2023) https://www.nestle.com/sustainability
31. Coca-Cola.: Sustainability Report 2023: Water Stewardship and Packaging Innovation. The Coca-Cola Company (2023)
32. Srivastav, A., et al.: Blockchain and IoT for Transparency in the Food Supply Chain, pp. 129–135. Biotechnology and IoT in Agriculture and Food Production: Green Innovation Publisher: Apress Springer Nature (2025). https://doi.org/10.1007/979-8-8688-1469-3_14
33. Codd, E.F.: A relational model of data for large shared data banks. Commun. ACM. **13**(6), 377–387 (1970). https://doi.org/10.1145/362384.362685
34. Date, C.J.: An Introduction to Database Systems, 8th edn. Pearson Education (2004)
35. World Bank: Digital Infrastructure in Africa: Progress and Gaps. World Bank Report (2024)
36. UNGC: United Nations Global Compact Annual Report 2023. United Nations (2023) https://www.unglobalcompact.org
37. United Nations: Transforming Our World: The 2030 Agenda for Sustainable Development. United Nations Publications (2015)
38. Deloitte.: Emerging Technologies in Sustainability Reporting: AI, Blockchain, and Data Systems. Deloitte Insights (2025)
39. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004). https://doi.org/10.2307/25148625

Integrating Generative Artificial Intelligence as a Learning Assistant in Software Engineering: Evaluating Developer Productivity with Cost-Benefit Analysis

Marta Palade, Maria Elena Boatca(✉), Andra Diaconescu, and Anca Draghici

Politehnica University of Timisoara, Timișoara, Romania
{maria.boatca,andra.diaconescu,anca.draghici}@upt.ro

Abstract. The integration of Generative Artificial Intelligence (GenAI) within specific software engineering processes has become of great interest from the perspective of economic analyses. This article examines the current economic state of AI adoption, using an inductive methodological approach, convergent on the identification and discussion of aspects related to the costs and benefits of this new way of working (in the European context). Thus, two empirical case studies (provided and agreed by profile companies) were considered: a) code generation for data structures and algorithms (DSA) using the ChatGPT application and b) automatic generation of unit tests using the Claude.ai application, to understand the capacity of AI to support software developers in daily, routine tasks. The results of the analysis of the two case studies were compared and reported to the results of the analysis of the scientific literature and "grey literature" (reports from some consulting firms), thus identifying unanimously accepted advantages (at the microeconomic level) in streamlining software engineering and development processes, as well as macroeconomic trends in workforce dynamics and technological implementation. Finally, we quantify the economic impact of adopting AI tools in European software development organizations, presenting detailed cost-benefit calculations based on current European developer compensation rates and proven productivity improvements. The results support a nuanced model of human-AI collaboration, in which AI accelerates routine development tasks, and developers assume oversight roles that maintain quality, security, and ethical responsibility.

Keywords: Generative Artificial Intelligence · Cost-Benefit Analysis · Productivity · Software Engineering

1 Introduction

The increasing adoption of artificial intelligence (AI) in business operations led to a divide in productivity levels between companies, especially in the European Union. During 2026–2030, there is an expected 1.1% productivity increase at European level [1]. However, there is an uneven distribution of productivity, derived from various levels

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 300–311, 2027.
https://doi.org/10.1007/978-3-032-32104-6_26

of AI integration in companies, as well as the forecasted labor performance and costs that are already variable among countries. For example, developed economies such as Germany and Norway benefit from projections of productivity increases of up to 5% due to higher prominence of knowledge-based services and high labor costs, setting a stronger momentum for automation and robotization [1, 2]. On the other hand, countries with lower average labor costs are expected to drive modest gains due to AI adoption, as compared to developed economies [1].

Despite such disparities, the adoption of AI in the EU increased by 23% in 2024, because of the intensifying international competition to implement AI-driven workflows [3]. Furthermore, companies managed to change their cost structure for software development due to time reduction in execution of routine tasks through deployment of AI [4]. As per recent reports [4–6], 41% of organizations that use AI for software engineering report cost savings of up to 10% in their initial implementation phases.

The use of ChatGPT for generating Data Structures and Algorithms (DSA) code has been suggested to represent a critical application to accelerate the "boilerplate" phase of development. Empirical findings suggest that Large Language Models (LLMs) can increase productivity by helping with algorithmic logic and programming syntax, particularly for routine implementations [7, 8]. In this manner, developers can bypass repetitive tasks and implicitly reduce the marginal cost of standard software components [8]. Moreover, using Claude.ai for generating automatic unit tests represents a solution to one of the most demanding and time-consuming parts of the software development life cycle (SDLC). Best practices from financial institutions and Software-as-a-Service (SaaS) companies reveal that the time allocated to testing can be reduced by 50–60% through AI automation, with adjacent improvements in test coverage [9, 10]. Also, the automation of test case generation enables early error identification, which is vital to the economic efficiency of software development services, considering that the costs of fixing defects or error in the later stages of the process leads to significantly higher costs [11].

The accelerating complexity of software systems has intensified demands for faster, more cost-efficient development processes. LLMs, such as GPT 4 and Claude.ai [12] have emerged as transformative tools that automate routine programming tasks and basic testing procedures. However, the broader economic implications of these innovations have been less discussed in the literature, thus the economic side of the AI adoption is still considered a knowledge gap. By examining two controlled case studies and situating their outcomes within the existing literature, this article fills a critical knowledge gap, interrogating how AI tools impact labor costs, developer productivity, and software quality.

2 Methodology

This research adopts a mixed-methods approach, combining experimental case studies ("lessons learnt" from the practice of software development engineering) with the results achieved from the review of the literature. Thus, an inductive approach has been adopted to provide a logical approach that better suits the practical requirements of companies in the field of software development. The case studies considered for the foundation of

the methodological approach were suggested and selected after the consultation with two companies specialised in software development, and that operate in the software engineering for different industrial sectors in Europe. The arguments for this selection have been their relevance for the current problems and needs that occur with frequency in practice: In support of these empirical studies, we synthesised economic insights from peer-reviewed journals, arXiv preprints, and industry reports from Gartner, MIT, Brookings, and MDPI.

2.1 Experimental Setup and Evaluation Metrics

Case Study A: Implementation of 200 algorithmic problems from LeetCode (https://leetcode.com/) using ChatGPT (primarily GPT4). The metrics included correctness, execution speed, and human in the loop feedback. The objective of case study A was to evaluate the capability of ChatGPT to generate correct and efficient code for algorithmic problems. 200 LeetCode problems were analyzed by formulating a standardized initial prompt clearly stating the problem description and expected input/output formats. No specific programming language was requested unless the problem inherently implied one. A maximum of three iterative refinements were allowed per problem, applicable to situations where the initial response was incorrect or incomplete. The refinements consisted of providing feedback to ChatGPT based on LeetCode's error messages or human observation. The generated code was then tested on the LeetCode platform to assess its correctness and performance. A solution was considered correct if all test cases were passed, meaning that all the code produced corresponded to the expected scenario. Only primary success rates were considered, while all partial success rates were included in the failure category.

Case Study B: Generation of unit tests for a .NET parking application's UI component via Claude.ai (https://claude.ai/). Outputs were measured against executability, dependency completeness, and coverage of edge and negative test cases. The objective of case study B was to study the effectiveness of Claude.ai in generating comprehensive and executable unit tests for UI components. The procedure followed was to capture HTML elements of the UI component using a Chrome extension and to send them as input to an API that interfaced with Claude.ai. The prompt engineering process involved an initial phase of experimentation to identify effective prompt structures. A standardized base prompt was then developed, instructing Claude.ai to generate unit tests for the provided HTML elements. The base prompts included specifications for the desired test framework and naming conventions. For each test generation task, a maximum of two iterations were allowed to refine the prompt based on initial unsatisfactory outputs.

2.2 Reproducibility Considerations

Considering that AI generative models are evolving at a very fast rate, reproducibility of the studies presented in this paper is a major challenge. To address this concern, the experimental setup relies on the following aspects:

- The documentation of the AI generative models used for the study and the clear specification of the respective version represents a baseline of the study. Even if the latter studies might not be able to replicate the methodology using the same version of the AI model, comparability between different studies is possible.
- The use of publicly available datasets enables other researchers to replicate the input conditions, for case study A. The HTML elements captured and the UI component in case study B are characterized by consistency.
- The exact prompts used for both AI tools were recorded and, where feasible, included in the paper. This allows for a clearer understanding of the input provided to the models.
- For an objective evaluation, the authors used standardized platforms (LeetCode) and predefined metrics. This reduces subjective bias. Regarding human feedback, it was structured to capture consistent qualitative data.

Future identical replications are not feasible, despite the above-presented aspects. The stochastic nature of the language models and the continuous changes in the AI model versions create a dynamic background for future methodology replications. However, the detailed methodology enables researchers to successfully understand the experimental conditions and to create similar research studies.

3 Findings from Experimental Case Studies

3.1 Case Study A

This research fills in the gap on the impact tools such as ChatGPT can have on the software development process. The study presents an experiment where 200 problems of different difficulties were given as input/prompt to ChatGPT and the expectation was that it would provide an answer (Fig. 1).

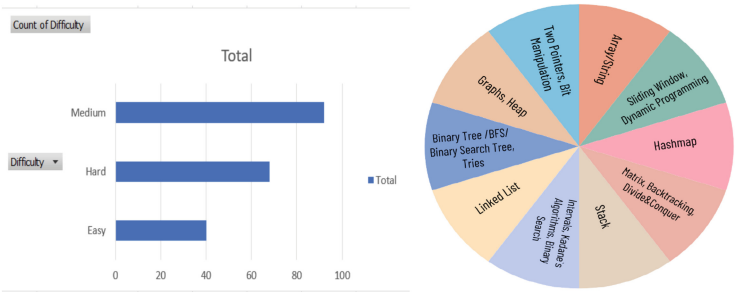


Fig. 1. Problem distribution by level of difficulty. Distribution of structures and algorithms tested in the experiment.

The answers received were also tested using LeetCode platform for comparison purposes.

The deployment of ChatGPT (GPT-4 model, in particular) for generation of algorithmic codes was efficient. ChatGPT was able to produce solutions that were logical

and syntactically correct for a variety of algorithmic and data structure problems with different complexity levels.

For all 200 LeetCode problems analyzed, the model generated functionally accurate code, demonstrating its potential in fields such as education and professional software development (Fig. 2). Another benefit of using ChatGPT was the clear and structured set of explanations for algorithmic strategies, along with detailed timeframes for test execution, further substantiating the contribution of AI deployment as a productivity tool. A notable capacity of incorporating developers' feedback was observed, as well as iterative improvements of its responses. These findings are consistent with the results of [9], which reveals that AI-based tools require critical thinking and evaluative mindset when used. Nevertheless, AI tools can be successfully leveraged to overcome software development bottlenecks.

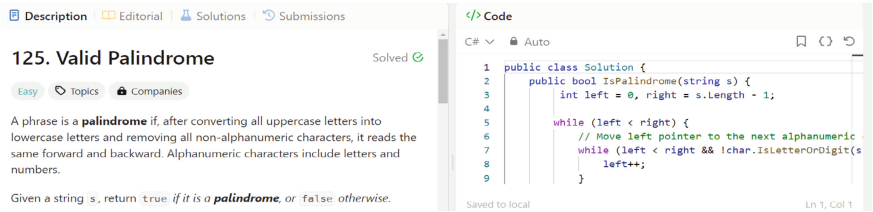


Fig. 2. Problem and solution pair on the LeetCode platform.

The success rate of ChatGPT was close to 97.5%; the chatbot failed to provide correct solutions for ~2.5% of the 200 problems provided. The failure cases were associated with high complexity tasks, such as combinations of several data structures and algorithms. The failure cases can be divided into three main categories:

- The solution passed initial tests but triggered out-of-memory exceptions caused by high execution time;
- The generated code failed tests;
- The solution was semantically incorrect despite syntactic validity.

The overall interaction with the chatbot (ChatGPT) did not reveal major faults, as prompts were in majority well understood. The programmers who performed the tests reported the experience as being similar to pair programming – a collaborative approach to software development where two developers work together to identify the proper solution and share acquired knowledge.

The ~2.5% failure rate set the loop on a major aspect: developers still require fundamental knowledge on data structure and algorithms to effectively interpret, evaluate, and refine AI-generated output. Without such expertise, providing constructive feedback to the model becomes challenging.

3.2 Case Study B

For the generation of unit tests, the aim was to address the gap in research regarding the usability of Generative AI for unit test generation, specifically exploring the capabilities

of Anthropic's Claude AI model in generating unit test cases. Outputs were measured against executability, dependency completeness, and coverage of edge and negative test cases.

The experiment was done based on a .NET parking application's UI component. To support the automation process, the html elements of the component have been captured using a chrome extension. The role of the chrome extension is to capture the component elements in the UI. These elements are given as an input to the backend (an API which is calling the Claude.ai model).

```
26      public async Task<string> GenerateTestsAsync(string htmlContent)
27      {
28          var requestBody = new
29          {
30              model = "claude-3-sonnet-20240229",
31              max_tokens = 4000,
32              messages = new[]
33              {
34                  new
35                  {
36                      role = "user",
37                      content = $@"Generate executable test code for this HTML component using xUnit and bUnit.
38                          Do not provide descriptions or explanations, only provide the actual test code that can be directly used.
39                          Use C# with xUnit and bUnit frameworks.
40                          Format the response as compilable C# code only.
41
42                          HTML Component:
43                          {htmlContent}"
44                  }
45              }
46          };
47
48          var json = JsonSerializer.Serialize(requestBody);
49          var content = new StringContent(json, Encoding.UTF8, "application/json");
```

Fig. 3. Code snip with the model used by Claude.ai.

The captured UI elements are sent as input to an API, which is calling in the background the Claude.ai model. The specific model used is "claude-3-sonnet-20240229", and the prompt used is visible in Fig. 3. This definitive version of the prompt has been chosen after couple trials and errors and seemed to be the one with the least number of hallucinations.

The evaluation indicates that the Claude.ai model is capable of effectively processing the input provided and generating unit tests for HTML elements. It demonstrates creativity in addressing a range of scenarios, appropriately selecting test types such as [Fact], generating meaningful method names, and adhering to conventional naming practices, particularly by employing the "Should" format (e.g., ChangePasswordButtonShouldExist). A major disadvantage observed was that the model produces predominantly positive test scenarios, whereas in practice both positive and negative test scenarios are required in certain contexts. An example is the complementary test "ChangePasswordButtonShouldNotExist()", which is an enhancer of test robustness. The test performed in the present case study was not executable without human assistance when negative tests were needed. The model failed to initiate mock services and essential dependencies, especially in cases where complex components were tested. For a deeper understanding of this phenomenon, a detailed analysis was conducted on one of the Profile tests to identify gaps that required manual intervention for functional test automation.

The analyzed test is called ChangePasswordButtonShouldExist() and it checks if the button for changing the password is available when rendering the Profile component. The reason for choosing to explore this test is its simplicity and commonality in software applications. The test created by Claude.ai is visible in Fig. 4.

```
110
111     [Fact]
        0 references
112     public void ChangePasswordButtonShouldExist()
113     {
114         // Arrange
115         var cut = RenderComponent<ChangePasswordForm>();
116
117         // Act
118         var changePasswordButton = cut.Find("button.rz-button-md.rz-variant-filled.rz-primary.rz-shade-default");
119
120         // Assert
121         Assert.NotNull(changePasswordButton);
122     }
123
```

Fig. 4. Example of GenAI generated test.

The test is arranged correctly, following the unit test framework Arrange, Act, Assert. However, the test is failing because it is missing the services used by Profile component. In this context, "services" refer to a modular, self-contained unit of functionality that performs a specific task and can be accessed independently, often over a network, to support a larger application. To make the test work properly, the mock-ups needed have been put in place. The test was successfully run after human intervention, which highlights that Generative AI can be a good tool for creating a basic starting point when creating unit tests, but it is not so good in generating full code in this scenario. The main reason behind this failure is that the AI model was unaware of the entire code base and could not have a high-level overview. This shows that Generative AI can be used as an assistant in generating Unit tests, but it needs human intervention to deliver a fully working test for this specific scenario.

3.3 Preliminary Conclusions from Experimental Case Studies

The algorithmic code generation study revealed more than 90% success rate for correct solutions, particularly effective in problems involving standard data structures and classical dynamic programming scenarios. The use of ChatGPT-4 enabled generation of complex explanations, thus reducing the actual time and effort put into the testing by the developer. However, some of the problems tested in case study A required refinement and a ~2.5% failure rate was observed. Therefore, developer's vigilance and assistance are still required, especially in unusual situations and for complex requirements. ChatGPT fails to completely assist the developer for applications such as unconventional method signatures and integration of multiple data structures and complex patterns. These findings substantiate the fact that human assistance is still required even though AI automation is largely efficient and enhances productivity. Human experts are still valuable assets in software development, even in the testing phase.

On the other hand, Claude.ai was able to perform testing iterations with accuracy and structural validity. Its proficiency was specifically observed for the creation of positive tests and the adherence to the naming convention. The main flaw identified was the lack of negative test capacity and the inability to initialise dependency services. This limits its immediate utility. The results required human intervention, especially in complex cases, particularly in the late stages of testing. Therefore, AI models act as accelerators, but cannot be deployed as a solution for complete automation of software testing. The

positive side of using Claude.ai lies in its creativity in formulating test scenarios, but it lacked certain contextual elements. This is another highlight of the already-stated idea that LLM tools such as Claude.ai or ChatGPT cannot success-fully solve complex problems without human assistance.

The results presented in this section present both opportunities and challenges. The capacity of AI tools is impressive, significantly reducing the time and effort required from developers at the testing stage. However, experts in the field re-main important: due to their expertise, knowledge, and awareness of potential flaws, they can immediately interfere in the process and remediate specific failures. AI tools remain a solid field for innovation and have the potential to fill the current gaps in software development.

4 Discussions

The main aim of the experimental case studies was to fill in the literature gap on the impact of adopting AI tools for decision making processes and testing in software development. The case studies highlight improvements in productivity, a finding consistent with [13], which is a study documenting a 55.8% reduction in time required for software development by using GitHub Copilot. A similar conclusion is presented in [14], which estimates that deployment of AI for software development generates savings ranging from 20 to 45% from the total spending for software engineering. The main source of savings is represented by the reduction in execution time for activities such as generating initial code drafts, code correction and refactoring, root-cause analysis, and generating new system designs. Although these might seem as small-size effects, the potential of AI deployment is considerable through replication of best practices and routines across organizational functions.

A task-based economic model, such as the one proposed by [12], is based on the scenario where AI-human collaboration is mandatory for reliable high-quality outputs. Human supervision is essential, as complex circumstances require human intervention (e.g., guidance, corrections) when AI tools are only partially reliable. However, supervision and immediate identification of errors are specific to experienced developers, leaving junior software developers in the position where they need to acquire knowledge and expertise through the design, architecture, and testing strategy crafting. Similarly, Brookings [15] indicates that AI augments software development capabilities by complementing the capabilities and knowledge of developers with the aim of increasing productivity and efficiency.

In 2027, there were more than 70% of software developers are expected to use AI coding tools in their regular activity [16]. However, there are major differences in the level of AI integration in a sense that superficial deployments are likely to generate modest effects, whereas formalised, complex use cases are likely to drive improved yields [15, 16]. For deeper AI integration, formal trainings are mandatory, and companies should invest in trainings focused on prompt engineering, model evaluation and responsible integration of AI outputs in the developed software.

In recent years, different reports of international studies have reported that AI software engineering synergy can reduce resource consumption, improve code quality, and reduce developer burnout [8–12]. The economic landscape is being reshaped by the

introduction of AI tools in industries, with expected disparities in labour demand and income distribution [12].

The intensified adoption of AI reflects on the one hand the investments made by companies to increase productivity and operational efficiency and, on the other hand, the expansion of the software market globally [16]. In 2024, software expenditure increased by 12.6% y-o-y due to AI adoption [16]. Nonetheless, early AI adopters have also raised concerns about error propagation, security vulnerabilities and ethical matters. The risk of hallucinations posed by AI use remains a major concern for professionals, as generative AI tools are, at times, misused. In this context, frameworks such as Value-based Engineering and AI governance protocols remain adamant for ethical, transparent, and safe adoption of AI tools. The need for governance structures that require human assistance and supervision for AI-based software development remains a major industry concern.

5 Analysis of the Economic Impact on European Organizations

In the prior sections, the authors debated the foundations and empirical evidence for assessing the economic impact of adopting AI tools in software development This section focuses on presenting a cost-benefit analysis for the European software development market based on developer compensation rates and proven improvements in productivity at European level.

The authors conducted research to collect comprehensive salary data for the European market by country. There are significant regional variations in European countries [17–19]. Based on the data collected, the average developer cost per hour is €30. This cost includes all salary expenses (gross salary, social contributions, benefits), infrastructure expenses, and administrative overhead. The standard working time per month in the European Union is 160 h, as per the EU Working Time Directive 2003/88/EC. This leads to an average salary of €4,800 per month for a software developer. According to the average increase in productivity of 25% identified by [13], in corroboration with the analysis of [14], an experienced software developer is able to deliver outputs equivalent to 200 h of work, executed within the normal 160 h of work per month. This translates into a labour productivity increase of €1,200 per developer per month. At the same time, the subscription for AI tools costs, on average €15 per month [20–22]. Hence, if an AI tool is used, the monthly productivity gain reduces to €1,185 per developer on average. To adopt AI tools, companies incur an initial one-time cost to set up the infrastructure and provide training. However, this cost can be considered negligible in comparison to the overall benefits obtained in the short and long term. Based on subscription costs, a company is expected to spend €180 per developer. The estimated annual gains add up to €14,220, leading to a return on investment (ROI) of 7,800%. It is mandatory to mention that this is a conservatory approach, that excludes set up costs.

Without a doubt, the breakeven point occurs at a very early stage of AI adoption due to the low costs compared to the estimated gains. A team of 10 developers requires an investment of €1,800 annually in AI subscriptions and records an estimated €142,200 in annual savings. This leads to an investment recovery period of only 4.6 working days. Companies differ based on a variety of criteria, from the number of employees and revenue to the regional and national particularities of the market in which they operate.

Return on Investment is an indicator typically calculated in percentage values by reporting the difference between annual savings and annual investments to the annual investments. Annual savings are the total cost savings per year obtained by investing. In the present cost-benefit analysis, the investment is represented by AI tools introduced in the company's operations. Annual Investment is the total cost per year for the solution implemented with the purpose of generating future benefits. The annual investment for adoption of AI tools is represented by subscriptions, trainings for employees who work with AI and infrastructure integration.

Table 1. Return on Investment (ROI) Scenarios for a Team of 50 Developers.

Scenario (AI adoption rate)	Productivity Gain (%)	Monthly Savings	Annual Savings	ROI (%)
Conservative (15%)	10	€35,250	€423,000	4,600
Moderate (25%)	25	€59,250	€711,000	7,800
Optimistic (40%)	40	€95,250	€1,143,000	12,600

The results of the three scenarios (Table 1) are undeniable: even in the conservative scenario, a 50-employee company is expected to achieve an ROI of 4,600%. This represents another proof that AI tools are an investment with reduced risks and high benefits. Considering the low adoption rate of AI across European companies, as compared to those in North America and South-East Asia, this study represents a simple proof of the growth potential they are failing to attract.

6 Conclusions

The paper presents a convergent set of theoretical, empirical, and economic arguments in favor of adopting AI tools for software development. Case studies A and B reveal the utility of AI tools for software testing in a variety of contexts and scenarios. Without a doubt, these tools are a reliable instrument for reducing the time allocated to simple repetitive tasks that are oftentimes highly time-consuming. The economic analysis performed in the paper further substantiates the financial benefits of adopting AI tools for software development, as the yields of such an investment are more than impressive. A modest increase of 15% in productivity is connected to a ROI of 4,600% in the case of a 50-employees European company. Also, AI tools can perform basic tasks independently, leaving professionals to focus on more strategic, value-added activities. However, AI tools have limitations when used for complex requirements, proving that human assistance remains important when deploying AI for software development-related tasks. The role of humans remains extremely important. AI tools cannot replace humans, as they are built to improve their performance and will generate changes in the current labour market.

The limitations of the study are represented by potential bias in analysing the impact, benefits, and usefulness of AI tools, since only several tools were tested for the purpose

of the presented research. Further exploration of AI tools and their impact should be performed to ensure completeness of the analysis. In addition, the rapid evolution of AI tools makes the present analysis obsolete if further updated studies are not performed. Nonetheless, future work should quantify ROI by organizational scale, country, and software maturity levels. AI models are continuously updated and refined. The performance and capabilities observed during our study might differ from those of the newer versions of the same tools. This inherent dynamism poses a challenge for long-term reproducibility and generalizability of results.

The economic case presented here provides a foundation for evidence-based decision-making regarding AI tool adoption in European software development organizations. Even though AI tools fail at times when unusual or complex situations require solutions, their level of independence remains satisfactory for the objectives of software development activities, as well as for other organizational activities characterised by predictability and low decision-making effort.

Humans remain at the core of all economic activities, whereas AI tools are valuable assistants that enable companies to leverage high ROI due to cost savings and efficiency gains.

References

1. IMF: How Europe Can Capture the AI Growth Dividend. https://www.imf.org/en/blogs/articles/2025/11/20/how-europe-can-capture-the-ai-growth-dividend (2025), last accessed 2026/01/21
2. European Commission: Artificial Intelligence: Economic Impact, Opportunities, Challenges. https://economy-finance.ec.europa.eu/document/download/10867a27-f49f-4aa9-89bb-feb752513f78_en?filename=dp210_en_artificial%20intelligence_0.pdf (2024), last accessed 2026/01/21
3. Stanford HAI: The 2025 AI Index Report: Economy. https://hai.stanford.edu/ai-index/2025-ai-index-report/economy, last accessed 2026/01/21
4. PIMCO: The Economics of AI Scale. https://www.pimco.com/gb/en/insights/the-economics-of-ai-scale (2025), last accessed 2026/01/21
5. World Economic Forum: Europe can still win with AI. https://www.weforum.org/stories/2026/01/europe-can-still-win-with-ai-the-key-is-focusing-on-physical-ai/ (2026), last accessed 2026/01/21
6. DHS: The Impact of Generative AI on Software Engineering Activities. https://www.dhs.gov/sites/default/files/2025-01/2024_1219_impact_of_genai_on_software_engineering_activities_minkiewicz.pdf (2025), last accessed 2026/01/21
7. ArXiv: Utilizing ChatGPT in a Data Structures and Algorithms Context. https://arxiv.org/html/2410.08899 (2025), last accessed 2026/01/21
8. Weber, T., et al.: Significant Productivity Gains through Programming with LLMs. https://www.mmi.ifi.lmu.de/pubdb/publications/pub/weber2024eics-llm/weber2024eics-llm.pdf (2024), last accessed 2026/01/21
9. TestQuality: AI's Impact on Software Testing in 2025. https://testquality.com/how-ai-is-transforming-software-testing/, last accessed 2026/01/21
10. Pandhare, H. V.: Future of Software Test Automation Using AI/ML. https://www.researchgate.net/profile/Harshad-Vijay-Pandhare-2/publication/391806293_Future_of_Software_Test_Automation_Using_AIML/links/68277cda026fee1034f8a449/Future-of-Software-Test-Automation-Using-AI-ML.pdf (2025), last accessed 2026/01/21

11. Software Testing Magazine: Early Testing Saves Time and Money in Software Development. https://www.softwaretestingmagazine.com/knowledge/how-early-stage-testing-saves-time-and-money-in-software-development/ (2025), last accessed 2026/01/21
12. Acemoglu, D.: The simple macroeconomics of AI. NBER Working Paper 32487. National Bureau of Economic Research, https://www.nber.org/papers/w32487 (2024), last accessed 2026/01/21
13. Peng, S., Kalliamvakou, E., Cihon, P., Demirer, M.: The impact of AI on developer productivity: evidence from GitHub Copilot. arXiv preprint https://arxiv.org/abs/2302.06590. (2023) https://arxiv.org/abs/2302.06590
14. McKinsey & Company: The economic potential of generative AI: The next productivity frontier. https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier (2023), last accessed 2026/01/21
15. Cihon, P., Demirer, M.: How AI-powered software development may affect labor markets. Brookings Institution. https://www.brookings.edu/articles/how-ai-powered-software-development-may-affect-labor-markets/ (2023), last accessed 2026/01/21
16. Gartner: Augmenting the software development life cycle with AI. https://www.gartner.com/en/articles/set-up-now-for-ai-to-augment-software-development (2024), last accessed 2026/01/21
17. CodeSubmit: Average software engineering salaries by country. https://codesubmit.io/blog/software-engineer-salary-by-country/ (2024), last accessed 2026/01/21
18. Glassdoor: Software developer salary in Germany 2025. https://www.glassdoor.com/Salaries/germany-software-developer-salary-SRCH_IL.0,7_IN96_KO8,26.htm, last accessed 2026/01/21
19. ICTS Custom Software: Software developer salary in top outsourcing countries. https://icts.io/2024/01/10/software-developer-salary-comparison-in-top-5-outsourcing-countries-2024 (2024), last accessed 2026/01/21
20. DX Research: AI coding assistant pricing 2025: Complete cost comparison. https://getdx.com/blog/ai-coding-assistant-pricing/ (2024), last accessed 2026/01/21
21. MarkTechPost: Codeium vs. Tabnine: Comparison of key features and benefits. https://www.marktechpost.com/2024/10/03/codeium-vs-tabnine-comparison-of-key-features-and-benefits/ (2024), last accessed 2026/01/21
22. TechCrunch: GitHub Copilot introduces new limits, charges for 'premium' AI models. https://techcrunch.com/2025/04/04/github-copilot-introduces-new-limits-charges-for-premium-ai-models/ (2025), last accessed 2026/01/21

Transforming Higher Education: A Case Study on the Integration of Coursera at the University of Hradec Králové

Petra Poulová(✉) and Miloslava Černá

University of Hradec Králové, Rokutanského 62, 500 03 Hradec Králové, Czech Republic
{Petra.Poulova,Miloslava.Cerna}@uhk.cz

Abstract. This study presents a qualitative case study on the integration of the Coursera for Campus platform at the Faculty of Informatics and Management, University of Hradec Králové. Triggered by the COVID-19 pandemic, the initiative evolved from a pilot program into a fully institutionalized form of 'virtual mobility,' governed by a Dean's Directive and integrated into the university's credit system. Using a mixed-methods approach, we analyzed quantitative platform data from October 2022 to May 2023 and qualitative testimonials from students and academic staff. The findings reveal high engagement, with 1,214 course enrolments and a 54.2% completion rate, significantly surpassing typical MOOC benchmarks. Qualitative data indicate that students value practical skills, professional certification, and language enhancement, while academic staff use the platform for continuous professional development and curriculum enrichment. The study concludes that formal, credit-bearing integration is a key factor in maximizing the value of MOOCs within higher education, offering a successful model for bridging the gap between the potential of online platforms and their practical application.

Keywords: Coursera · COVID-19 · MOOC · online learning · student satisfaction · higher education · technology integration

1 Introduction

The landscape of modern education is undergoing a profound transformation, driven by technological advancements and a growing demand for accessible, flexible learning opportunities. In this context, Massive Open Online Courses (MOOCs) have emerged as a significant force, democratizing access to quality education for learners worldwide [1]. Platforms like Coursera offer a vast array of courses from leading universities and companies, enabling individuals from diverse socioeconomic and geographical backgrounds to acquire new skills and knowledge. These platforms facilitate self-paced learning, effectively bridging educational gaps and fostering a culture of lifelong learning [2]. As traditional higher education institutions (HEIs) adapt to these new paradigms, integrating MOOCs into their curricula presents both a compelling opportunity and a complex challenge.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 312–321, 2027.
https://doi.org/10.1007/978-3-032-32104-6_27

The COVID-19 pandemic served as a powerful catalyst, accelerating the adoption of online learning and compelling HEIs to re-evaluate their pedagogical models [3, 4]. This sudden shift not only ensured educational continuity during a global crisis but also laid the groundwork for a more resilient, technologically integrated future of education [5]. For institutions like the Faculty of Informatics and Management (FIM) at the University of Hradec Králové (UHK), which has a long-standing history of engagement with e-learning, this period highlighted the potential of leveraging external platforms to enrich its educational offerings.

This paper presents a case study of the integration of the Coursera for Campus program at the Faculty of Informatics and Management, University of Hradec Králové. It examines the implementation process, from the initial pilot phase during the pandemic to its full integration into the faculty's curriculum, guided by a formal Dean's directive. The study aims to provide practical insights into the benefits and challenges of such an integration, drawing on both quantitative data and qualitative feedback from students and academic staff. By analyzing this real-world example, we seek to answer the following research questions:

- RQ1: What are the key factors and procedural steps that facilitate the successful integration of a MOOC platform like Coursera within a university faculty?
- RQ2: How do students and academic staff perceive the value, benefits, and challenges of using Coursera courses as a supplement to traditional university education?

Through this investigation, the study contributes to the broader discourse on the future of higher education, offering a detailed account of how to bridge the gap between the theoretical potential of MOOCs and their practical application in a specific institutional context.

2 Literature Review

The evolution of Massive Open Online Courses (MOOCs) is a story of rapid adaptation to the changing needs of both learners and educators. Initially rooted in the open education movement, the first cMOOCs (connectivist MOOCs) emphasized networked learning and knowledge co-creation. However, the landscape shifted dramatically with the rise of xMOOCs, platform-centric courses offered by providers like Coursera, edX, and Udacity, which focused on scaling content delivery from established institutions [6]. Researchers have chronologically divided the evolution of MOOCs into distinct phases, such as the starting phase (2009–2012), a period of high growth (2013–2015), a plateau (2016–2018), and a peak phase (2019–2021), with research topics shifting from initial explorations to a deeper focus on learner continuance and the integration of MOOCs into blended learning programs [7].

Technological advancements have been a primary driver of this evolution. Modern MOOCs incorporate a variety of interactive tools, mobile learning capabilities, and data-driven personalization to enhance learner engagement and accessibility [1]. These technological interventions have significantly expanded educational choices, allowing for more interactive and flexible learning environments [2]. However, the effective design and delivery of MOOCs require a generalized framework that addresses underlying pedagogical issues to meet the diverse needs of a global audience [2].

In parallel with technological growth, the pedagogical frameworks surrounding MOOCs have also matured. The focus has shifted from simple content dissemination to creating more engaging and motivating learning experiences. Research has increasingly focused on understanding learner motivation, expectations, and the factors that contribute to course completion rates, which have historically been a significant challenge for MOOCs [8]. New taxonomies have been proposed to classify MOOCs based on pedagogical criteria, aiming to better align course design with disciplinary knowledge and learners' specific needs, such as those in teacher training programs [8]. This reflects a broader trend towards recognizing MOOCs not just as standalone courses, but as valuable components within larger educational ecosystems, including blended learning models in traditional HEIs [7, 9].

The COVID-19 pandemic acted as a significant accelerant in this process. As universities worldwide were forced to transition to remote instruction, MOOC platforms became a critical resource for ensuring educational continuity [10]. This period highlighted the key advantages of MOOCs—namely, their flexibility, accessibility, and the vast repository of high-quality content available for immediate deployment [11]. While the emergency transition to online learning was not without its challenges, particularly regarding disparities in ICT support and digital literacy, it also demonstrated the viability of integrating MOOCs into formal university curricula at scale. This study builds upon this context, moving beyond the general discussion of MOOCs to examine the specific, practical steps and outcomes of a structured integration within a university faculty, thereby addressing a gap in the literature concerning detailed, institution-level case studies of post-pandemic MOOC adoption.

3 Methodology

This study adopts a mixed-methods single-case study design to examine the integration of the Coursera for Campus program at the Faculty of Informatics and Management (FIM) at the University of Hradec Králové (UHK). The design is qualitatively driven. The primary aim is to capture processes, experiences, and perceived value from the perspective of key stakeholders, while incorporating descriptive quantitative indicators to contextualize engagement and outcomes at the aggregate level. The case study methodology is well-suited to exploring a real-life, contemporary bounded system (the integration process). It allows for a rich, detailed understanding of the context, processes, and stakeholder perspectives.

3.1 Context of the Study

The Faculty of Informatics and Management at UHK provided a fertile ground for this initiative. With over two decades of e-learning experience, dating back to the turn of the millennium, the faculty had a well-established technological infrastructure and experienced personnel dedicated to online education. This substantial background, encompassing hundreds of internally developed e-courses and a long-standing use of blended learning, gave the faculty a significant advantage in adapting to the abrupt shift to fully online instruction mandated by the closure of Czech universities on March 11, 2020, due to the COVID-19 pandemic.

3.2 The Implementation Process

The integration of Coursera was a phased process, evolving from a reactive measure into a strategic, fully integrated component of the faculty's educational offerings. The chronology of the implementation is as follows:

1. *Initial Pilot Phase (Summer Semester 2021/22):* During the final semester of pandemic-related teaching restrictions, the faculty capitalized on an offer from Coursera for free licenses for students. This pilot phase allowed both students and staff to gain valuable firsthand experience with the platform's content and interface, serving as a proof of concept.
2. *Strategic Investment (Academic Year 2022/23):* Building on the positive feedback and experience from the pilot, the faculty leadership made the strategic decision to invest in the purchase of Coursera licenses for its students, despite a general sense of "online fatigue" following the pandemic. This marked the transition from a temporary solution to a long-term educational enhancement.
3. *Formalization and Regulation (Dean's Directive No. 2/2022):* To ensure a structured and transparent integration, the process was formalized through a Dean's Directive. This key document established the rules for "virtual mobility" and clarified the process for both students and employees. It outlined the selection procedure for license allocation (prioritizing students with strong academic results and those who had not previously participated), the obligations of participants (e.g., completing at least one course), and, most importantly, the mechanism for recognizing successfully completed courses and converting them into ECTS credits within the university's official study system (up to 10 ECTS per academic year).

3.3 Data Collection and Analysis

To answer the research questions, this study used a mixed-methods approach to data collection, combining quantitative platform data with qualitative user feedback.

Quantitative data comprised aggregated Coursera for Campus administrative metrics. These data included the total number of courses users enrolled in and the number of courses successfully completed. User activity patterns were also analyzed to identify trends in platform usage over time.

Qualitative data were obtained through unsolicited written testimonials from a purposive sample consisting of students and academic staff member who were active users of the platform. These narratives provided rich, detailed accounts of their personal experiences, perceived benefits, and challenges.

Data analysis was conducted in two streams. The quantitative data were analyzed using descriptive statistics to summarize enrolment, completion rates, and activity trends. The qualitative data from the written testimonials were analyzed using a thematic approach to identify and categorize recurring themes, such as platform usability, course quality, skill development, and language enhancement.

3.4 Limitations of the Design

The qualitative core relies on a small, self-selected sample, which constrains breadth and generalizability; quantitative indicators are administrative and descriptive, not inferential, and cannot establish causality or control for confounds, e.g., prior motivation. These constraints are acknowledged and addressed in the Discussion and Conclusion through careful scoping of claims.

4 Results

This section presents the findings from data collected during the implementation of the Coursera for Campus program at FIM UHK, divided into quantitative platform metrics and qualitative participant feedback.

4.1 Quantitative Findings

The quantitative data reveal a significant level of engagement with the Coursera platform among both students and staff. Between October 2022 and May 2023, a total of 1,214 courses were enrolled in by users. Of these, 658 courses were successfully completed by May 10, 2023, representing an overall completion rate of approximately 54.2%. This rate is notably higher than the single-digit completion rates often reported for open, unintegrated MOOCs. As a direct result of the formal integration, 50 students successfully met the requirements outlined in the Dean's Directive and earned ECTS credits for their virtual mobility.

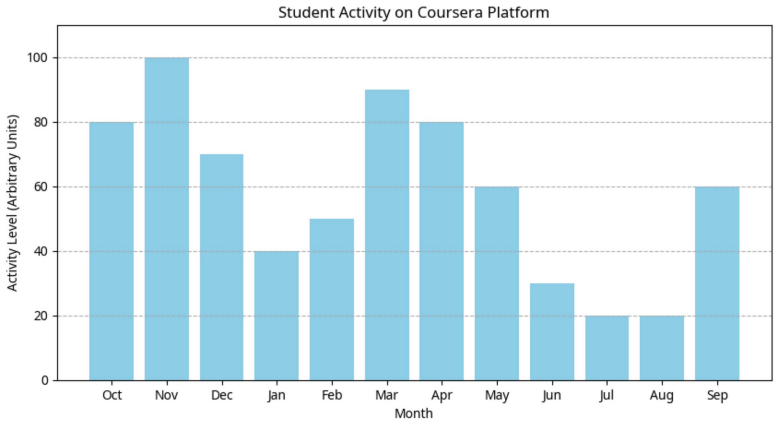


Fig. 1. Student Activity on Coursera Platforms.

Analysis of user activity logs revealed distinct engagement patterns for students and academic staff, as illustrated in Figs. 1 and 2.

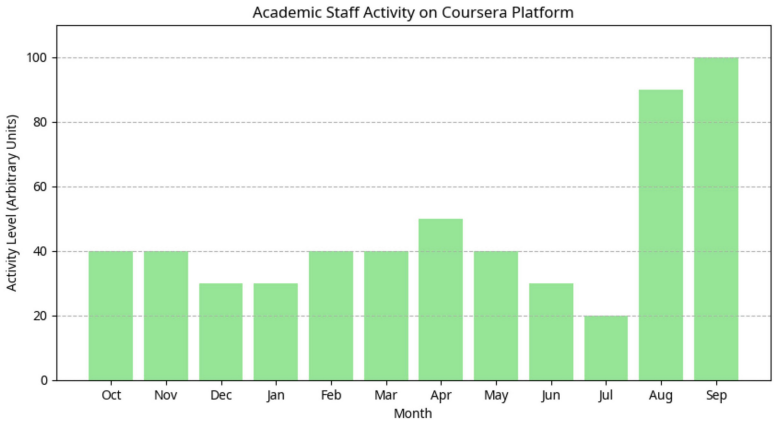


Fig. 2. Academic Staff Activity on Coursera Platform

4.2 Qualitative Findings

The qualitative data, derived from written testimonials, provides a deeper understanding of the participant experience. The feedback was overwhelmingly positive, with several key themes emerging from the narratives of students and academic staff.

Student Perceptions

The student perspectives expressed through the testimonials resonate closely with several patterns widely discussed in MOOC research. Both Lucie and Jaroslav emphasized the breadth and academic quality of the courses they completed, noting the wide range of topics and reputable institutions. This aligns with existing literature describing MOOCs as vehicles for expanding learners' educational choices and providing access to high-quality content beyond the constraints of local curricula. The students' references to courses ranging from HR and marketing to digital footprints and project management thus illustrate, at the micro level, what the literature characterizes as one of the defining strengths of MOOC ecosystems: their diversity, scalability, and disciplinary breadth.

Practical relevance emerged as another central theme in the students' accounts. Jaroslav's appreciation of the Digital Marketing course from Meta—particularly its structured explanations, video-based demonstrations, and real world examples—reflects a broader trend identified in empirical MOOC research, which suggests that learners tend to value courses that offer clear applicability to their academic or professional trajectories. Prior studies repeatedly emphasize that practical scaffolding and authentic tasks enhance learner motivation and persistence, helping bridge the gap between theoretical knowledge and workplace demands. The students' feedback thus illustrates how practice-oriented Coursera courses can function as complementary components within formal higher education programmes.

The testimonials also highlighted a benefit often acknowledged but less frequently foregrounded in the literature: language development. Lucie's observation that her

English proficiency improved as a by-product of course participation underscores findings from studies noting that EMI MOOCs can serve as informal, immersive environments for language exposure. Although not originally designed as language-learning tools, English-medium MOOCs often contribute to students' receptive skills, technical vocabulary acquisition, and increased confidence, thereby indirectly enhancing their academic preparedness.

Professional signalling was another recurrent theme. The students noted that the certificates they obtained are valuable for their CVs and online professional profiles, such as LinkedIn. This aligns with a substantial body of research indicating that credentialing—particularly microcredentials and industry-issued certificates—plays an important motivational role for learners and may support early-career transitions. The students' comments thus illustrate how external certification offered by platforms like Coursera can augment employability narratives, functioning as portable evidence of specific competences.

Finally, the emphasis on platform usability corresponds to a well-documented determinant of MOOC engagement and learner satisfaction. Lucie's remarks about the intuitive interface and the convenience of mobile access reflect findings that ease of navigation and technological reliability significantly reduce cognitive load, enabling learners to focus more fully on content rather than on managing the learning environment. In this sense, the students' experiences reaffirm that positive user experience is not a marginal feature but a critical component of effective online learning design.

Taken together, the testimonials not only offer insight into individual learner experiences but also provide concrete examples of broader patterns reflected in con-temporary MOOC research: the value of breadth and quality of content, the significance of practical relevance, the incidental yet meaningful role of language enhancement, the importance of credentialing in shaping professional identity, and the centrality of platform usability in supporting learner engagement.

Academic Staff Perceptions

The testimonial from the academic staff member highlights several dimensions of professional and pedagogical value associated with integrating Coursera, which closely align with themes identified in contemporary research on MOOCs and technology-enhanced professional development.

A notable benefit is access to specialized, high-level training that is often unavailable or financially inaccessible in the local context. The staff member emphasized the opportunity to complete a suite of certified courses offered by the International Information System Security Certification Consortium (ISC^2), which would typically require participation in costly commercial programmes. This reflects broader findings suggesting that MOOC platforms function as important democratizing tools for faculty development, expanding access to niche or industry specific expertise that institutions may struggle to provide internally.

Another prominent theme concerns the direct transfer of newly acquired knowledge into teaching practice. The academic noted that course completion had significantly strengthened their professional competencies, which they subsequently integrated into their own instruction and specialized subjects. This aligns with literature indicating

that MOOCs can serve as catalysts for pedagogical renewal: by engaging with up-to-date disciplinary content and observing diverse instructional models, educators often enrich their teaching materials, enhance curricular relevance, and align more closely with emergent industry standards. In this sense, the platform operated not merely as a means of individual upskilling but as a mechanism for maintaining and elevating the overall currency and quality of teaching.

The testimonial also underscored the value of benchmarking—an element frequently highlighted in studies on academic engagement with MOOCs. By encountering instructional approaches used by leading universities worldwide, the academic staff member gained comparative insights that offered a basis for reflecting on local pedagogical practices at UHK. Such exposure supports what the literature describes as "informal global peer review," in which educators adopt, adapt, or critically evaluate methodologies observed elsewhere. This process fosters pedagogical innovation and contributes to a broader institutional culture of quality enhancement.

Finally, the academic emphasized the importance of flexibility. The asynchronous and self-paced nature of the courses enabled the staff member to pursue professional development despite a demanding academic schedule. This finding mirrors research demonstrating that temporal flexibility is a central determinant of participation in online learning for faculty members, for whom heavy teaching loads, administrative duties, and research commitments often limit opportunities for structured training. By allowing educators to progress through course materials at their own pace, MOOCs provide a sustainable avenue for continuous professional growth.

Taken together, the academic staff member's perspective illustrates how Coursera can play a multifaceted role in supporting faculty development: by widening access to specialized expertise, enhancing teaching practice, facilitating pedagogical reflection through international comparison, and accommodating the scheduling realities of academic life. These individual experiences reinforce the broader argument that MOOC integration, when strategically implemented, can contribute meaningfully not only to student learning but also to the professional and pedagogical evolution of university educators.

5 Discussion

The results of this case study offer valuable insights into the successful integration of a MOOC platform within a university setting and address the research questions posed in the introduction. The findings suggest that a structured, formalized approach is key to leveraging the full potential of platforms like Coursera, moving them from a peripheral resource to a core component of the educational ecosystem.

Regarding the first research question (RQ1), which focused on the factors and steps for successful integration, our experience points to several critical elements. The pre-existing e-learning infrastructure and culture at FIM UHK undoubtedly provided a solid foundation, mitigating some of the technological and pedagogical challenges noted in the literature [12]. However, the most crucial step was formalizing the process through the Dean's Directive. By establishing clear rules for participation and, most importantly, creating a direct link between course completion and the university's credit system

(ECTS), the faculty provided a powerful extrinsic motivator for students. This formal recognition likely played a significant role in the high completion rate (54.2%) observed in our study, which stands in stark contrast to the typically low rates for standalone MOOCs. The phased approach, starting with a pilot before a full-scale investment, proved prudent, allowing the institution to gauge interest and gather feedback before committing significant resources.

The second research question (RQ2) explored the perceived value and benefits for students and academic staff. The qualitative findings clearly indicate that both groups derived significant value from the integration, albeit in different ways. For students, the primary benefits included access to a wide range of high-quality, practical content that complemented their formal studies, the ability to enhance their professional profiles with recognized certificates, and the opportunity to improve their English language skills. This aligns with research highlighting the importance of MOOCs for skill development and career enhancement [7].

For academic staff, the platform served as a vital tool for continuous professional development. It provided access to specialized, up-to-date knowledge that could be directly integrated into their own teaching, thereby enhancing the quality and currency of the faculty's curriculum. Furthermore, it offered a unique window into the pedagogical practices of other leading global institutions, fostering informal benchmarking. The platform's flexibility was a key enabler, allowing busy academics to pursue learning opportunities at their own pace. These findings underscore the dual role of MOOC platforms in HEIs: not only as a resource for students but also as a powerful mechanism for faculty development.

This case study demonstrates a successful model of bridging theory and practice. While the literature has extensively discussed the potential of MOOCs, this study provides a concrete example of how that potential can be realized through thoughtful and strategic implementation. The synergy between the university's internal e-learning expertise and Coursera's external resources created a learning environment that was richer and more flexible than either could have provided on its own.

6 Conclusion

This paper has presented a case study on the integration of the Coursera for Campus platform at the Faculty of Informatics and Management, University of Hradec Králové. The study detailed the implementation process, from an initial pandemic-driven pilot to a fully formalized and credit-bearing component of the curriculum. The results demonstrate a high level of engagement among both students and academic staff, with a completion rate that significantly exceeds typical MOOC benchmarks.

The key takeaway from this study is that the value of MOOC platforms for higher education institutions is maximized through deep and formal integration. Simply providing access is not enough; the true potential is unlocked when external courses are woven into the fabric of the university's own academic structures, complete with credit recognition and clear administrative support. The experience at FIM UHK shows that such an integration can successfully supplement the curriculum, enhance student skills, and provide a valuable channel for faculty professional development.

However, this study is not without its limitations. As a single-institution case study, the findings may not be universally generalizable to all HEIs, particularly those without a comparable pre-existing e-learning culture. The qualitative data, while rich, was drawn from a small, self-selected sample of participants. Future research could build upon this work by conducting larger-scale surveys to gather more representative data on student and faculty perceptions. Longitudinal studies could also track the long-term impact of this integration on student career outcomes and faculty teaching practices. Comparative studies across institutions and disciplines would also provide a broader understanding of the contextual factors shaping the success of such initiatives.

In conclusion, the integration of platforms like Coursera represents a promising pathway for the evolution of higher education. By thoughtfully combining their own strengths with the vast resources of the digital world, universities can create more dynamic, relevant, and accessible learning experiences for the 21st century.

References

1. Ronkowitz, K., Ronkowitz, L.C.: MOOCs: evolution and revolution. In: McKay, E., Lenarcic, J. (eds.) Macro-Level Learning through Massive Open Online Courses (MOOCs): Strategies and Predictions for the Future, pp. 183–211. IGI Global Scientific Publishing, Hershey, PA (2015). https://doi.org/10.4018/978-1-4666-8324-2.ch011
2. Zhu, M., Sari, A.R., Lee, M.M.: Trends and issues in MOOC learning analytics empirical research: a systematic literature review (2011–2021). Educ. Inf. Technol. **27**, 10135–10160 (2022). https://doi.org/10.1007/s10639-022-11031-6
3. Rof, A., Bikfalvi, A., Marques, P.: Pandemic-accelerated digital transformation of a born digital higher education institution. Educ. Technol. Soc. **25**(1), 124–141 (2022)
4. Fauzi, M.A.: E-learning in higher education institutions during the COVID-19 pandemic: current and future trends through bibliometric analysis. Heliyon. **8**(5) (2022)
5. El Galad, A., Betts, D.H., Campbell, N.: Flexible learning dimensions in higher education: aligning students' and educators' perspectives for more inclusive practices. Front. Educ. **9**. Frontiers Media SA (2024)
6. Martin, F., Oyarzun, B.: Distance learning: standards, perspectives, and best practices of online education. In: Foundations of Learning and Instructional Design Technology: Historical Roots & Current Trends. EdTech Books (2023)
7. Zhang, S., et al.: MOOCs as a research agenda: changes over time. Int. Rev. Res. Open Distrib. Learn. **23**(4), 193–210 (2022)
8. Menendez-Alvarez-Hevia, D., et al.: Contributions of futures studies to education: a systematic review. Comunicar: Media Educ. Res. J. **30**(73), 9–20 (2022)
9. Stracke, C.M., Bozkurt, A.: Evolution of MOOC designs, providers and learners and the related MOOC research and publications from 2008 to 2018. In: International Open and Distance Learning Conference Proceedings Book: 14-15-16 kasım/november 2019, pp. 13–20. Anadolu Universitesi, Eskişehir (2019). https://doi.org/10.5281/zenodo.3598418
10. Simonova, I., Faltynkova, L., Kostolanyova, K.: On the reflection of online distance instruction into blended teaching and learning. In: Blended Learning: Lessons Learned and Ways Forward: 16th International Conference on Blended Learning, ICBL 2023, Hong Kong, China, July 17–20, 2023, Proceedings, pp. 97–110. Springer-Verlag, Berlin, Heidelberg (2023). https://doi.org/10.1007/978-3-031-35731-2_10
11. Rulinawaty, R., et al.: Massive open online courses (MOOCs) as catalysts of change in education during unprecedented times: a narrative review. Jurnal Penelitian Pendidikan IPA. **9**(SpecialIssue), 53–63 (2023). https://doi.org/10.29303/jppipa.v9iSpecialIssue.6697

Digital Education as a Tool for Shaping Modern Human Capital: A Comparative Analysis of Eastern European Countries

Viktoriia Savitska[1](✉), Oksana Gomotiuk[1], Mykola Karmaz[1], Tomáš Lengyelfalusy[2], Ervín Weiss[3], and Filip Makan[4]

[1] West Ukrainian National University, 11 Lvivska st., 46009 Ternopil, Ukraine
Savitska.Vika@ukr.net, o.homotiuk@wunu.edu.ua
[2] DTI University, Sládkovičova 533/20, 018 41 Dubnica nad Váhom, Slovakia
lengyelfalusy@dti.sk
[3] Constantine the Philosopher University in Nitra, Trieda Andreja Hlinku 1, Chrenová, 949 74 Nitra, Slovakia
eweiss@ukf.sk
[4] Catholic University in Ružomberok, Spišská Kapitula 12, 053 04 Spišské Podhradie, Slovakia

Abstract. Within the framework of this study, digital education is conceptualized as a mechanism for the contemporary formation of human capital; however, rather than focusing solely on aggregate accessibility, the article emphasizes the internal symmetry of digital infrastructure within national education systems. Using officially reported internationally comparable indicators aligned with SDG 4.a.1, the dataset covers Eastern European countries over the period 2020–2024. Projected values for 2025 are included only in the illustrative materials and not in the statistical testing. Digital infrastructure is operationalized through access to computers and the Internet in schools across three educational levels (primary, lower secondary, and upper secondary), on the basis of which composite indices and intra-system inequality measures are calculated.

Descriptive results indicate a persistent hierarchy of Internet access across educational levels: primary education demonstrates a lower mean level of access compared to lower secondary and upper secondary education, accompanied by substantially higher variability. Repeated-measures analysis of variance confirms statistically significant within-system differences by educational level, as well as a significant interaction between educational level and country.

Post hoc comparisons reveal that inequality is concentrated at the primary education level, while differences between lower and upper secondary education are not statistically significant. The results indicate that a high level of national digitalization does not eliminate internal disparities, and that digital isolation at early stages may limit the cumulative development of competencies. These findings highlight the importance of targeted policy interventions aimed at ensuring a balanced digital infrastructure at all levels of education, particularly at the initial stage. Furthermore, the study emphasizes the need for methodological consistency based on observed data, which enhances the robustness and reliability of the obtained empirical conclusions.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 322–336, 2027.
https://doi.org/10.1007/978-3-032-32104-6_28

Keywords: digital education · human capital · SDG · structural digital inequality · digital divide · ICT · Eastern Europe

1 Introduction

The concept of human capital occupies a central position in contemporary economic and educational research and is defined as the aggregate of knowledge, competencies, skills, and personal attributes that individuals acquire through education and professional training and that contribute to economic productivity and social well-being. Traditionally, investments in education and healthcare have been understood as critically important determinants of human capital formation; however, the rapid diffusion of digital technologies is transforming the ways in which education contributes to these processes.

According to the updated Human Capital Index (HCI) developed by the World Bank for 2020, Eastern European economies – such as the Czech Republic, Poland, Slovakia, Hungary, Bulgaria, Romania, Ukraine, and Moldova – demonstrate moderate levels of human capital (ranging from 0.65 to 0.78) in comparison with global benchmarks, with values generally lower than those observed in high-performing Organisation for Economic Co-operation and Development (OECD) countries (for example, scores above 0.7 in Australia, Canada, Sweden, Finland, the United Kingdom, and Germany) [1]. This indicates disparities in learning outcomes and health conditions, as well as differences in how effectively education systems transform schooling into skills relevant for future productivity. At the same time, the United Nations Sustainable Development Goal (SDG) 4, which aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all," emphasizes the importance of building education systems that provide adequate infrastructure and learning environments [2]. Within SDG 4, indicator 4.a.1 specifically tracks the proportion of schools offering basic services, such as access to the Internet and computers for pedagogical purposes, disaggregated by level of education. According to the official metadata of the UNESCO Institute for Statistics (UIS), this indicator is designed to measure the infrastructural readiness of education systems to support effective teaching and learning through ICT tools, which constitutes a key prerequisite for digital education.

Considering these international trends and analytical frameworks, Eastern Europe represents a particularly relevant context for examining the role of digital education in human capital formation. Although national strategies across the region increasingly prioritize digitalization, the structural distribution of digital resources—especially across different levels of education – may significantly influence cumulative learning trajectories and labor market readiness. Therefore, this study adopts a structural perspective by analyzing cross-national data on access to ICT in schools in order to assess whether digital education infrastructure is equitably distributed within national education systems and how this distribution may contribute to the contemporary development of human capital.

1.1 Digital Technologies in Human Capital Development

Sairmaly [3] notes that digital information and communication technologies (ICT) in the educational process contribute to human capital development by expanding access to learning, improving the quality of pedagogical interactions, and providing learners with competencies necessary for participation in the contemporary knowledge-based economy.

Similar conclusions are drawn by Bahri [4], and Shatila [5], who emphasize that the digitalization of education fosters human capital through the development of key competencies. The literature also extensively addresses the potential of advanced AI-based technologies [6], Learning Management Systems (LMS) [7], cloud computing solutions [8], and other digital tools as drivers of educational effectiveness and skills formation.

1.2 Lifelong Learning as a Driver of Human Capital Development

Markozova analyzes the essence of the concept of lifelong learning and demonstrates that educational capital constitutes the primary resource for individual self-realization in the information society [9]. Rykovska, in turn, synthesizes European experience in the formalization of education, emphasizing its economic and social feasibility as a key factor in human capital development and quality enhancement [10]. The role of lifelong learning in the context of human capital development is also examined by Hendricks [11], and highlight the contribution of education to economic growth and improvements in the quality of human capital, substantiates the necessity of developing human potential through continuous learning across the life course.

The formation of human capital as a specific economic resource enables countries not only to increase the rate of economic growth by expanding national output, but also to ensure a more balanced distribution of economic gains [12]. This indicates the presence of effective financial mechanisms that support the regulatory function of the state.

Moreover, recent empirical evidence suggests that the digital economy may contribute to reducing regional income inequality; however, this effect is highly heterogeneous and strongly dependent on the level and quality of human capital. This implies that digital transformation alone is insufficient without adequate investment in human capital development. At the same time, the ambiguous role of educational investments highlights a potential mismatch between existing educational resources and the demand for digital skills.

Given that Onopriienko et al. [13] emphasize the opportunities provided by digital technologies, it becomes necessary to analyze structural inequality in access to technologies within schools as a prerequisite for comprehensive human capital formation. Thus, the present study addresses a gap in the literature that predominantly focuses on the availability and potential of new technologies by shifting attention to the equity of access, which is critical for the development of competencies among all learners.

2 Methods

2.1 Theoretical Background

The empirical analysis is based on official internationally comparable indicators of school-level digital infrastructure [2] corresponding to SDG Indicator 4.a.1, which captures access to digital and information and communication technologies (ICT) for pedagogical purposes.

The study covers the period from 2020 to 2025, with values for 2025 designated as author-generated forecasts in cases where official data were not yet available. Forecasting was performed in Microsoft Excel using the formula MIN(100; FORECAST.ETS(…)), which estimates future values through exponential triple smoothing (seasonal forecasting) while constraining the output to a maximum value of 100. The dataset presented in Appendix A includes selected Eastern European countries (Bulgaria, Moldova, Poland, Romania, Slovakia, Hungary, Ukraine, and the Czech Republic). However, due to the absence of consistent data, Hungary, Romania, Bulgaria, and Poland were excluded from the statistical analysis and considered instead as illustrative case studies. Provided that relevant data become available, the analytical methodology can be scaled to a broader range of countries or applied in a global context with regional disaggregation.

For each country and year, six core indicators were employed: (1) access to computers for pedagogical purposes at the primary, lower secondary, and upper secondary education levels; and (2) access to the Internet for pedagogical purposes at the same three educational levels. All indicators were systematized using data from the official website of the UNESCO Institute for Statistics (UIS) [2] and are expressed as percentages of schools.

2.2 Analytical Logic

This study focuses on the internal structure and symmetry of digital infrastructure within national education systems. Structural digital inequality is conceptualized as the uneven distribution of access to digital resources across different levels of general secondary education.

To operationalize this concept, inequality indices and composite indices of overall digital infrastructure were constructed. In particular, inequality in access to computers for pedagogical purposes is calculated as follows:

$$\text{Inequality}_{\text{Computer}} = \sqrt{\frac{\left(\text{Comp}_{\text{Pri}} - \overline{\text{Comp}}\right)^2 + \left(\text{Comp}_{\text{LowSec}} - \overline{\text{Comp}}\right)^2 + \left(\text{Comp}_{\text{UpSec}} - \overline{\text{Comp}}\right)^2}{3}} \tag{1}$$

where:

$$\overline{\text{Comp}} = \frac{\text{Comp}_{\text{Pri}} + \text{Comp}_{\text{LowSec}} + \text{Comp}_{\text{UpSec}}}{3} \tag{2}$$

$\overline{\text{Comp}}$ – Computer_School_Index.

Similarly, inequality in access to the Internet is calculated as follows:

$$\text{Inequality}_{\text{Internet}} = \sqrt{\frac{\left(\text{Int}_{\text{Pri}} - \overline{\text{Int}}\right)^2 + \left(\text{Int}_{\text{LowSec}} - \overline{\text{Int}}\right)^2 + \left(\text{Int}_{\text{UpSec}} - \overline{\text{Int}}\right)^2}{3}} \quad (3)$$

where:

$$\overline{\text{Int}} = \frac{\text{Int}_{\text{Pri}} + \text{Int}_{\text{LowSec}} + \text{Int}_{\text{UpSec}}}{3} \quad (4)$$

$\overline{\text{Int}}$ – Internet_School_Index.

A zero value of the inequality indices (Inequality_Computer, Inequality_Internet) indicates perfect structural equality, meaning uniform access across all education levels, whereas higher values reflect increasing asymmetry within the education system.

The mean value (DEI) of the composite indices of overall digital infrastructure (Computer_School_Index, Internet_School_Index) was used for descriptive purposes.

During the analysis of the results, it is important to test the following research ***hypotheses:***

H1. It is expected that statistically significant differences in access to digital infrastructure will be observed across different levels of education within national education systems.
H2. It is assumed that structural digital inequality is primarily concentrated at the level of primary education, where access to digital resources is significantly lower than at subsequent stages of education.
H3. It is expected that a high level of development of digital infrastructure at the national level does not preclude the existence of internal structural digital inequality within education systems.
H4. It is assumed that the scale and configuration of structural digital inequality will not differ substantially across countries of Eastern Europe, but will instead reflect country-specific institutional and policy contexts.

2.3 Statistical Methods

The empirical study design combined descriptive analysis, statistical methods, and within-subject comparisons. Descriptive statistics were calculated to assess mean levels of access and variability across different educational stages.

Separate repeated measures analyses of variance (ANOVA) were conducted for access to Internet and computers. Educational level (primary, lower secondary, upper secondary) was treated as a within-subject factor, allowing for a direct assessment of statistically significant differences between educational levels within a single country in a given year. When the assumption of sphericity was violated, Mauchly's test was applied. Effect sizes were reported using partial eta-squared ($\eta^2 p$) to evaluate the magnitude of structural inequalities. Additionally, a Bonferroni-adjusted post hoc test was performed to identify specific pairwise differences between educational levels. Data preprocessing and index construction for the creation of the dataset were carried out in Microsoft Excel, while subsequent statistical analyses were conducted using JASP.

3 Result

3.1 The Case of Eastern Europe

Eastern Europe provides a theoretically and empirically relevant context for studying digital education as a mechanism for developing modern human capital, as the region combines: (1) high coverage of basic school education, (2) ongoing digital transformation reforms, and (3) heterogeneous national capacities and political trajectories.

Cross-country differences in digitalization strategies indicate diverse models of human capital development through digital education, varying in terms of institutional coherence, scale of investment, and emphasis on equitable development of digital competencies. Specifically, Poland has implemented large-scale programs to equip schools with digital resources, including centrally funded initiatives aimed at strengthening ICT infrastructure and enhancing digital skills for both students and teachers (e.g., national programs operating from 2020–2024) [14].

The Czech Republic has adopted a structured approach to digital education through national strategies (e.g., Education Policy Strategy 2030) and long-term political roadmaps for an advanced digital education system [15].

Slovakia's approach aligns digital education transformation with European digital competence frameworks and policy documents. National initiatives emphasize capacity-building structures (e.g., Centers for Digital Technologies in Education) as part of a broader systemic transformation, providing an institutional basis for more equitable access across educational levels [16].

Hungary's digital education policy [17] focuses on promoting digital literacy at all education levels, integrating digital competencies into curricula, teacher development, and infrastructure modernization. International comparative monitoring [18] also highlights that improvements in digital skills can coexist with significant internal productivity gaps, suggesting that policy attention should go beyond aggregate digitalization levels and consider patterns of access and competence distribution.

Romania provides a notable example of policy explicitly aimed at building a digital ecosystem and reducing existing inequalities through the SMART.Edu strategy (2021–2027) [19]. At the same time, a PESTLE analysis by Constantin [20] identifies systemic implementation challenges, including infrastructure and connectivity limitations, particularly affecting rural and disadvantaged regions.

Bulgaria has undertaken targeted interventions to ensure equitable access to schooling under crisis conditions, including measures focused on device access and the development of basic digital skills for vulnerable groups [21].

In Moldova, the political framework links ICT in education to long-term development objectives (e.g., the implementation goals of the "Education 2030" Strategy, which include standards for equipment, teacher training in digital skills, and institutional capacity for digital tools) [22].

Ukraine presents a unique case due to the full-scale military context, which has accelerated reliance on remote and hybrid learning while highlighting infrastructure and equity issues. Official analytical reports emphasize the central role of equal access and quality in distance learning under wartime constraints, whereas peer-reviewed studies

examine how remote education practices evolved during the conflict and how digital technologies became a resilience mechanism.

3.2 Formation of Source Data

The empirical analysis is based on official internationally comparable statistical data corresponding to SDG Indicator 4.a.1, which measures access to information and communication technologies (ICT) for pedagogical purposes in schools.

For each Eastern European country during 2020–2024, data were collected separately for primary, lower secondary, and upper secondary education on access to educational technologies in schools, specifically: (1) the share of schools with access to computers for pedagogical purposes, and (2) the share of schools with access to the Internet for pedagogical purposes. All indicators are expressed as percentages. The data for 2025 were projected by the author and included only in the illustrative materials.

Based on these indicators, composite indices (Inequality_Computer, Inequality_Internet) were constructed. The dataset used for the calculations is presented in Appendix A. The analysis results are reported in Appendix B and visualized in Fig. 1.

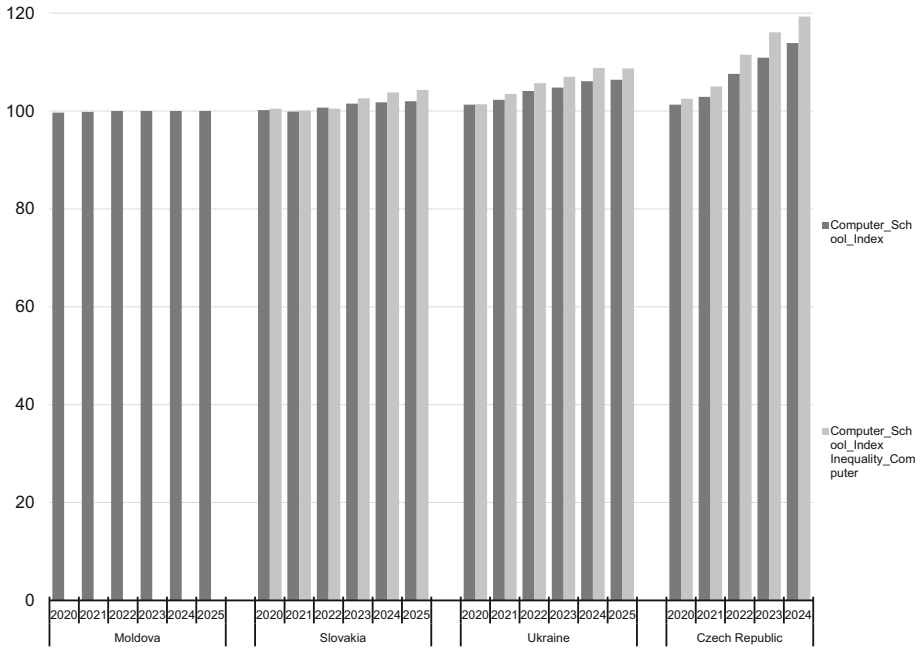


Fig. 1. Results of the composite access indices calculations [2]

A subsequent two-dimensional analysis will allow for a comparison of overall digital infrastructure levels across countries and an assessment of within-system structural inequality at different educational levels.

Table 1. Descriptive statistics of internet access by education level.

Education_Level	N	Mean	SD	SE	Coefficient of variation
Primary	24	95.54	7.146	1.459	0.075
LowerSec	24	99.31	1.693	0.346	0.017
UpperSec	24	99.08	2.060	0.421	0.021

Source: own research

Table 1 presents descriptive statistics for internet access across education levels aggregated over the sample.

The results of the descriptive statistics indicate a clear hierarchical pattern in the distribution of digital infrastructure across educational levels. The average level of Internet access in primary education is 95.54%, which is substantially lower than in lower secondary (99.31%) and upper secondary education (99.08%). The absolute gap between primary and lower secondary education reaches 3.77 percentage points, while the gap between primary and upper secondary education is 3.54 percentage points.

In addition to lower mean values, primary education exhibits considerably higher variance. The standard deviation at the primary level (SD = 7.15) is more than four times higher than at the lower secondary level (SD = 1.69), resulting in a coefficient of variation of 0.075 compared to 0.017 and 0.021 for lower and upper secondary education, respectively. This indicates that access to digital technologies at the primary level is not only lower on average but also significantly more uneven across countries and years.

These descriptive results already point to the presence of structural digital inequality (H2), concentrated at the earliest stage of education, which represents a critical bottleneck for the development of cumulative digital competencies.

3.3 Structural Inequality Between Education Levels: Repeated Measures ANOVA

To formally test whether differences between education levels are statistically significant, a repeated measures ANOVA were conducted (Table 2).

Table 2 Repeated Measures ANOVA (Within Subjects Effects)

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Education_Level	214.3[a]	2[a]	107.131[a]	12.324[a]	< 0.001[a]	0.381
Education_Level ✻ Country	240.1[a]	6[a]	40.019[a]	4.603[a]	0.001[a]	0.408
Residuals	347.7	40	8.693			

Note. Type III Sum of Squares;
[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < 0.05$).

The results demonstrate a statistically significant main effect of educational level on access to digital technologies ($F(2,40) = 12.324$, $p < 0.001$) with a large effect size ($\eta^2p = 0.381$). This indicates that approximately 38.1% of the within-system variance in access to digital technologies is explained solely by educational level, which represents a substantial magnitude in comparative education research.

Moreover, a significant interaction between educational level and country ($F(6,40) = 4.603$, $p = 0.001$, $\eta^2p = 0.408$) suggests that the structure and intensity of digital inequality differ markedly across national education systems. In other words, while structural inequality is a shared characteristic, its configuration is country-specific, reflecting differences in policy priorities and institutional capacity (H4).

A Repeated Measures ANOVA (Between-Subjects Effects) was then conducted. The results are presented in Table 3.

Table 3. Repeated Measures ANOVA (Between Subjects Effects)

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Country	295.7	3	98.56	4.335	0.017	0.394
Residuals	454.7	20	22.74			

Note. Type III Sum of Squares

The between-subjects analysis demonstrates statistically significant differences between countries in overall digital infrastructure levels ($F(3,20) = 4.335$, $p = 0.017$, $\eta^2p = 0.394$). This indicates that nearly 39.4% of the variance in average access to digital technologies can be attributed to national-level differences, confirming that digital education infrastructure remains strongly influenced by country-specific conditions in Eastern Europe.

In other words, the coexistence of substantial within- and between-subject effects suggests that high national averages do not preclude the presence of within-system structural inequality (H1). This underscores the need to examine the distribution within the system rather than relying solely on aggregate indicators.

3.4 Pairwise Comparisons Between Education Levels

Post hoc tests provide a more granular understanding of where structural inequality is concentrated. To identify which education levels, differ significantly, Bonferroni-adjusted post hoc tests were conducted (Table 4).

The difference between primary and lower secondary education is statistically significant (Mean Difference $= -3.77$, $p = 0.002$, Cohen's $d = -0.86$), as is the difference between primary and upper secondary education (Mean Difference $= -3.54$, $p = 0.004$, Cohen's $d = -0.80$). Both effect sizes fall within the large effect range, indicating substantial and meaningful disparities.

In contrast, the comparison between lower and upper secondary education is not statistically significant (Mean Difference $= 0.23$, $p = 1.000$), suggesting that digital infrastructure is relatively balanced across the upper levels of secondary education.

Table 4. Post hoc comparisons between education levels (Bonferroni-adjusted).

			95% CI for Mean Difference						95% CI for Cohen's d		
		Mean Difference	Lower	Upper	SE	df	t	Cohen's d	Lower	Upper	p_{bonf}
Primary	Lower Sec	−3.768	−6.332	−1.204	1.032	46	−3.652	−0.856	−1.512	−0.199	0.002**
	Upper Sec	−3.540	−6.104	−0.976	1.032	46	−3.430	−0.804	−1.452	−0.156	0.004**
Lower Sec	Upper Sec	0.228	−2.336	2.792	1.032	46	0.221	0.052	−0.527	0.630	1.000

Thus, structural digital inequality is unevenly distributed across the education system and is sharply concentrated at the primary education level (H3), where access lags behind subsequent stages.

3.5 Digital Access and the Formation of Human Capital

The results of the analysis are consistent with the findings of Onopriienko et al. [13], which emphasize the potential of digital technologies to enhance human capital development. However, they add an important qualification: the positive impact of digitalization is not automatic and largely depends on the structural organization of access to technologies. The high variance in Internet access at the primary education level (SD = 7.15; coefficient of variation = 0.075), compared to lower variance at higher educational levels (SD = 1.69 and 2.06, respectively), indicates uneven digital opportunities at early stages of education, limiting the cumulative development of digital competencies.

The results of the Repeated Measures ANOVA statistically confirm the presence of structural digital inequality across educational levels. A significant main effect of educational level ($F(2,40) = 12.324$, $p < 0.001$, $\eta^2 p = 0.381$) indicates that over 38% of the within-system variance in access to digital technologies is explained solely by educational level. These results empirically confirm H1, demonstrating the presence of statistically significant differences across levels of education. Furthermore, they empirically support the claim advanced by Sairmaly [3] regarding the key role of digital technologies in the development of human capital. At the same time, our results refine his conclusions by highlighting that the effect of digital technologies is structurally differentiated within education systems.

3.6 Threshold Effects of Early Digital Exclusion in the Context of Lifelong Learning

The results of the post hoc tests are particularly illustrative, showing that statistically significant differences are concentrated between primary and subsequent educational levels (Primary/Lower Secondary: Mean Difference = −3.77, p = 0.002, Cohen's d = −0.86; Primary/Upper Secondary: Mean Difference = −3.54, p = 0.004, Cohen's d = −0.80), whereas no significant differences were found between lower and upper

secondary education. These results confirm H2, indicating that structural inequality is primarily concentrated at the level of primary education. In other words, structural digital inequality is not linear but threshold-based, being concentrated at the earliest stage of education.

From the perspective of the lifelong learning framework [9, 10], these results are critically important. Since lifelong learning assumes the accumulation of educational and human capital throughout the life cycle, limited access to digital resources at the primary education level creates a long-term gap in development opportunities, which cannot be fully compensated at later stages.

The observed cross-country differences ($F(3,20) = 4.335$, $p = 0.017$, $\eta^2 p = 0.394$) are consistent with the findings of Rykovska [10] regarding the role of national institutional and political factors in human capital development. At the same time, a significant interaction between educational level and country ($\eta^2 p = 0.408$) indicates that structural digital inequality is not a universal phenomenon but a context-dependent one, reflecting diverse models of educational policy and digital transformation priorities across Eastern European countries.

3.7 Conclusion

Eastern Europe combines (1) politically driven digitalization programs, (2) heterogeneous institutional capacity, and (3) varying impacts of crises and reforms. This makes the region particularly well-suited for moving from a perspective of "digitalization as access" to a structural perspective, examining whether digital infrastructure is distributed symmetrically across all educational levels – a question directly linked to equitable opportunities for accumulating digital competencies and developing modern human capital. The study's results confirm the presence of within-system structural digital inequality in the education systems of Eastern European countries and highlight its concentration at the primary education level, a critical stage for the cumulative development of digital competencies. These results highlight the importance of targeted policy interventions aimed at ensuring a balanced digital infrastructure across all levels of education, particularly at the initial stage. Furthermore, the study emphasizes the need for methodological consistency grounded in observed data, which enhances the robustness and reliability of the empirical findings. A key practical implication is that investments in educational digitalization should be evaluated not only by coverage but also by the structural balance of access across educational stages. Therefore, prioritizing the alignment of digital infrastructure at the primary school level is essential for the inclusive development of modern human capital. The obtained results have important policy implications, highlighting the need to prioritize improvements in digital infrastructure at the early stages of education. Addressing inequality at the primary level is crucial for ensuring the overall development of digital competencies and reducing long-term disparities in human capital formation across countries.

Appendix

A. Source Data

Country	Period	Comp_Primary	Comp_LowerSec	Comp_UpperSec	Internet_Primary	Internet_LowerSec	Internet_UpperSec
Bulgaria	2021	68.72					
	2023	72.5					
Moldova	2020	99.26	99.73	100	92.49	94.41	97.96
	2021	99.76	100	99.74	94.47	95.73	97.96
	2022	100	100	100	96.96	97.41	99.24
	2023	100	100	100	97.01	97.75	99.48
	2024	100	100	100	98.07	99.09	99.74
Poland	2020	100	100	100	100	100	100
	2021	100	100	100	100	100	100
	2023	97.26					
Romania	2024	100					
Slovakia	2020	100	100	100	100	100	100
	2021	100	100	100	100	100	100
	2022	100	100	100	100	100	100
	2023	100	100	100	100	100	100
	2024	100	100	100	100	100	100
Hungary	2023	91.93					
Ukraine	2020	96.86	99.8	99.84	93.27	100	100
	2021	98.93	99.72	99.9	100	100	100
	2023	86.02	96.22	95.26	98.99	98.53	97.56
	2024	88.24	95.1	94.09	99.44	99.97	97.64
Czech Republic	2020	98.31	99.94	99.85	98.17	99.94	99.77
	2021	98.23	99.71	99.69	98.11	99.71	99.54
	2022	98.76	99.78	99.54	98.71	99.72	99.39
	2023	85.08					

B. Results of Calculations

Country	Period	Comp_Primary	Comp_LowerSec	Comp_UpperSec	Computer_School_Index**	Inequality_Computer**
Moldova	2020	99.26	99.73	100	99.66333333	0.003757424
	2021	99.76	100	99.74	99.83333333	0.001449251
	2022	100	100	100	100	0
	2023	100	100	100	100	0
	2024	100	100	100	100	0
	2025*	*100*	*100*	*100*	100	0
Slovakia	2020	100	100	100	100	0
	2021	100	100	100	100	0
	2022	100	100	100	100	0
	2023	100	100	100	100	0
	2024	100	100	100	100	0

(*continued*)

(*continued*)

Country	Period	Comp_Primary	Comp_LowerSec	Comp_UpperSec	Computer_School_Index**	Inequality_Computer**
	2025*	*100*	*100*	*100*	100	0
Ukraine	2020	96.86	99.8	99.84	98.83333333	0.017292484
	2021	98.93	99.72	99.9	99.51666667	0.00518484
	2022*	*100*	*99.64*	*99.96*	99.86666667	0.001975788
	2023	86.02	96.22	95.26	92.5	0.06089011
	2024	88.24	95.1	94.09	92.47666667	0.040049577
	2025*	*81.43262*	*93.7817*	*92.36596*	89.19342662	0.075770509
Czech Republic	2020	98.31	99.94	99.85	99.36666667	0.009220456
	2021	98.23	99.71	99.69	99.21	0.008555224
	2022	98.76	99.78	99.54	99.36	0.005367267
	2023	85.08	100	*99.3837*	94.821236	0.089028409
	2024*	*83.57869*	*99.98071*	*99.22862*	94.26267347	0.098238674
	2025*	*78.5027*	*100*	*99.07321*	92.52530236	0.131345337

Country	Period	Internet_Primary	Internet_LowerSec	Internet_UpperSec	Internet_School_Index	Inequality_Internet
Moldova	2020	92.49	94.41	97.96	94.95333333	0.029226796
	2021	94.47	95.73	97.96	96.05333333	0.018399401
	2022	96.96	97.41	99.24	97.87	0.012338826
	2023	97.01	97.75	99.48	98.08	0.012924506
	2024	98.07	99.09	99.74	98.96666667	0.008505931
	2025*	*99.6367*	*100*	*100*	99.87890446	0.002099979
Slovakia	2020	100	100	100	100	0
	2021	100	100	100	100	0
	2022	100	100	100	100	0
	2023	100	100	100	100	0
	2024	100	100	100	100	0
	2025*	*100*	*100*	*100*	100	0
Ukraine	2020	93.27	100	100	97.75666667	0.039747338
	2021	100	100	100	100	0
	2022*	*100*	*100*	*100*	100	0
	2023	98.99	98.53	97.56	98.36	0.007421716
	2024	99.44	99.97	97.64	99.01666667	0.012334534
	2025*	*100*	*98.34685*	*96.31401*	98.22028759	0.018797039
Czech Republic	2020	98.17	99.94	99.77	99.29333333	0.009834915
	2021	98.11	99.71	99.54	99.12	0.008866082
	2022	98.71	99.72	99.39	99.27333333	0.005187779
	2023*	*98.8947*	*99.579*	*99.1897*	99.22112816	0.003459095
	2024*	*99.1592*	*99.467*	*98.999*	99.20840424	0.002397392
	2025*	99.4016	99.34694	98.80565	99.18472995	0.003321376

Note. * predicted by the author; ** calculated by the author

References

1. Our World in Data: Human Capital Index. World Bank (2018). https://ourworldindata.org/grapher/human-capital-index-in-2018
2. UNESCO Institute for Statistics: UIS data browser. UNESCO (2026). https://databrowser.uis.unesco.org/
3. Sairmaly, F.A.: Human capital development and economic growth: a literature review on information technology investment, education, skills, and productive labour. J. Minfo Polgan. **12**(1), 679–693 (2023). https://doi.org/10.33395/jmp.v12i1.12491
4. Bahri, E.S.: The identification of competency-based human resource development strategy. Int. J. Strateg. Stud. **2**(1), 1–7 (2025). https://doi.org/10.59921/icestra.v2i1.50
5. Shatila, K., Aránega, A.Y., Soga, L.R., Hernández-Lara, A.B.: Digital literacy, digital accessibility, human capital, and entrepreneurial resilience: a case for dynamic business ecosystems. J. Innov. Knowl. **10**(3), 100709 (2025). https://doi.org/10.1016/j.jik.2025.100709
6. Nurwendari, W., Yusrizal, Y., Musnadi, S., Yunus, M.: Development of online video-based learning media as an alternative for accounting education: a conceptual approach. J. Educ. Cult. Soc. **16**(2), 461–482 (2025). https://doi.org/10.15503/jecs2025.3.461.482
7. Savitska, V., Matveieva, N., Chinchoy, A., Romanyshyna, N., Sakh, Y.: The effectiveness of multimedia boards/projection systems in teaching humanities. Rev. Eduweb. **19**(2), 234–251 (2025). https://doi.org/10.46502/issn.1856-7576/2025.19.02.16
8. Alam, S., Hameed, A., Balogh, Z., Haque, S.A.: Digitalisation of education: challenges and issues in virtual pedagogical competence. J. Educ. Cult. Soc. **16**(1), 165–177 (2025). https://doi.org/10.15503/jecs2025.2.165.177
9. Markozova, O.O.: Learning for life –a prerequisite for achieving life success. Visn. Int. Slav. Un-tu. Kharkiv. Ser. Sociol. Sci. **16**(1), 5–13 (2013)
10. Rykovska, O.V.: Adult education as a factor in human capital development: European experience and Ukrainian trends. Econ. Sci. **22**, 96–101 (2016)
11. Hendricks, L., Schoellman, T.: Human capital and development accounting: new evidence from wage gains at migration. Q. J. Econ. **133**(2), 665–700 (2017)
12. Ryabchuk, O.G., Gozhulovsky, S.S.: Peculiarities of human capital development as the main factor of economic growth in Ukraine. Econ. Bull. Ser. Finan. Account. Taxat. **4**, 182–192 (2020). https://doi.org/10.33244/2617-5932.4.2020.182-192
13. Onopriienko, K., Duranowski, W., Petrushenko, Y., Artyukhov, A.: Development of regions, industries and types of economic activity. Bull. Cherkasy Natl. Univ. Econ. Sci. **1**, 49–56 (2021). https://doi.org/10.31651/2076-5843-2021-1-49-56
14. ReferNet Polska: Digital education strategy and policy overview (Poland) (2025). https://www.refernet.de/dokumente/pdf/2025_SD_PL.pdf
15. European Commission: Czech Republic—education policy strategy 2030 (Digital Skills and Jobs Coalition). https://digital-skills-jobs.europa.eu/en/actions/national-initiatives/national-strategies/czech-republic-education-policy-strategy-2030
16. Ftáčnik, M., Šveda, D., Kireš, M.: Digital transformation of education in Slovakia within the context of European documents. In: Proceedings of the 2020 International Conference on Emerging eLearning Technologies and Applications (ICETA). IEEE (2020). https://doi.org/10.1109/ICETA51985.2020.9379154
17. European Commission: Hungary—digital education strategy (Digital Skills and Jobs Coalition) (2024). https://digital-skills-jobs.europa.eu/en/actions/national-initiatives/national-strategies/hungary-digital-education-strategy
18. European Commission: Education and training monitor 2025: Hungary country report (2025). https://op.europa.eu/webpub/eac/education-and-training-monitor/en/country-reports/hungary.html

19. DiploFoundation & Geneva Internet Platform: Strategy for the digitalization of education in Romania 2021–2027 (2021). https://dig.watch/resource/strategy-for-the-digitalization-of-education-in-romania-2021-2027
20. Constantin, A.: Digitalization of Romania's education sector: a PESTLE analysis. In: Proceedings of the International Conference on Business Excellence, vol. 18. Bucharest University of Economic Studies (2024). https://doi.org/10.2478/picbe-2024-0009
21. Lesková, A., Lenghart, P.: Post-Covid media behaviour patterns of the generation Z members in Slovakia. J. Educ. Cult. Soc. **14**(1), 503–513 (2023). https://doi.org/10.15503/jecs2023.1.503.513
22. Ministry of Education and Research of the Republic of Moldova: Standards and ICT in education: country framework and policy (2023). https://mded.gov.md/wp-content/uploads/2023/11/STD_EN.pdf

Mapping Citizen Science Platforms for Environmental Education in Slovakia

Katarína Skokanová(✉) and Imrich Jakab

Katedra ekológie a environmentalistiky, FPVaI UKF, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia
katarina.skokanova@ukf.sk

Abstract. Citizen science represents a promising approach to address persistent challenges in Slovak environmental education, particularly low student engagement and the disconnect between theory and practice. Despite the 2023 curriculum reform elevating environmental education to cross-curricular literacy, systematic implementation tools remain scarce. This study develops a comprehensive catalogue of citizen science platforms for formal education, presented in tabular and interactive flashcard formats. The catalogue evaluates mobile applications and web platforms based on cross-subject applicability, language accessibility, and pedagogical potential. Analysis of the current state of citizen science in Slovakia reveals a significant gap: while citizen science projects exist, teachers lack accessible resources for classroom implementation. Although involving students as citizen scientists has great yet not well discovered potential for developing environmental responsibility and scientific literacy. The platform catalogue addresses this need by providing structured guidance for educators. It will be disseminated through CVTI SR website and the Špirála environmental education network in Slovak schools.

Keywords: Mobile Applications · Citizen Science · Third Cycle · Smartphones · Environmental Literacy · Educational Platforms · Project-Based Learning · Teacher Resources · Teacher Professional Development · Slovakia

1 Environmental Education in Slovakia

The need to systematically address environmental education (EE) on a global scale was first officially emphasized at the 1977 Intergovernmental Conference on Environmental Education in Tbilisi, Georgia. There, environmental education was defined as "a process aimed at helping individuals become more aware of their environment and develop responsible environmental behaviour, as well as the skills necessary to improve the quality of their environment" [1].

In Slovakia, environmental education has been incorporated into the formal education system since 1996, when cross-curricular guidelines for environmental education came into effect [2]. Currently, it is defined as a cross-curricular theme [3, 4] aimed at developing students' relationship with nature and the environment, building environmental awareness, and promoting responsible behaviour towards the environment. Following the 2023 curriculum reform for elementary education [5], environmental education was

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 337–348, 2027.
https://doi.org/10.1007/978-3-032-32104-6_29

transformed into environmental literacy, integrating EE standards directly into individual educational fields. This approach assumes that students will acquire environmental knowledge not in an isolated subject, but across the entire curriculum through authentic situations and practical activities.

In the Slovak education system, non-formal education plays a crucial complementary role to the formal system. Non-governmental organizations dominate the implementation of projects, programmes, and methodological support [6]. The third sector is more effective in working with experiential learning than the formal system and better fosters a personal relationship with environmental issues.

The most recent analysis of the state of environmental education in Slovakia, conducted by Bodáczová et al. [7] under the title "Čo vás v tej škole učia" ("What They Teach You in That School"), identified problems and solutions in three key areas: (a) establishing a concept for formal EE, (b) more targeted and consistent integration of EE content, and (c) improving teacher preparation for EE. Unsatisfactory and outdated coverage of environmental education and awareness (EEA) is also highlighted in documents from the Slovak Environmental Agency [8] and the Regional Concept of EEA in Trenčín Region [9] issued by the Ministry of the Interior of the Slovak Republic.

Low levels of active engagement among students and the public in environmental issues remain one of the main problems of the Slovak education system. People do not participate without intrinsic meaning and motivation, underscoring the need for new methods that accommodate diverse motivations – from volunteering and purpose to the joy of action. One such method is citizen science, introduced in the following section.

2 Citizen Science

Citizen science (CS), also referred to as community-based monitoring, involves scientific work carried out by members of the general public – termed citizen scientists – in collaboration with professional scientists and expert institutions [10–12]. Participants contribute to various stages of research, thereby expanding scientific knowledge and scientific literacy [13–15]. Citizen science thus bridges the academic scientific community with a broad spectrum of everyday environmental challenges. Scientists gain large-scale field data, while the public receives opportunities to actively engage in addressing issues that directly affect them [16]. According to ECSA [17], citizen science involves the public in the scientific process in various ways – from data collection to co-formulation of research questions and participation is typically unpaid voluntary work.

The most common form of public involvement is the collection of diverse research data over relatively short periods, with modern technologies, particularly mobile phones, playing a significant role [18]. These tools are integrated into numerous educational programmes worldwide. Implementing these platforms into the learning process yields multiple benefits: students gain practical experience with scientific methods and contribute to real research [15], develop digital skills [19], build connections to their local environment, and experience increased motivation and sense of purpose in learning [20].

Traditional scientific research often encounters limitations in data scale and frequency of data collection, which CS effectively overcomes [21]. Citizen science mostly follows a bottom-up approach, where greater involvement yields not only more data but

also higher societal legitimacy of the research [22]. Topics supported by community-collected data enter public and political discourse more rapidly, having stronger influence than data from closed laboratories [22]. Transforming environmental policy requires thorough mapping of the current state – only by clearly understanding point A we can meaningfully define and reach point B. Citizen science thus provides a critical evidence base for informed environmental policy and decision-making [23].

Citizen science functions as: (1) a tool for engaging the public in real research, (2) a means of building a relationship with nature through personal experience, and (3) a way to "experience problems" rather than merely "learn about problems" [24]. For students, it is crucial that CS allows them to move beyond abstract issues and become part of the solution – contributing to scientific knowledge with concrete purpose and meaning [25].

The vision of positive change offered by CS projects serves as a powerful source of intrinsic motivation [26]. These projects provide benefits beyond scientific value, including contributions to environmental protection [27]. Research also indicates that interest and perceived self-efficacy are key predictors of pro-environmental behaviour [28], making CS a powerful tool not only for scientific education but also for building environmental responsibility.

Citizen science as a pedagogical approach remains a relatively new and unfamiliar area for many teachers, with low awareness of available educational tools [29]. Yet citizen science is more than a data-collection method – it is a pathway for students to explore their living environment and contribute meaningfully to it [30]. Certain studies emphasize that meaningful outcomes require students to go beyond mere data collection and actively analyze and interpret information [31, 32].

Implementation of citizen science in education also has an important dimension often overlooked in traditional research – the emotional dimension of children's engagement. While (adult) scientists approach research primarily rationally and analytically, children experience the entire process of data collection and subsequent changes much more intensely. Research shows that emotional engagement is key to children's success in school [33]. When children successfully complete data collection leading to tangible change – such as protecting a critical habitat, reducing air pollution, or creating green infrastructure at school – this change extends beyond the scientific community. Another research [34] demonstrated that children transfer environmental knowledge to their parents, measurably changing household behaviour.

3 Methodology: Availability of Citizen Science in Slovakia

This article builds on the authors' previous work [35], which focused on selecting mobile CS applications specifically for teaching biodiversity. The current study expands this approach by developing a comprehensive catalogue of citizen science platforms – encompassing both mobile and web-based applications – organized by subject areas rather than specific biodiversity topics.

This shift reflects two key insights: first, modern web applications offer tablet-friendly, intuitive interfaces suitable for field use, making the mobile-only restriction unnecessary; second, teachers in the third cycle of elementary education need a broad, subject-based overview as a starting point before developing topic-specific materials.

Although some applications are available only in English, however younger generations – often referred to as "digital natives" - quickly adapt to English interfaces due to their experience with mobile technologies [36]. This creates an opportunity to support teachers: while students handle the technological aspects relatively easily, teachers require methodological guidance for integrating these tools into teaching.

The need to compile a list of platforms – rather than merely a list of projects – became evident during the author's internship at Slovak Centre of Scientific and Technical Information (CVTI SR). CVTI SR is the leading institution supporting citizen science in Slovakia, offering a list of CS projects in SR, educational platform, online webinars, and various publications. For teachers, selecting directly from CS platforms is more appealing, as students perceive the use of mobile applications in school as attractive and motivating.

Another important source confirming the need for such a list was a pilot quantitative descriptive survey carried out from May–December 2025 via an online questionnaire with preliminary results based on 95 responses. The survey targeted primary and secondary school teachers of all subjects and school types across Slovakia. Using purposive sampling from all Slovak schools [37, 38], the questionnaire included 47 questions (20 closed, 20 semi-open, 10 open) across 5 thematic sections: respondent background, CS awareness, basic CS experience, opinions on CS, and expectations, reduced to 19 for those without prior CS experience. A parallel questionnaire for expert institutions to map school-institution collaborations in CS was also developed, though it is beyond the scope of this paper. Despite outreach through municipal and regional school authorities, the response rate reached only 7.4%.

The survey revealed a paradox: while only 10 teachers had practical experience with CS and 66 had no overview of available projects, the same number 66 expressed interest in receiving information about CS programmes. This indicates that the limited usage of CS is not due to lack of interest but rather to insufficient information and systematic support.

The compiled platform list will, upon publication, be made available for use in both formal and non-formal education. In non-formal education, results will be disseminated via the CVTI SR portal (otvorenaveda.cvtisr.sk). In formal education, the selection was presented at the Špirála network board meeting on 30 January 2026; the board approved dissemination through the Špirála newsletter, social media, and presentation at the general assembly in Žilina on 12–13 March 2026, thereby making it accessible to a wide teacher community. Špirála is a nationwide network of 15 organizations (e.g., CEA, ŽIVICA, DAPHNE, SOSNA, and others) and 14 individual members dedicated to EEA [39].

4 Results: Practical Implementation of CS in School Settings

Mobile applications and web platforms for citizen science have become widely available tools for educational purposes. Applications such as iNaturalist, eBird, and Pl@ntNet utilize GPS technology, artificial intelligence for species identification, and cloud storage, enabling even novice citizen scientists to collect high-quality scientific data directly [40]. Citizen science is not limited to biodiversity monitoring; it encompasses meteorology, astronomy, history, sociology, and urban planning etc.

The main output of this study is: Selection of CS freely available platforms covering a wide spectrum of subject areas, primarily intended for the 3rd cycle of elementary education - which we consider the most suitable age for involvement in CS. Platforms were selected based on three criteria: alignment with educational standards, application functionality, and usability in the area of Slovakia. To ensure thematic diversity, one to two platforms were selected per subject area. From 42 platforms considered, 27 were selected. The list was created in two formats tailored to teachers' needs:

1. **Tabular version** – designed for printing and offline use (see Table 1). The table lists the platform name, type (mobile application/web platform), citizen science domain, possible applications across school subjects, and language.
2. **Interactive web version** – designed as flip flashcards (see Fig. 1). Each card displays the platform logo, relevant domain, and subject on the front, the back provides a detailed description, usage possibilities, download links for iOS and Android, and language information. This format was chosen to increase attractiveness and usability for practicing teachers (https://view.genially.com/68b07ff594d891d96607ffaf).

Fig. 1. Preview of the interactive web version of the citizen science platform selection

Table 1. Overview of freely available citizen science applications suitable for the third cycle of elementary education.

Platform Name	Platform Type	Citizen Science Domain	School Subjects	Lang--uage
iNaturalist	mobile app	Biodiversity	Biology, Natural Science	SK
EyeOnWater	mobile app	Water quality	Chemistry, Geography	EN
Invasive Species	mobile app / web platform	Invasive species	Biology, Geography	EN
GLOBE Observer	mobile app	Earth observation	Geography, Physics, Biology	EN
Hush City	mobile app	Noise pollution	Physics, Geography, Civics	EN
Loss of the Night	mobile app	Light pollution	Physics, Geography	EN

(*continued*)

Table 1. (*continued*)

EBMS	mobile app / web platform	Butterfly monitoring	Biology	EN
CSI-COP: Cookie, pozor!	web platform	Food and health	Chemistry, Biology, Health Education	CZ
Seislok	mobile app	Seismology	Physics, Geography	CZ
TrashOut	mobile app	Waste management	Geography, Civics	SK
Historypin	web platform	Local history	History, Geography, Civics	EN
MapSwipe	mobile app	Territory mapping	Geography, Informatics	CZ
Chronolog	digital tool + web	Landscape monitoring	Biology, Geography	EN
Agroecology Map	web platform	Agroecology	Biology, Geography	EN
Senzor kvality ovzdušia	physical sensor + web	Air quality	Physics, Chemistry, Geography	CZ
Foldit	online game	Molecular biology	Biology, Informatics	CZ
INCREASE	mobile app + experiments	Agro- - biodiversity	Biology	EN
Mushroom Fans	web platform	Fungi and biodiversity	Biology, Natural Science	SK
Spot-A-Bee	mobile app	Pollinators	Biology	EN
Zastav sršňa	mobile app	Invasive species	Biology	SK

(*continued*)

Table 1. (*continued*)

Bicizen	mobile app	Mobility	Geography, Civics	EN
Lesodiverzita	mobile app	Forest ecosystems	Biology	CZ
Merlin Bird ID	mobile app	Birds	Biology, Natural Science	EN
Echo Bat Detector	mobile app / device	Bioacoustics	Biology, Physics	EN
RadonEye	physical sensor	Radioactivity	Physics, Chemistry	EN
Tea Bag Index	mobile app	Organic matter decomposition	Biology, Chemistry	EN
PlantNet	mobile app	Plant identification	Biology, Natural Science	SK

While technological tools are readily available, successful integration of citizen science into teaching requires systematic teacher support, as discussed below.

5 Discussion: Interest, Barriers, and Need for Support

Teachers are key actors in implementing innovative educational methods, including CS in school practice [41]. International research shows that many teachers already conduct activities that could be classified as CS but are often unaware that their work falls into this category [42]. In Slovakia, this phenomenon is illustrated by the oldest CS programme: the White Stork monitoring coordinated since 1976 by RNDr. Miroslav Foltín (Nature Protection School Varín), who retrospectively identified his long-term work as an example of citizen science (otvorenaveda.cvtisr.sk).

In recent years, several public-involvement research projects have emerged in Slovakia, yet CS still lacks the systemic support it deserves, given its potential for science and society. Although tools and projects exist, CS is not systematically embedded in the curriculum or routine teacher's practice. Despite the mentioned benefits, teachers face multiple barriers when integrating citizen science into regular lessons. Lack of time is the most frequently cited obstacle – teachers perceive citizen science more as suitable for extracurricular clubs than standard classroom instruction.

The author's own research highlighted the following main barriers: lack of time (70% of responses), limited funding, and inadequate teacher training. Teachers emphasized the need for methodological materials, introductory workshops, and systematic support

from NGOs. While this concern is partly justified, international experience shows that integrating citizen science into the curriculum is both feasible and effective [43–45].

The key to successful implementation lies in aligning CS projects with existing educational standards [46, 47]. American studies demonstrate successful integration across school subjects, e.g., natural science (data collection), mathematics (data processing), native language (report writing), and geography (mapping), across the entire curriculum [14, 43]. This cross-curricular nature makes citizen science naturally suited to project-based learning, where a single long-term project can integrate various subjects and teaching forms in different phases.

Research in environmental education unequivocally confirms that project-based learning incorporating citizen science is among the most effective tools for developing environmental awareness [48]. The reason is clear: PBL engages emotions, authentic experiences, and real problems, leading not only to better cognitive outcomes but also to higher motivation and deeper internalization of environmental values [49–51]. Investment in citizen science thus returns to teachers in a form of meaningful learning process. Further studies confirm that PBL increases motivation and self-efficacy – students' belief in their own abilities [52].

6 Conclusion

Citizen science is offered but not embedded in the curriculum or routine teachers' practice. The topic of citizen science was presented at the Špirála association assembly, indicating growing interest from environmental education organizations. School environmental coordinators represent a potential group for implementing citizen science, yet they require adequate support and methodological materials.

For successful implementation, teachers need systematic support, as citizen science is not merely data collection – it is a scientific method with rigorous procedures ensuring result quality and reliability [53]. This support should include training in scientific methods, mobile applications handling, and student engagement methodology [54]. Motivation for developing such materials should stem from teachers' needs and real practice in the Slovak context. Equally important is creating spaces for teachers to share experiences and inspire one another with best practices.

NGOs and scientific institutions, with their expertise and practical experience, can play a key role in providing this support [55]. Collaboration between schools and external experts not only enhances educational quality but also brings prestige to the school and creates opportunities for long-term partnerships between educational institutions and the scientific community [56, 57]. Contact with real scientists and experts boosts students' self-confidence and sustains long-term interest in scientific topics [58].

Citizen science offers significant potential as a complement to environmental education in Slovakia. It provides an opportunity to overcome long-standing issues in the formal system – low active engagement, lack of personal connection to environmental themes, and weak linkage between theory and practice. Citizen science also builds a bridge between the academic research community and the public, connecting expert knowledge with everyday problems and subjective, emotional perceptions of issues and hopes.

Acknowledgements. This article was written within the KEGA project 037UKF-4/2025.

References

1. UNESCO: Tbilisi declaration. Intergovernmental Conference on Environmental Education, Tbilisi (1977)
2. Čižmárová, M.: Environmentálna výchova a vzdelávanie detí a mládeže – aktuálna situácia na Slovensku. Výskum mládeže (2012)
3. Ministerstvo školstva SR: Štátny vzdelávací program pre základné vzdelávanie. Bratislava (2008)
4. Ministerstvo školstva SR: Štátny vzdelávací program pre základné vzdelávanie. Inovovaná verzia. Bratislava (2015)
5. Ministerstvo školstva SR: Štátny vzdelávací program pre základné vzdelávanie. Bratislava (2023)
6. Ministerstvo životného prostredia SR: Rezortná koncepcia environmentálnej výchovy, vzdelávania a osvety do roku 2025. Bratislava (2015)
7. Bodáczová, B., et al.: Čo vás v tej škole učia. Analýza environmentálneho vzdelávania na Slovensku (2021)
8. SAŽP: Koncepcia rozvoja environmentálnej výchovy, vzdelávania a osvety v SAŽP do roku 2030. Slovenská agentúra životného prostredia (2020)
9. Medal, P., et al.: Krajská koncepcia environmentálnej výchovy, vzdelávania a osvety v Trenčianskom kraji. Ministerstvo vnútra SR (2019)
10. Irwin, A.: Citizen Science: A Study of People, Expertise and Sustainable Development. Routledge (1995)
11. Heigl, F., Kieslinger, B., Paul, K.T., Uhlik, J., Dörler, D.: Toward an international definition of citizen science. Proc. Natl. Acad. Sci. **116**(17), 8089–8092 (2019). https://doi.org/10.1073/pnas.1903393116
12. European Commission: Citizen science: elevating research and innovation through societal engagement. DG RTD, European Commission (2020)
13. Bonney, R., et al.: Citizen science: a developing tool for expanding science knowledge and scientific literacy. Bioscience. **59**(11), 977–984 (2009). https://doi.org/10.1525/bio.2009.59.11.9
14. Bonney, R., Phillips, T.B., Ballard, H.L., Enck, J.W.: Can citizen science enhance public understanding of science? Public Underst. Sci. **25**(1), 2–16 (2016). https://doi.org/10.1177/0963662515607406
15. Phillips, T.B., Ballard, H.L., Lewenstein, B.V., Bonney, R.: Engagement in science through citizen science: moving beyond data collection. Sci. Educ. **103**(3), 665–690 (2019). https://doi.org/10.1002/sce.21501
16. Hidalgo, E., Perelló, J., Becker, F., Bonhoure, I., Legris, M., Cigarini, A.: Participation and co-creation in citizen science. In: The Science of Citizen Science. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58278-4_11
17. ECSA: 10 Principles of CS. European Citizen Science Association (2015)
18. Graham, H., et al.: The experiences of everyday participation: comparing two typologies. J. Civ. Soc. **7**(4), 297–316 (2011)
19. Aristeidou, M., Herodotou, C.: Online citizen science: a systematic review of effects on learning and scientific literacy. Citiz. Sci. Theory Pract. **5**(1), 1–12 (2020). https://doi.org/10.5334/cstp.224

20. Herodotou, C., Aristeidou, M., Miller, G., Ballard, H.: Youngsters as citizen scientists: exploring engagement with science and learning outcomes. Citiz. Sci. Theory Pract. **5**(1), 2 (2020). https://doi.org/10.5334/cstp.248
21. IUCN: Citizen Science—Resource. International Union for Conservation of Nature (2025). https://iucn.org/resources/explainer-brief/citizen-science
22. OECD: What is citizen science and why should policymakers care? OECD Policy Briefs, Paris (2025)
23. de Sherbinin, A., et al.: The critical importance of citizen science data. Front. Clim. **3**, 650760 (2021)
24. US Forest Service: Citizen science. United States Department of Agriculture (2024). https://www.fs.usda.gov/working-with-us/citizen-science
25. Ballard, H.L., et al.: Contributions to conservation outcomes by natural history museum-led citizen science. Biol. Conserv. **208**, 87–97 (2017)
26. RAND Corporation: Citizen science helps communities participate in, understand, and trust science. Santa Monica (2022)
27. McKinley, D.C., et al.: Citizen science can improve conservation science, natural resource management, and environmental protection. Biol. Conserv. **208**, 15–28 (2016)
28. Szczytko, R., et al.: Connecting outdoor science education to community through citizen science. J. Experiential Educ. **42**(2), 107–121 (2019)
29. Phillips, T., et al.: Engagement in science through citizen science. Bioscience. **69**(2), 112–123 (2018)
30. Herodotou, C., et al.: Youth-focused citizen science: examining the role of environmental science learning and agency for conservation. Biol. Conserv. **208**, 65–75 (2023)
31. Kloetzer, L., et al.: Learning through citizen science: a framework for examining outcomes. Environ. Educ. Res. (2021)
32. Pizzolato, T., Tsuji, K.: School-based citizen science: moving beyond citizen-sensors. J. Sci. Educ. (2022)
33. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School engagement: potential of the concept, state of the evidence. Rev. Educ. Res. **74**(1), 59–109 (2004)
34. Damerell, P., Howe, C., Milner-Gulland, E.J.: Child-orientated environmental education influences adult knowledge and household behaviour. Environ. Res. Lett. **8**(1), 015016 (2013)
35. Skokanová, K., Jakab, I.: Mobile citizen science applications for teaching biodiversity within the third cycle of elementary education. In: International Scientific Conference on Distance Learning in Applied Informatics. Springer (2024)
36. Chandler, M., et al.: Useful biodiversity data were obtained by novice observers using iNaturalist during college orientation retreats. Citiz. Sci. Theory Pract. **6**(1), 32 (2021)
37. Gavora, P.: Úvod do pedagogického výskumu, 4. vyd. edn. Univerzita Komenského (2008)
38. Švec, Š., et al.: Metodológia vied o výchove: kvantitatívno-scientické a kvalitatívno-humanitné prístupy. IRIS, Bratislava (1998)
39. Enviroportál: Správa o stave životného prostredia Slovenskej republiky v roku 2019. Ministerstvo životného prostredia SR (2019)
40. Smithsonian Science Education Center: Mobile apps for citizen science (2015). https://ssec.si.edu/stemvisions-blog/mobile-apps-citizen-science
41. Oberhauser, K., LeBuhn, G.: Insects and plants: engaging undergraduates in authentic research. Front. Ecol. Environ. **10**(6) (2012)
42. Kermish-Allen, R., et al.: Characteristics and motivations of teachers participating in citizen science. J. Res. Sci. Teach. **56**(9) (2019)
43. Carrier, S.J., et al.: Citizen science in elementary classrooms: a tale of two teachers. Front. Educ. **9** (2024)
44. Shah, H.R., Martinez, L.R.: Current approaches in implementing citizen science in the classroom. J. Microbiol. Biol. Educ. **17**(1), 17–22 (2016)

45. Vázquez-Alonso, Á., Manassero-Mas, M.A.: Citizen science for scientific literacy and the attainment of sustainable development goals in formal education. Sustainability. **12**(10), 4283 (2020)
46. UNH Extension: Guide to choosing locally relevant citizen science projects for the classroom. University of New Hampshire Extension (2023)
47. California Academy of Sciences: Citizen science toolkit. https://www.calacademy.org
48. Lozano, A., López, R., Pereira, F.J., Blanco Fontao, C.: Impact of cooperative learning and project-based learning through emotional intelligence. Int. J. Environ. Res. Public Health. **19**(24), 16977 (2022)
49. Ardoin, N.M., et al.: Early childhood environmental education: a systematic review. Educ. Res. Rev. **31**, 100353 (2020)
50. Hernández-Barco, M., Sánchez-Martín, J., Corbacho-Cuello, I., Cañada-Cañada, F.: Emotional performance of a low-cost eco-friendly project based learning methodology. Sustainability. **13**(6), 3385 (2021)
51. PLT: Teaching and learning ecologically/Fostering SEL through nature education. Project Learning Tree (2021)
52. Kim, K., Lee, J.: Effects of project-based learning on students' motivation and self-efficacy. Engl. Teach. **73**(1), 95–114 (2018)
53. Cooper, C.B., et al.: The invisible prevalence of citizen science in global research. PLoS One. **9**(9) (2014)
54. National Geographic Society: Learn the ins and outs of citizen science (2023). https://www.nationalgeographic.org/article/citizen-science
55. Ciasullo, A., et al.: Citizen science in the classroom: challenges and opportunities. J. Sci. Educ. (2019)
56. Hod, Y., et al.: Learning with and from others' expertise. J. Learn. Sci. **27**(3), 306–343 (2018)
57. Kermish-Allen, R., et al.: Citizen science: schoolteachers' motivation, experiences, and recommendations. Int. J. Sci. Math. Educ. **21**, 527–547 (2022)
58. Haywood, B.K., Besley, J.C.: Education, outreach, and inclusive engagement. Front. Ecol. Environ. **12**(3), 186–192 (2014)

Bridging the Policy–Practice Gap in Primary Informatics Education: Ukrainian Teachers' Practices and the Polish Curricular Framework

Nataliia Morze[1] and Eugenia Smyrnova-Trybulska[2](✉)

[1] V. N. Karazin Kharkiv National University, 4 Svoboda Square, Kharkiv 61022, Ukraine
[2] University of Silesia in Katowice, Bankowa, 12, 40-007 Katowice, Poland
esmyrnova@us.edu.pl

Abstract. This study examines the gap between informatics education policy and classroom practice in Ukrainian primary schools, and identifies elements of the Polish curricular framework that may support evidence-informed reform. It combines an online survey of 78 Ukrainian primary teachers (conducted in January 2026) with a documentary analysis of Poland's core curriculum (*Podstawa programowa*) within an asymmetric comparative design. Survey data were analysed using descriptive statistics, including; variability measures for all quantitative items. The findings indicate Results show that Ukrainian informatics education in Ukraine typically begins in Grade 2, with 0.5–1 h of instruction per week and focuses primarily on algorithmic thinking and visual programming tools such as Scratch, Code.org. In contrast, robotics and AI literacy are addressed inconsistently and depend heavily on available resources. While teachers demonstrate high levels of motivation, they rely largely on self-directed professional development. By comparison, the Polish framework introduces, informatics from Grade 1 and ensures coherent progression, supported by national infrastructure and systematic teacher training, leading to stronger alignment between policy and classroom practice. The study is limited by its reliance on empirical data from a single country (Ukraine) and on document-level analysis for the Polish system. However, it identifies several transferable elements, including an earlier formal start to informatics education, targeted AI literacy training, and increased investment in infrastructure. These findings provide an evidence base to support the development of AI-ready primary informatics education in reform contexts.

Keywords: informatics education · primary school · computational thinking · teaching practices · digital competence · AI literacy · comparative analysis

1 Introduction

The digital transformation of society has reshaped educational goals and methods, positioning informatics as a key subject for developing algorithmic thinking, digital literacy, problem-solving skills, and early awareness of artificial intelligence. These competencies are essential for pupils growing up in technologically saturated and AI-mediated

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 349–361, 2027.
https://doi.org/10.1007/978-3-032-32104-6_30

environments. Across Europe, education systems respond by revising curricula, strengthening digital competence frameworks, and investing in infrastructure and teacher training. Frameworks such as DigComp, DigCompEdu, and the UNESCO AI Competency Framework highlight early computational thinking, responsible technology use, and ethical AI awareness, though implementation differs among countries.

Poland and Ukraine share regional and historical contexts but differ in the pace and systemic support of digital transformation. Poland's model is characterised by a stable, structured curriculum and national initiatives such as *Laboratoria Przyszłości*, ensuring access to robotics and digital technologies. Ukraine's New Ukrainian School (NUS) emphasises competence-based learning and cross-curricular digital literacy but often operates with uneven resources. Poland was selected as the comparator for three systematic reasons: (1) both countries share post-Soviet educational reform trajectories and have implemented competence-based curriculum revisions in the 2017–2018 period; (2) Poland's accession to the EU has driven structured digital education policy that Ukraine is increasingly aligning with through its EU candidacy commitments; (3) the Polish core curriculum (*Podstawa programowa*) is extensively documented and accessible, enabling reliable document-level analysis. These factors make Poland a methodologically sound and contextually relevant reference point.

Despite growing international research, few studies connect curriculum intentions with classroom realities, especially in transforming education systems. This study addresses this gap through an asymmetric comparative design combining survey data from Ukrainian primary informatics teachers with an analysis of the Polish curriculum. Instead of comparing countries directly, it contrasts Ukrainian classroom practices with the Polish normative model to examine how policy frameworks translate into practice.

The study analyses teaching organisation, methods, digital tools, and readiness for AI-oriented education, identifying which elements of the Polish approach may support further development of Ukrainian primary informatics education. The results contribute empirical evidence and offer insights for aligning curriculum and practice in the context of AI-ready primary education.

***Research questions were identified*:**

R1. How is primary informatics education implemented in Ukraine at the classroom level, and how does it relate to the Polish curriculum framework?
R2. Which digital and AI-related competencies are declared in the Polish model and realised in Ukrainian teachers' practices?
R3. What teaching methods and digital tools do Ukrainian primary informatics teachers use?

The study hypothesizes that there is a gap between the normatively defined model of teaching informatics in primary education in Poland and the actual pedagogical practices of Ukrainian teachers, particularly regarding the use of digital resources, project-based approaches, and the integration of AI education. At the same time, Ukrainian teachers demonstrate the potential to adopt selected components of the Polish model, provided that contextual and resource-related constraints are taken into account.

2 Literature Review

2.1 Primary Informatics Education: Global Trends

Recent developments in primary informatics education reflect a shift from isolated "digital skills" training toward a comprehensive understanding of computer science as a foundational element of early education. The dominant trends include the integration of computational thinking (CT), early programming, and problem-solving skills, introduced increasingly earlier through age-appropriate, block-based programming environments [1]. Systematic reviews demonstrate that both plugged and unplugged activities effectively support CT development, provided that the instructional design is coherent and aligned with learning outcomes [2, 3]. Educational robotics has emerged as another significant pathway for CT; meta-analyses confirm its positive effect on learning, while emphasising the importance of teacher readiness and classroom resources as moderating factors [4, 5]. Across studies, successful implementation is linked to teacher professional development, adequate infrastructure, and curricular continuity, with teachers' self-efficacy playing a central role in enabling project-based and inquiry-oriented learning formats [6].

2.2 Digital and AI Competence Frameworks (DigComp, DigCompEdu, AI4K12)

European policy increasingly relies on competence frameworks to systematise digital and AI-related learning outcomes. DigComp 2.2 defines domains of digital competence – including data literacy, digital creation, and safety – offering curriculum designers a flexible scaffold for mapping digital competence to primary informatics learning [7]. DigCompEdu expands this perspective to educators, identifying competencies related to the use of digital resources, differentiation, assessment, and fostering learners' digital skills [8]. In the emerging field of AI literacy, the AI4K12 initiative provides a structured progression built on the "Five Big Ideas in AI," promoting developmentally appropriate, ethically informed AI education rather than early technical specialization [9, 10]. Collectively, these frameworks underline that successful implementation depends not only on clearly articulated competencies but also on teacher preparation and institutional support.

2.3 Overview of Primary Informatics Education in Poland

Poland's 2017 curriculum reform formally introduced computer science as a core area of early education, emphasising logical and algorithmic thinking, programming, digital safety, and competent ICT use [11]. Polish informatics education combines cognitivist principles – structured content, progression, and clear objectives – with constructivist elements that stress collaboration, active learning, and guided discovery. The curriculum is reinforced by national commentary documents – [12] and is frequently cited in comparative international studies as a model of structured, progression-based K–12 informatics [13]. Policy initiatives further strengthen implementation: *Laboratoria Przyszłości* equips schools with robotics and STEAM resources [14], while programmes

such as *Lekcja: Enter* focus on upskilling teachers in active, technology-enhanced pedagogies [15]. Although these initiatives create a coherent model, research notes that local conditions and teacher readiness still influence implementation outcomes [6].

2.4 Ukrainian Context and Previous Studies

The New Ukrainian School (NUS) reform positions digital competence as a cross-curricular priority, while informatics is typically delivered as a separate subject starting in primary grades. The State Standard defines expected learning outcomes and ensures alignment with competence-based education goals [16]. National programmes provide curricular guidance, permitting some school-level flexibility [17]. Research in Ukraine often focuses on teacher readiness, highlighting the role of interdisciplinary integration and practical digital tools in implementing CT [18]. Further studies stress the need for capacity building to ensure teachers can design CT-oriented activities effectively [19]. Existing analyses emphasise variability in implementation quality due to resource disparities, a factor that gains particular significance when comparing Ukraine with more structurally supported systems such as Poland.

2.5 Theoretical Framework for Comparative Analysis

Comparing primary informatics education across countries requires analysing both normative-curricular structures and classroom-level practices. National standards set expectations for learning outcomes and competencies, but real-world implementation depends on teacher preparation, school infrastructure, and organisational conditions. In Poland, the *Podstawa programowa* defines informatics as compulsory from Grade 1, with a coherent progression of algorithmic, digital, and programming competencies. In Ukraine, digital competence is conceptualised as cross-curricular, while informatics is delivered as a separate subject with variable instructional time. These systemic differences affect classroom practices, instructional methods, and resource availability. The comparison of Poland's structured curriculum with empirically documented Ukrainian practices enables identification of divergences between policy intentions and practical realities, as well as potential areas for transferring successful solutions. Table 1 synthesises the key comparative parameters and serves as a foundation for interpreting the study's empirical findings.

Table 1 highlights substantial differences between the Polish normative model of primary informatics education and the classroom practices reported by Ukrainian teachers. In Poland, the *Podstawa programowa kształcenia ogólnego* defines informatics as a compulsory subject from Grade 1, specifying clear digital and programming competences [20]. Ukraine, by contrast, follows the State Standard of Primary Education and the New Ukrainian School (NUS) framework, where digital competence is conceptualised as cross-curricular, while informatics is taught as a separate subject with variable instructional time [21].

A major point of divergence concerns the starting age and instructional time. The Polish system requires informatics from Grade 1, ensuring continuous development of algorithmic thinking and digital competence [11]. Ukrainian teachers report that

Table 1. Comparison of Ukrainian teachers' practices in primary informatics education with the Polish curriculum and policy framework

Criterion/Parameter	Poland (normative–curricular level)	Ukraine (pedagogical practice based on survey results)
1. Regulatory framework	*Podstawa programowa kształcenia ogólnego [Core curriculum for general education]*; clearly defined digital and programming competences	State Standard of the New Ukrainian School (NUS); digital literacy integrated cross-curricularly; informatics as a separate subject
2. Start of informatics education	Formally introduced in Grade 1	Mostly from Grade 2; some schools start from Grade 1
3. Instructional time	1 h per week, with the possibility of extension through the school-based component	0.5–1 h per week, depending on the educational programme
4. Who teaches informatics	Teacher trained in informatics/digital education	Primary school teacher or informatics teacher (depending on the school)
5. Curriculum content	Computer literacy, algorithmics, Scratch, robotics, elements of 3D thinking	Algorithmic thinking, Scratch, digital safety, media literacy, basic robotics
6. Algorithmic thinking	Structured and compulsory component from Grade 1	Clearly present, but usually starts from Grades 2–3
7. Programming tools	ScratchJr, Scratch, Blockly, Scottie Go!, Ozobot	Scratch, Blockly, Code.org; in some schools Lego, micro:bit
8. Robotics	Systematically integrated (supported by the *Laboratoria Przyszłości* programme)	Implemented unevenly; depends on resources and project participation
9. STEM infrastructure	Basically ensured in most schools	Ranges from well-equipped to almost non-equipped institutions
10. Access to devices	Often a 1:1 or 1:2 student–device ratio	Most commonly 1:2–1:4; sometimes 1:10
11. Curricular flexibility	High level of school autonomy within the standard	More formalised programmes, but with some room for variability
12. Digital safety	Dedicated module; focus on ethics and cyberbullying	Integrated cross-curricularly; provided for by the NUS standard

(*continued*)

Table 1. (*continued*)

Criterion/Parameter	Poland (normative–curricular level)	Ukraine (pedagogical practice based on survey results)
13. Media literacy	Systematically integrated	Implemented mainly through projects and interdisciplinary links
14. Teacher training	Centralised state-funded professional development programmes	NUS courses, EdEra, Prometheus; strong reliance on self-education
15. Teacher motivation	Supported by standards and funding mechanisms	High, but dependent on access to resources
16. Educational platforms	Librus, Lekcja: Enter, Microsoft 365, Minecraft Education	All-Ukrainian Online School, Google Classroom, Office 365, Code.org
17. Assessment	Formative assessment, rubrics, projects	Formative assessment; no summative grading in Grades 1–4
18. Age differentiation	Strong emphasis on early start (Grades 1–3)	Gradual increase in complexity from Grade 2
19. Integration with other subjects	STEAM projects, robotics	Cross-disciplinary NUS projects
20. STEM/STEAM projects	Systematically supported at the state level	Implemented selectively
21. National digital initiatives	*Laboratoria Przyszłości*, *Cyfrowa Szkoła [Laboratories of the Future, Digital School]*	"Laptop for Every Teacher?

instruction begins mostly in Grade 2, which aligns with typical NUS programmes but reduces opportunities for earlier CT development.

In both countries, curricula include visual programming (Scratch, Blockly). However, Poland integrates robotics and STEM infrastructure more systematically, supported by national programmes such as *Laboratoria Przyszłości* and *Cyfrowa Szkoła* [7, 22]. In Ukraine, robotics and emerging technologies are introduced inconsistently and often depend on grants or donor-funded projects, as confirmed by survey responses.

Teacher preparation further distinguishes the two systems. Poland benefits from a centralised, curriculum-aligned system of professional development, resulting in comparatively high teacher digital competence [23]. In Ukraine, professional development relies on NUS training modules, large online platforms (EdEra, Prometheus), and self-directed learning, contributing to uneven readiness for advanced digital or AI-oriented practices.

Differences also emerge in AI literacy and readiness for future technologies. Polish policies already incorporate algorithmic thinking, data use, and digital ethics, forming a foundation for early AI education [7]. The survey indicates that in Ukraine, AI literacy

is at an initial stage and is implemented mainly through individual teacher initiative, lacking a unified curricular framework. This aligns with UNESCO's call for structured, age-appropriate AI education in primary schools [8].

Overall, Table 1 confirms a systemic gap: Poland's model is structured, resource-supported, and consistent, while Ukrainian practice is flexible but uneven and highly context-dependent. Nevertheless, survey data show strong professional motivation among Ukrainian teachers and a clear willingness to adopt selected features of the Polish model – especially robotics integration, project-based learning, and explicit competence progression – if institutional support and infrastructure improve.

The comparative analysis of normative documents in both countries clarifies the structural differences in how informatics is conceptualised: Poland emphasises early, compulsory, and systematically supported CS education, whereas Ukraine embeds digital competence broadly across subjects while allowing variation in informatics implementation. However, policy documents alone do not fully reveal how informatics is enacted in classrooms. To address this gap, the study uses empirical data from a questionnaire administered to Ukrainian primary teachers, offering insights into actual pedagogical practices, the digital tools used, and the extent of AI-related integration.

This dual-level analysis – policy comparison and empirical teacher data – provides a more complete understanding of informatics education in both contexts. It demonstrates that while the Polish model offers strong structural coherence, Ukrainian teachers demonstrate adaptability and potential for rapid advancement provided that systemic support, resource equity, and AI-oriented professional development are strengthened.

3 Research Methodology

The study employed a quantitative descriptive research design, with a questionnaire survey as the primary method of empirical data collection. The aim of the research was to identify current practices of teaching informatics in Ukrainian primary schools and to assess teachers' readiness to implement contemporary digital and AI-oriented instructional approaches.

Empirical data were collected through an author-designed online questionnaire developed on the basis of the State Standard of Primary Education of Ukraine, the New Ukrainian School framework, and European digital competence frameworks (DigComp, DigCompEdu). The questionnaire included blocks of items addressing the organisation of informatics teaching, instructional methods and digital tools used, experience with programming and robotics, the development of computational thinking and elements of AI literacy, and teachers' self-assessment of professional readiness.

The survey was administered in January 2026 using the Google Forms platform. A total of 78 educators from general secondary education institutions in Ukraine participated in the study. Respondents were recruited via professional teacher networks and online communities; participation was voluntary and anonymous. The sample comprised teachers from both urban (61%) and rural (39%) schools across multiple Ukrainian regions, with teaching experience ranging from 1 to 30 years ($M = 12.4$, $SD = 7.8$). The sample does not claim statistical representativeness; findings should be interpreted as exploratory and contextually bounded. Data analysis was performed using descriptive

statistical methods. The findings were interpreted in comparison with the Polish normative and curricular model of informatics education within an asymmetric comparative research design, which provides the analytical framework for the presentation of results in the following section.

3.1 Analysis of Survey Results

The results of the survey of educators from general secondary education institutions in Ukraine indicate significant variability in the organisation of informatics education at the primary school level, which directly relates to the **RQ1** and **RQ3** (see Fig. 1). The majority of respondents are primary school teachers and informatics teachers, reflecting a widespread model in Ukrainian practice whereby informatics in Grades 1–4 is taught both by subject-specialist informatics teachers and by generalist primary educators. This situation contributes to heterogeneity in instructional approaches, as the level of specialised training in informatics varies considerably among teachers, which, in turn, affects the content focus, pedagogical methods, and the depth of learning experiences provided to pupils.

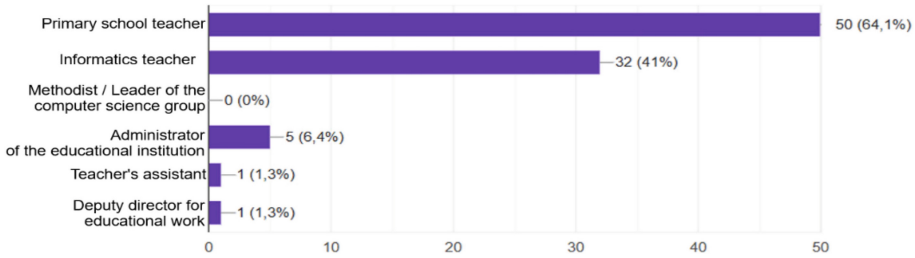


Fig. 1. Distribution of respondents by professional role

Analysis of responses related to organisational aspects of informatics education (see Fig. 2) shows that the subject is most commonly introduced from Grade 2 and less frequently from Grade 1. This confirms the presence of a temporal asymmetry compared with the Polish curricular model, which mandates an early start to informatics education. This observation is consistent with the research hypothesis that there is a gap between normatively defined models of informatics education and actual pedagogical practice in the Ukrainian context.

With regard to curriculum content and the development of key competences **(RQ2)**, the survey results reveal a clear focus among Ukrainian teachers on fostering algorithmic thinking, basic skills in using digital tools, and the application of visual programming environments, primarily Scratch and platforms such as Code.org (see Fig. 3).

At the same time, the integration of more advanced components – such as systematic work with data, robotics, or elements of AI literacy – remains fragmented and largely depends on the material and technical capacity of individual schools as well as teachers' personal initiative. In this study, AI literacy was operationalised through three survey items: (1) frequency of AI-related topics in lessons (5-point Likert scale), (2) self-reported familiarity with AI4K12 framework concepts, and (3) use of LLM-based tools

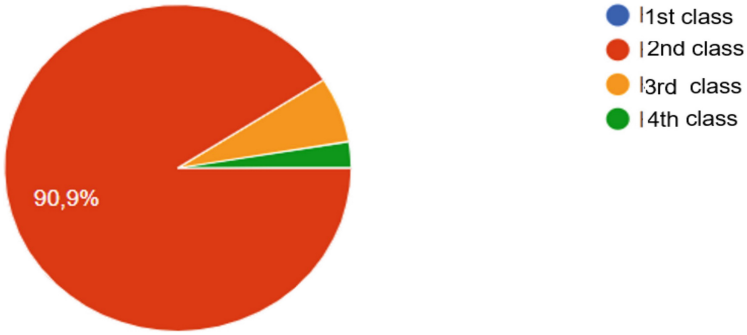


Fig. 2. Grade level at which informatics education begins in Ukrainian primary schools

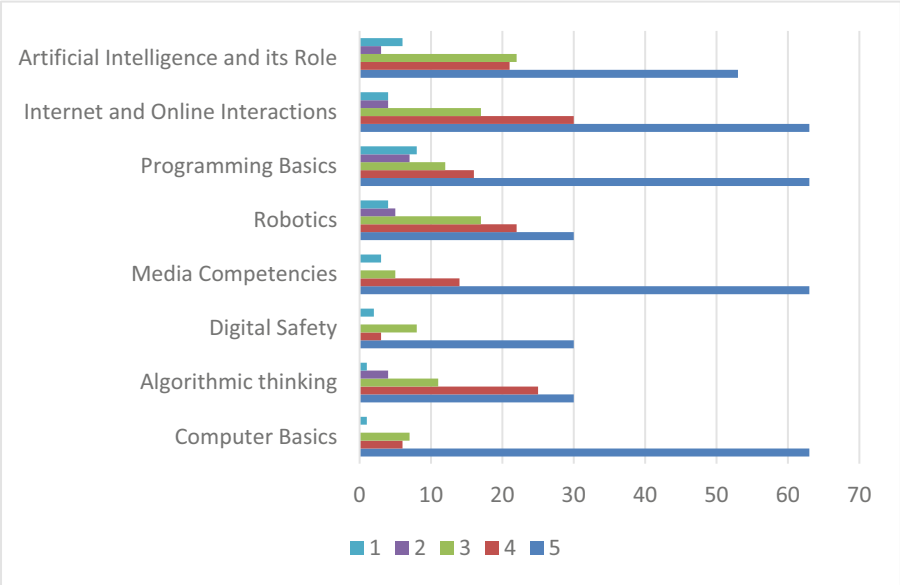


Fig. 3. Teachers' perceptions of the importance of key content areas in primary informatics education

in instructional design (frequency scale). These items do not constitute a validated scale; they offer an initial mapping of AI integration rather than a psychometric measurement of AI literacy competence. This finding supports the hypothesis that Ukrainian practices partially align with the competence-oriented goals articulated in European educational models, while lacking sufficient systemic coherence in their implementation.

4 Discussion of Results and Prospects for Further Research

The study's findings confirm the hypothesis of a clear gap between the competence-oriented intentions of Ukrainian educational policy and the actual practices observed in primary informatics classrooms. Although national documents emphasise the development of algorithmic and digital literacy, their implementation remains uneven and depends strongly on teachers' preparation, school infrastructure, and local management decisions. This aligns with international studies showing that pedagogical and institutional conditions critically influence the effectiveness of primary informatics education. The observed variability in implementation is primarily driven by structural conditions rather than by insufficient teacher motivation or competence: the absence of a mandatory, grade-specific informatics curriculum for Grades 1–4, unequal access to devices and robotics equipment, and reliance on self-directed rather than institutionally supported professional development. These systemic factors create a situation in which individual teachers absorb the burden of curriculum interpretation, resource acquisition, and methodological innovation – tasks that in better-supported systems are addressed at the policy level.

A comparison with the Polish normative model highlights the role of systemic support. In Poland, programming, robotics, and project-based learning are consistently integrated due to national initiatives and significant infrastructural investment. In Ukraine, however, the development of informatics education is driven largely by teacher initiative and flexibility, creating space for innovation but also contributing to unequal learning opportunities across schools.

Survey results reveal high motivation among Ukrainian teachers to develop professionally, including interest in AI literacy. Yet the absence of targeted, age-appropriate training in informatics didactics limits the realisation of curricular goals. This underscores the need for specialised professional development addressing programming methodologies, computational thinking, and gradual AI integration.

Future research should expand the empirical sample, conduct parallel international surveys with standardised instruments, and employ qualitative methods to analyse classroom practice in greater depth. Such work would support the design of more coherent educational policies amid digital transformation and the growing relevance of artificial intelligence. Beyond the Polish case, comparative data from European leaders in primary digital education – such as Estonia and Finland – suggest that the decisive enablers are not curriculum sophistication per se, but coherent national professional development systems, universal device access, and stable long-term funding for school infrastructure [6, 23]. This broader European context strengthens the transferability argument: Ukrainian teachers already demonstrate the motivational and adaptive capacity that, combined with systemic support, has driven rapid reform progress in comparable contexts. Future work involving parallel surveys in Poland and other European countries, using standardised instruments, would allow more rigorous cross-national comparison and further validate the findings reported here.

The chapter also highlights the importance of teacher preparation, drawing on the author's experience with the Computer Science Education Methodology module. The module equips future teachers for ICT use in education, to conduct informatics classes in Grades 1–3 in line with the curriculum [12, 24] and digital competence frameworks [8].

Its content includes core informatics concepts, visual programming (Scratch, Baltie), digital safety, lesson planning, multimedia task design, the use of LLM-based tools, and microteaching with peer assessment. Evaluation tasks include comparing educational projects, designing lesson plans, delivering micro-lessons, and creating safe computer-work guidelines.

Module implementation demonstrates that students successfully develop both technical and methodological competences, integrating ICT with subjects such as mathematics and Polish [24]. Microlearning supports reflective practice and the design of engaging, collaborative activities. Participants also recognise the need for continued development in visual programming and in combining informatics content with media education and online safety.

5 Conclusions

The study provides a comprehensive picture of informatics education in Ukrainian primary schools and its comparison with the Polish curricular model. Findings for RQ1 show substantial variability in Ukraine regarding the starting grade, instructional time, and teacher qualifications, contrasting with Poland's more standardised and systemically supported approach. Analysis of learning content (RQ2) indicates that Ukrainian teachers prioritise algorithmic thinking, basic digital skills, and visual programming, while the integration of robotics, data-oriented tasks, and AI literacy remains irregular and limited.

In terms of pedagogy and digital tools, instructional, practical, game-based, and project-oriented methods dominate Ukrainian classrooms, consistent with the New Ukrainian School philosophy. However, the use of STEM/STEAM solutions and technologically rich practices depends heavily on local resources and teacher initiative, restricting scalability.

Teacher preparation results reveal strong motivation for professional growth but a lack of systematic training in informatics didactics and AI-oriented methods, which hinders effective implementation of current expectations. Overall, the Ukrainian model shows considerable potential due to its flexibility and teachers' openness to innovation, yet faces challenges including unequal access to technology, fragmented methodological support, and limited integration of AI.

The findings confirm the hypothesis that combining empirical teacher data with the Polish curricular model provides a solid basis for adapting effective solutions to the Ukrainian context. These insights can inform educational policy, the design of teacher-training programmes, and future comparative research on digital transformation in primary education.

Preparing future teachers requires integrating technical, methodological, and pedagogical knowledge. Blending cognitivist and constructivist principles [25], along- side practical work, e-learning, and microlearning, supports competencies aligned with DigCompEdu [8]. Well-prepared teachers are essential for developing pupils' computational thinking, programming skills [26], and safe digital practices [27].

Broader comparative studies are planned in the near future with the participation of scientists from other countries to obtain more objective information on the formal,

substantive, methodological, organizational-technical, and legal-ethical aspects of computer science teaching in schools in different countries, as well as the determinants of teacher and child satisfaction with teaching and learning computer science in the era of AI and robots. The supplemented and extended data of the research results will be included in a new article, published later.

Acknowledgement. This research was supported by the Statutory funds, Ministry of Science and Higher Education, Poland, topic is „Conditions for the successful development and implementation of E-learning and innovative methods of teaching learning in a modern university in a digital society.", 2024–2026, Z.89 50203-1-33-23031-001.

References

1. Dagienė, V., Jevsikova, T., Stupurienė, G., Juškevičiene, A.: Teaching computational thinking in primary schools: worldwide trends and teachers' attitudes. Comput. Sci. Inf. Syst. **19**(1), 1–24 (2022). https://doi.org/10.2298/CSIS201215033D
2. Chen, P., Lai, Y.-C., Wang, S.-K.: Fostering computational thinking through unplugged activities: a systematic review and meta-analysis. Int. J. STEM Educ. **10**(1), 34 (2023). https://doi.org/10.1186/s40594-023-00434-7
3. Lin, Y., Zhu, C., Gardner, M.: Comparing the effects of plugged-in and unplugged activities on computational thinking. Educ. Inf. Technol. **29**(8), 1–34 (2024). https://doi.org/10.1007/s10639-023-12181-x
4. Wang, Y., Yin, Z., Liu, H.: Can robot-supported learning enhance computational thinking? A meta-analysis. Think. Ski. Creat. **52**, 101528 (2024). https://doi.org/10.1016/j.tsc.2024.101528
5. Tengler, K., Nemeth, L., Dörnyei, Z.: The effect of robotics-based storytelling activities on computational thinking. Educ. Sci. **12**(1), 10 (2021). https://doi.org/10.3390/educsci12010010
6. Stupurienė, G., Lucas, M., Bem-Haja, P.: Teachers' perceptions of the barriers and drivers of teaching computational thinking in primary school. Comput. Educ. **208**, 104939 (2024). https://doi.org/10.1016/j.compedu.2023.104939
7. Vuorikari, R., Kluzer, S., & Punie, Y.: DigComp 2.2: the digital competence framework for citizens—with new examples of knowledge, skills and attitudes. EUR 31006 EN, Publications Office of the European Union, Luxembourg (2022). https://doi.org/10.2760/115376
8. Redecker, C., Punie, Y.: European Framework for the Digital Competence of Educators: DigCompEdu. Publications Office of the European Union, Luxembourg (2017). https://doi.org/10.2760/178382
9. Touretzky, D., Gardner-Morse, M., Martin, F.: Machine learning and the five big ideas in AI. Int. J. Artif. Intell. Educ. **33** (2022). https://doi.org/10.1007/s40593-022-00314-1
10. AI4K12: AI4K12: sparking curiosity in AI. https://ai4k12.org/. Accessed 30 Jan 2025
11. Kwiatkowska, A.B., Syslo, M.M.: Informatyka w edukacji. Edukacja informatyczna a rozwój spoleczenstwa. Ksztalcenie na odleglosc seria: Informatyka w edukacji [Computer science in education. IT education and the development of society. Distance learning series: IT in education], red: Maciej M. Syslo, Torun, Poland: Nicolaus Copernicus University Press. 396 pp. ISBN 978-83-231-4437-3
12. ORE (Education Development Center): Informatyka: Podstawa programowa z komentarzem. Szkoła podstawowa (2017). https://dziennikustaw.gov.pl/du/2017/356

13. Oda, M., Horita, T., Noborimoto, Y.: International trends in K–12 computer science curricula through comparative analysis: implications for the primary curricula. Int. J. Comput. Sci. Educ. Schools. **4**(4), 24–59 (2021). https://doi.org/10.21585/ijcses.v4i4.102
14. Gov.pl: Laboratoria Przyszłości (program portal). https://www.gov.pl/web/laboratoria?
15. Gov.pl: Lekcja:Enter (program description). https://www.gov.pl/web/cppc/lekcja-enter
16. Cabinet of Ministers of Ukraine: On approval of the state standard of primary education (Resolution No. 87) (2018). https://zakon.rada.gov.ua/go/87-2018-?text=87-2018-%CF
17. Ministry of Education and Science of Ukraine: Навчальні програми для 1–4 класів / Primary education programs (Grades 1–4). https://mon.gov.ua/osvita-2/.../navchalni-programi-dlya-1-4-klasiv
18. Morze, N.V., Barna, O.V., Boiko, M.A.: The importance of computational thinking training for primary school teachers. Educ. Dimens. **6**, 22–39 (2022). https://doi.org/10.31812/educdim.4466
19. Morze, N., Barna, O., Boiko, M.: The relevance of training primary school teachers computational thinking. In: Proceedings of the VI International Workshop on Professional Retraining and Life-Long Learning Using ICT (3L-Person 2021) Co-located with ICTERI 2021, CEUR Workshop Proceedings, vol. 3104, pp. 141–153 (2021). http://ceur-ws.org/Vol-3104/
20. Ministerstwo Edukacji Narodowej: Podstawa programowa kształcenia ogólnego (2017). https://www.gov.pl/web/edukacja/podstawy-programowe-ksztalcenia-ogolnego
21. Ministry of Education and Science of Ukraine: Official Ukrainian state standard for primary education (2018). https://mon.gov.ua/osvita-2/zagalna-serednya-osvita/osvitni-programi/navchalni-programi-dlya-1-4-klasiv
22. Ministerstwo Edukacji i Nauki: Laboratoria Przyszłości—programme documentation. https://www.gov.pl/web/laboratoria?utm_source=chatgpt.com (2021)
23. UNESCO: AI competency framework for teachers and learners. UNESCO (2023). https://unesdoc.unesco.org/ark:/48223/pf0000391104; https://doi.org/10.54675/ZJTE2084
24. Smyrnova-Trybulska, E.: Development of prospective preschool and primary school teachers in the area of ICT use in education. E-mentor. **3**(90), 32–42 (2021). https://doi.org/10.15219/em90.1520
25. Morze, N., Smyrnova-Trybulska, E.: From programmed teaching to constructivism and personally-oriented approach to e-learning. EduAkcja Mag. Eduk. Elektronicznej. **2**, 15–28 (2016). https://www.researchgate.net/publication/315834783_From_Programmed_Teaching_to_Constructivism_and_Personally-Oriented_Approach_to_E-Learning
26. Bers, M.U.: Coding as another language: a pedagogical approach for teaching computer science in early childhood. J. Comput. Educ. **6**, 499–528 (2019). https://doi.org/10.1007/s40692-019-00147-3
27. Flogie, A., Lipovec, A., Škrobar, J.: Integrating computer science and informatics education in primary schools: insights from a Slovenian professional development initiative. Sustainability. **17**(20), 9068 (2025). https://doi.org/10.3390/su17209068

Reproducible Synthetic Tabular Data Generation for Education with Gemini

Olga Martinčíková Sojková[1](✉), Miroslav Adámek[2], and Mikuláš Gangur[1]

[1] Faculty of Economics, University of West Bohemia, Pilsen, Czech Republic
{sojkova,gangur}@fek.zcu.cz
[2] STATS LAB s.r.o., Považská 1709/61, 911 01 Trenčín, Slovakia
miroslav.adamek@statslab.eu

Abstract. The scarcity of real-world datasets often limits hands-on teaching of statistics and data mining due to privacy, security, and documentation issues, so controlled synthetic tabular data can provide realistic yet shareable training material. This paper proposes and evaluates a reproducible, code-driven Gemini pipeline that generates synthetic record-variable datasets directly in a structured format without an intermediate Matlab step, while ensuring that repeated runs produce different yet comparable files. The pipeline combines JSON-schema enforcement (Pydantic), a generate–validate–repair loop with strict local checks of types, ranges, and logical dependencies, targeted row-level repair of invalid records, and run-level comparability control via feature vectors and Mahalanobis-distance outlier detection with covariance shrinkage. In experiments on domain-consistent synthetic car datasets, the method reliably produced valid structured outputs and maintained stable comparability across repeated generations. In a stability test across 5 datasets generated over 10 runs, no outlier dataset was created in any run, meeting the goal of generating multiple datasets with consistent structure and data that meet the monitored similarity criteria.

Keywords: Synthetic tabular data · LLM · Gemini · Schema validation · Mahalanobis distance

1 Introduction

Artificial intelligence (AI) and machine learning (ML) have become integral to many educational activities, including the teaching of statistical analysis, data mining, and predictive modeling. A persistent obstacle in these contexts is the limited availability of suitable datasets: real-world data are often inaccessible due to privacy and security constraints, poorly documented, or difficult to process - important considerations when verifying students' knowledge and practical skills.

Controlled synthetic data generation offers a scalable solution. Synthetic datasets can emulate realistic patterns while adhering to predefined statistical profiles, enabling practical training even in sensitive domains such as finance, corporate environments, or healthcare, where authentic data cannot be shared.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 362–371, 2027.
https://doi.org/10.1007/978-3-032-32104-6_31

This paper continues our previous study on language-model-based generation of synthetic tabular data, in which we investigated model behavior in an MS Copilot workflow and documented typical failure modes, especially those related to prompt formulation. While the previous paper demonstrated that LLMs can produce usable tabular datasets, it also revealed two practical limitations - limited reproducibility in an interactive workflow and insufficient systematic verification of statistical and structural requirements - which motivate the methodology proposed here.

Building on these findings, we propose and evaluate a code-driven pipeline based on Gemini that generates record × variable datasets directly in a structured output format, eliminating an intermediate Matlab step. The main contribution is the integration of automated quality control: the generated data are systematically checked for validity (types, ranges, logical constraints), statistical fidelity (targeted univariate properties and dependencies), and suitability for downstream multivariate tasks. By combining structured generation with iterative validation, the approach increases reproducibility and reduces the risk of producing datasets that are formally correct but analytically unusable.

2 Related Work

Research on synthetic data contains (i) pedagogical uses in teaching/assessment, (ii) algorithmic generators for tabular data, and (iii) frameworks for evaluating utility and privacy. In education, synthetic datasets are advocated as a means to provide realistic yet shareable data for hands-on learning; importantly, recommendations emphasize that educational synthetic data should preserve multivariate structure ("richness") and realism while remaining aligned with learning objectives [1].

Beyond generic classroom datasets, recent work proposes generating purpose-built ("bespoke") teaching datasets from already cleared analytical outputs, enabling learners to develop code and workflows even when original data cannot leave secure environments [2]. Complementarily, studies on learning analytics show that synthetic student data can be compared across generators using downstream-model consistency (utility) rather than only marginal similarity, and that privacy constraints shape what can be shared [3, 4].

On the generation side, classic and widely adopted tabular synthesizers include sequential conditional approaches such as synthpop [5], broader tool ecosystems such as SDV (including multi-table synthesis and benchmarking tooling) [6]. At the same time, multiple surveys and evaluation frameworks argue that the key open problem is not "how to generate rows," but how to verify that synthetic data meets requirements on fidelity, utility for specific tasks, and privacy risk - preferably with interpretable, structured metric suites rather than ad hoc checks [7, 8].

LLM-based generation adds a new axis: models such as GReaT treat tabular rows as token sequences and can produce realistic mixed-type tables [9], while later work improves correlation and label consistency via fine-tuning and conditional sampling strategies [10]. However, LLMs are also prone to schema violations and failures to respect functional dependencies, which motivates the use of structured-output adapters and validator layers that measure both schema accuracy and content fidelity [11, 12].

These results suggest that for educational multivariate analysis tasks, the practical contribution lies in end-to-end, code-driven pipelines that couple LLM-based synthesis with schema enforcement and automated quality control, enabling datasets to be generated directly in the target environment without manual post-processing.

3 Methodology

An effective synthetic dataset for teaching should preserve the key analytical properties of real data (plausible distributions and dependencies) while remaining controlled, shareable, and aligned with learning objectives. In practice, LLM-generated tables rarely reproduce full real-world complexity - especially when many inter-variable constraints are required - so the goal in education is not perfect realism but predictable structure, controlled difficulty, and reliable statistical behavior.

Building on our previous study, where we used an interactive MS Copilot workflow and observed prompt-design failure modes and limited reproducibility, this paper introduces a code-driven Gemini pipeline that generates record $\times$ variable datasets directly in structured formats (e.g., CSV), without an intermediate Matlab step, and integrates an automated generate-validate-regenerate quality-control loop. Gemini was selected over Copilot primarily because it is better suited for programmatic use via an API and is supported by a broader technical ecosystem for this purpose, including services such as Google Workspace and Vertex AI. The reference to Copilot draws on the authors' previous related study, in which Copilot was employed at an earlier stage of the research. The process starts with an explicit data profile (a schema, hard constraints such as ranges and logical rules, and soft targets such as means/SDs, class proportions, and correlation structure). Gemini then generates the dataset using reusable prompt templates that encode this profile and enforce machine-parseable output, optionally adding conditioning rules to strengthen intended relationships.

Quality is ensured by automated checks in three layers: validity (schema/type/range/logic compliance), fidelity (alignment with target marginals and dependencies), and utility (whether the data reliably support intended multivariate tasks such as PCA, regression, and clustering). Datasets that fail validity are regenerated. By combining direct generation with systematic QC and logging, the pipeline reduces format drift and run-to-run variability, improves comparability of assignment difficulty, and lowers manual intervention compared to the prior workflow.

The main goal is to create detailed prompts for directly generating data files that can be used to practice multivariate statistical analysis tasks. This has not been possible before due to the ambiguous outputs at different times. The solution was to create Matlab code using LLM, modify it "manually", and then use it to generate your own data files as a synthetic data generator. One of the requirements is to skip the Matlab code step and generate the data file directly. Another requirement is that the data files should be different yet similar in the monitored criteria when generated repeatedly. The following metrics were included among the currently monitored criteria:

- file size, i.e. number of records,
- number of categorical attributes
- number of dichotomous attributes,

- number of numeric attributes,
- number of attributes with normal distribution,
- number of attributes without identifiable distribution.

The generated data files should be simultaneously checked for

- the required quality, i.e. whether when generating one data file it corresponds to the specified criteria,
- the required comparability with each other, i.e. whether when generating at different times two different data files correspond to the specified criteria.

The point is therefore to construct a code in the used Python system that, using Gemini AI, will use constructed prompts that generate the required data on a given topic, fulfill the quality criteria, further check the fulfillment of the prescribed criteria and at the same time try to solve the problems addressed in the already published article [13], i.e. meaningful values of categorical attributes and corresponding values of related attributes, whether categorical or numerical in an imaginary hierarchical structure (e.g. car brand - model - engine - power - torque - price).

Furthermore, the second requirement for comparability with already generated data files would be implemented using the Python command structure.

3.1 General Description of the Methodology: Idea

The proposed methodology for generating and checking synthetic datasets is based on two complementary principles:

1. Controlled generation of data based on an explicit plan and
2. subsequent multivariate quality control of datasets using outliers.

Generation takes place in two phases. In the first phase, a structured plan (catalog) is created from the textual specification of the required data, explicitly defining permitted attribute combinations (e.g., brand-model-permitted drives and gearbox types) and realistic intervals for numerical values (e.g., price, mileage). This plan serves as a so-called "contract", significantly reducing the risk of inconsistencies typical of free generation (e.g., nonexistent models or unrealistic property combinations). In the second phase, based on the plan, the records themselves are generated so that each line respects the defined restrictions and preserves statistically and domain-realistic values.

Since records that violate the rules can occur sporadically even with controlled data generation, the pipeline includes automatic repair of only erroneous lines. Instead of repeating the entire generation, invalid records are identified, to which error diagnostics are added, and the model is asked to minimally correct only these items. This increases stability and scalability for large numbers of records while maintaining consistency across datasets.

At the dataset quality level, the methodology is extended to detect outlier datasets using the Mahalanobis distance. In one run (rune), n datasets (e.g., 5) are generated. In contrast, each dataset is represented by a multidimensional vector of summarization characteristics (e.g., the number of records, the number of numerical/categorical attributes, and, for each numerical variable, the mean, median, and standard deviation).

The Mahalanobis distance then quantifies how far a particular dataset is from the "typical" dataset in a given run, accounting for correlations among characteristics. This step allows you to automatically identify a different dataset (e.g., too high prices, atypical raid distribution, or variation in variance) and may require regeneration or a more detailed audit. The entire process thus combines domain-controlled generation, local validation and correction, and global dataset quality control at the run level.

3.2 Technical Description of the Methodology (Implementation Procedure)

The implementation is implemented in the Python/Jupyter environment and consists of three main layers:

1. structured outputs of the model,
2. local strict validation and line correction,
3. multidimensional detection of outlier datasets.

In the first layer, the "controlled generation" mechanism of the generative model is used, in which a JSON schema is passed to the model, and strictly valid JSON must be returned. In practice, the schema is generated from typed data models (Pydantic) and sent via the response_json_schema parameter along with response_mime_type = "application/json". This approach forces the model to produce output with well-defined types and fields (e.g., Price is an integer within a prescribed interval; MileageKm is an integer in km; FuelType is a value from an enumeration). The first phase (PLAN) uses a plan schema in which the catalog contains real Make/Model combinations and structured motorizations in pair form (Motorization, FuelType) to eliminate ambiguous textual descriptions.

```
PLAN_SYSTEM_INSTRUCTION = """
You are a data architect. Create a plan for synthetic car data.

Return only valid JSON according to the schema. No markdown, no
comments.

Plan rules:
- Catalog must contain real existing Make/Model combinations.
- Motorizations MUST be a list of objects, each object has:
- Motorization: either "EV" or "X.YL" format (e.g. "1.5L", "2.0L")
- FuelType: one of {Gasoline, Diesel, Hybrid, Electric}
Examples:
{"Motorization":"2.0L","FuelType":"Gasoline"}
{"Motorization":"2.0L","FuelType":"Hybrid"}
{"Motorization":"EV","FuelType":"Electric"}
- Do not use texts like "2.0L Hybrid" or anything similar.
- Specify Transmissions and DriveTrains as a list of allowed values
for the given model.
- Specify realistic PriceMin/PriceMax and MileageMinKm/Mileage-
MaxKm ranges for the given model.
- Catalog should have 20-60 items.
""".strip()
```

The generated catalog in json format is used in the next step of data generation.

```
You are a synthetic car data generator for students.

Return only valid JSON, no markdown and no comments.

Use ONLY combinations and allowed values from the catalog:
{catalog_json}

Rules:
- Select Make+Model from the catalog.
- Select Motorization+FuelType as one pair from the Motorizations
item for the given model.
- Select Transmission and DriveTrain from the allowed lists for the
given model.
- Price must be in [PriceMin, PriceMax], MileageKm in [Mile-
ageMinKm, MileageMaxKm].
- Electric => Motorization="EV" and EngineSizeCc=0.
- Non-electric => EngineSizeCc approximately corresponds to liters
(liters*1000 ±250).
- Number of rows exactly as requested by the user.
""".strip()
```

The second layer ensures that, even with formally correct JSON output, records remain in compliance with domain consistency rules. Each generated line is validated locally by a strict model that, in addition to types, also checks logical links: with FuelType = “Electric” Motorization = “EV” and EngineSizeCc = 0 must be present; for internal combustion and hybrid drives, Motorization must be in the form of X.YL and EngineSizeCc approximately corresponds to liters (conversion liters$\times$1000 with tolerance). The output is generated in batches due to stability and output token limits. If validation identifies invalid lines, a “repair payload” containing the original line, its position, and a summary of the validation error is created. Subsequently, a correction step is initiated, which provides the model with only erroneous records, the relevant catalog section (subset by Make/Model), and a request for minimal correction while preserving the order. The corrected lines are revalidated by the strict model and reinserted into their original positions. This mechanism reduces costs and time compared to repeating the entire batch and increases the determinism of the resulting datasets.

```
repair_system = f"""
You are a "repair agent" for synthetic car data.

Return only valid JSON according to the CarDataset schema
({{"rows":[...]}}). No markdown, no comments.
The number of returned rows must be exactly {len(chunk)} and in the
same order as the input list.

Repair rules:
- Repair minimally (change only necessary fields).
- Make+Model must be in the catalog. (If invalid, choose the closest
one from the catalog.)
- Motorization+FuelType must be an allowed pair according to the
catalog for the given model.
- Transmission and DriveTrain must be from the allowed lists for
the given model.
- Price and MileageKm must be in the ranges for the given model.
- Electric => Motorization="EV" and EngineSizeCc=0.
- Non-electric => EngineSizeCc ~ litre*1000 ±250.

CATALOG (subset):
{json.dumps(catalog_subset, ensure_ascii=False, indent=2)}
""".strip()

user_prompt = f"""
Fix the following rows. Each item has the original row and a val-
idation error.
Return only the corrected rows in the format: {{"rows":[...]}} and
preserve the order of the items.

INPUT:
{json.dumps(chunk, ensure_ascii=False, indent=2)}
""".strip()
```

The third layer concerns quality control at the dataset level. A feature vector is created for each dataset that summarizes its structure and distribution characteristics: the number of records (file size), the number of categorical, dichotomous, and numerical attributes, and, for each numerical column, three statistics (mean, median, SD). These vectors are standardized (z-scores), and at the level of a single run (e.g., 5 datasets), a covariance matrix is calculated with regularization (Ledoit-Wolf shrinkage) due to the small sample size. The Mahalanobis distance of each dataset from the centroid (average feature vector) is interpreted through approximation

$$MD^2 \sim \chi^2_p \tag{1}$$

, where p is the number of features used, and the dataset is marked as an outlier when the threshold is exceeded

$$\chi^2_{(p,1-\alpha)} \tag{2}$$

In practice, dimensionality is limited for stability with a small number of datasets (e.g., selection of the most variable 3–4 features). The output is a clear table with the Mahalanobis distance and p-value, which allows you to identify an outlier dataset within the run and then discard it, regenerate it, or analyze the cause of the deviation.

In this way, the methodology combines controlled generation with an explicit scheme, local-domain validation with automatic error correction, and statistical outlier detection across datasets, leading to robust synthetic data suitable for pedagogical and experimental purposes.

4 Results

The required datasets were generated using the procedure outlined in the methodology. The following is an example of the generating process's output as a single record in the json format.

```
{
  "Make": "Volkswagen",
  "Model": "Golf",
  "Motorization": "1.0L",
  "FuelType": "Gasoline",
  "Transmission": "Manual",
  "DriveTrain": "FWD",
  "Price": 18500,
  "MileageKm": 75000,
  "Color": 3,
  "EngineSizeCc": 900,
  "WarrantyYears": 5
}
```

The output of the third phase of data quality control is also shown, i.e., the comparison of generated datasets and the determination of the Mahalanobis distance, along with the result of the remoteness test, given by the p-value. In the outlier detection procedure, the significance level was set to $\alpha = 0.05$ (Table 1).

Table 1. Comparison of generated datasets

Dataset	MD	MD2	p_value	is_outlier	n_features_used
run1_ds3	2.662921	7.091146	0.131149	False	4
run1_ds2	2.375038	5.640804	0.227627	False	4
run1_ds4	2.056613	4.229658	0.375816	False	4
run1_ds1	1.737224	3.017946	0.554827	False	4
run1_ds5	0.142990	0.020446	0.999948	False	4

The stability of the dataset generator was finally verified through a test in which five datasets were generated over 10 runs. At $\alpha = 0.05$, the system did not generate a single

outlier dataset in any run. This met the goal of generating multiple datasets with stable structure and data with respect to the measured similarity of the individual datasets to each other.

5 Conclusion

The experiments on generating datasets with the required structure confirmed the functionality of the proposed solution. Compared to our earlier workflow, the two-stage approach - first producing an explicit data set structure and generating record-level content only then - proved essential for reducing inconsistencies and improving reproducibility. A key benefit is the integrated validation loop: by combining strict schema constraints with a validation prompt and self-correcting code, the system not only detects violations (types, ranges, and logical dependencies) but can also repair erroneous rows. At the run level, the final quality-control step provides an additional safeguard by quantifying dataset-to-dataset compactness using predefined similarity criteria and multivariate outlier detection, enabling automatic identification (and potential regeneration) of datasets that deviate too strongly from the target. The stability of the dataset generator was verified in an experiment where five datasets were generated across 10 runs; using a significance level of $= 0.05$, the system did not produce a single outlier dataset in any of the runs. This result supports the objective of generating multiple datasets that differ in content yet remain comparable in terms of the monitored structural and statistical similarity criteria.

Future work can focus on generalizing the assignment by formulating more open rules for generating the structure and for creating meaningful restrictions on the attribute structure generatively. A central research challenge is to prevent AI model "hallucinations" while preserving comparability across repeated runs, even as the schema becomes richer and the dependency structure more complex. In addition, future research should include a systematic comparison of Gemini with other large language models to evaluate whether the observed performance is model-specific or generalizable across LLM-based systems.

Acknowledgements. This paper was supported by the project SGS-2024-030, "Research on quantitative approaches for economic problems solving" of the Faculty of Economics of the University of West Bohemia in Pilsen.

References

1. Bulmer, M., Coote, L.: The role of synthetic data in teaching and learning statistics. In: Bridging the gap: empowering and educating Today's learners in statistics. Proceedings of the eleventh international conference on teaching statistics (2022). https://doi.org/10.52041/iase.icots11.t14i2
2. Elliot, M., Little, C., Allmendinger, R.: The production of bespoke synthetic teaching datasets without access to the original data. In: Privacy in Statistical Databases. PSD 2024. Lecture Notes in Computer Science, pp. 144–157. Springer (2024). https://doi.org/10.1007/978-3-031-69651-0_10

3. Zhan, C., Deho, O.B., Zhang, X., Joksimovíc, S., de Laat, M.: Synthetic data generator for student data serving learning analytics: a comparative study. Learn. Lett. **1**(5) (2023). https://doi.org/10.59453/khzw9006
4. Vie, J.-J., Rigaux, T., Minn, S.: Privacy-preserving synthetic educational data generation. arXiv preprint https://arxiv.org/abs/2207.03202. (2022)
5. Nowok, B., Raab, G., Dibben, C.: synthpop: bespoke creation of synthetic data in R. J. Stat. Softw. **74**(11) (2016). https://doi.org/10.18637/jss.v074.i11
6. Patki, N., Wedge, R., Veeramachaneni, K.: The synthetic data vault. In: IEEE DSAA 2016, pp. 399–410 (2016). https://doi.org/10.1109/dsaa.2016.49
7. Stoian, M.C., Giunchiglia, E., Lukasiewicz, T.: A survey on tabular data generation: utility, alignment, fidelity, privacy, and beyond. arXiv preprint https://arxiv.org/abs/2503.05954. (2025)
8. Yang, S.C.-H., Eaves, B., Schmidt, M., Swanson, K., Shafto, P.: Structured evaluation of synthetic tabular data. arXiv preprint https://arxiv.org/abs/2403.10424. (2024)
9. Borisov, V., Seßler, K., Leemann, T., Pawelczyk, M., Kasneci, G.: Language models are realistic tabular data generators. arXiv preprint https://arxiv.org/abs/2210.06280. (2022)
10. Nguyen, D., Gupta, S., Do, K., Nguyen, T., Venkatesh, S.: Generating realistic tabular data with large language models. arXiv preprint https://arxiv.org/abs/2410.21717. (2024)
11. Wang, D.Y.-B., Shen, Z., Mishra, S.S., Xu, Z., Teng, Y., Ding, H.: SLOT: structuring the output of large language models. arXiv preprint https://arxiv.org/abs/2505.04016. (2025)
12. Xu, S., Lee, C.-T., Sharma, M., Yousuf, R.B., Muralidhar, N., Ramakrishnan, N.: Why LLMs are bad at synthetic table generation (and what to do about it). arXiv preprint https://arxiv.org/abs/2406.14541. (2024)
13. Gangur, M., Sojková, O.M.: Using an AI-based language model to generate synthetic statistical data. In: EAI/Springer Innovations in Communication and Computing, pp. 97–107. Springer (2025). https://doi.org/10.1007/978-3-031-81261-3_8

Analysis of the Pilot Verification of the Questionnaire on the Use of Digital Teaching Aids by Teachers in the School Education Process

Júlia Sudorová[1](✉), Ján Záhorec[2], and Michal Munk[3]

[1] Department of Pre-Primary and Primary Education, Faculty of Education, Comenius University in Bratislava, 813 34 Bratislava, Slovakia
sudorova8@fedu.uniba.sk

[2] Department of Didactics of Mathematics and Natural Sciences, Faculty of Education, Comenius University in Bratislava, 813 34 Bratislava, Slovakia
zahorec@fedu.uniba.sk

[3] Department of Informatics, Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, 949 74 Nitra, Slovakia
mmunk@ukf.sk

Abstract. The paper builds research examining the implementation of digital didactic tools in pedagogical practice and focuses on the pilot verification of the research instrument used in this study. The questionnaire was administered to identify current needs of educational practice related to the development of teachers' professional competencies within the Slovak regional school system, particularly their digital literacy and didactic-technological competencies affecting the quality and effectiveness of instruction. It was designed to provide valid data on the level of digital competencies of primary and secondary school teachers, especially in relation to their use of selected digital tools and technologies in teaching. The pilot verification aimed at assessing the instrument's psychometric properties, with emphasis on reliability. The paper presents the results of reliability and item analyses. Internal consistency was examined using reliability coefficients, and potentially problematic items were identified through item analysis. The findings indicate satisfactory internal consistency of the analysed scales and confirm the instrument's suitability for the main research phase, while suggesting minor refinements of selected items.

Keywords: Didactic-technological competencies · digital teaching aids · pilot verification · research instrument reliability · Cronbach's alpha · item analysis

1 Introduction

The digital transformation of education fundamentally reshapes pedagogical practice and increases demands on teachers' professional competencies. The effective use of digital teaching aids in the school education process requires not only technical proficiency but,

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 372–384, 2027.
https://doi.org/10.1007/978-3-032-32104-6_32

above all, the ability to integrate these tools didactically into instruction. Monitoring the level of these competencies therefore represents an important prerequisite for the systematic planning of pedagogical innovation and the professional development of teachers within the regional school system.

However, empirical investigation of didactic-technological competencies requires a methodologically sound and psychometrically validated research instrument. Without verification of its reliability and validity, the interpretability of collected data and the credibility of subsequent analytical conclusions cannot be ensured. Pilot verification of the instrument thus constitutes an essential phase of the research process, aimed at identifying potential weaknesses in the questionnaire design prior to the implementation of the main study.

This paper focuses on the analysis of the pilot verification of a questionnaire designed to examine the use of digital teaching aids by teachers in the school education process. Particular attention is devoted to the assessment of content validity through expert evaluation and the calculation of the Content Validity Index (CVI), as well as to the examination of internal consistency of selected dimensions using Cronbach's alpha and item analysis. The objective is to evaluate the psychometric readiness of the instrument for application in the main quantitative research across different levels of regional education.

2 Theoretical Background

Teachers' digital competencies are currently operationalised through various competence frameworks and empirically validated measurement instruments. The European Reference Framework of Key Competences emphasises digital literacy as one of the fundamental competences for lifelong learning [1]. In educational research, the DigCompEdu framework is widely used, as it structures teachers' digital competencies into clearly defined domains and provides a basis for their assessment through questionnaire-based instruments [2].

Empirical validation studies highlight the necessity of multi-stage verification of the psychometric properties of research instruments. The validation of the DigCompEdu Check-in questionnaire was conducted using exploratory and confirmatory factor analysis, and the results confirmed its construct validity and satisfactory internal consistency [2]. Similarly, instruments based on the TPACK model have been validated through factor analysis and the calculation of Cronbach's alpha, thereby confirming their structural stability and reliability [3].

The development of a research instrument in the social sciences requires a systematic procedure that includes defining the construct, generating items, expert evaluation, and subsequent verification of reliability and validity [4]. Reliability refers to the degree of internal consistency of an instrument and is most commonly assessed using Cronbach's alpha [5]. The interpretation of this coefficient must consider the number of items and their homogeneity; values above 0.70 are generally regarded as acceptable for research purposes [5, 6].

Validity concerns the extent to which an instrument measures the construct it is intended to measure [6]. For newly developed or adapted questionnaires, the assessment of content validity through expert evaluation is particularly important. The Content Validity Index (CVI) enables the quantification of expert agreement regarding the relevance

of individual items [12]. This approach has also been applied in recent validation studies within educational research [7].

Pilot verification of a research instrument therefore represents an essential phase in ensuring its content adequacy, structural consistency, and psychometric readiness prior to the main study. The combination of qualitative pilot testing and quantitative reliability analysis is recommended in the validation of questionnaire-based instruments in educational research [4, 7].

3 Methodology

3.1 Aim of the Pilot Verification

The aim of the pilot verification was to assess the psychometric properties of the research instrument designed to examine the use of digital teaching aids by teachers in the school education process. The pilot phase focused on:

- verifying the content validity of the items,
- identifying potential interpretative ambiguities,
- assessing the internal consistency of selected dimensions,
- conducting item analysis to detect potentially problematic items.

Pilot testing represents a standard phase in instrument development, enabling the optimisation of its structure prior to the implementation of the main study [4].

3.2 Instrument Characteristics

The questionnaire was designed as a structured instrument consisting of:

- identification items,
- scale-based items focused on the frequency and manner of using digital teaching aids,
- items assessing self-reflection of digital competencies.

Most scale-based items were formulated using a Likert-type scale. The construction of the items was informed by established frameworks of teachers' digital competencies [1, 2] and by models of technology integration in education [3].

In the quantitative phase, the pilot verification focused on selected dimensions of the instrument (Dimension *B* and Dimension *D*), representing key areas of the use of digital teaching aids and self-assessment of digital competencies. The remaining sections of the questionnaire will be analysed in the main phase of the research.

3.3 Research Design and Sample

In line with the aforementioned principles, the research study was designed along two complementary lines: qualitative and quantitative.

Qualitative Phase
The qualitative pilot study was carried out with a sample of five female teachers from the

regional school system. The aim was to verify item comprehension, identify terminological ambiguities, and assess content relevance. The qualitative phase primarily focused on the analysis of non-scale items in the questionnaire, intended to detect potential interpretative inconsistencies, terminological issues, and possible deviations in understanding the meaning of the questions.

Methodologically, this phase was based on the principles of cognitive interviewing [8, 9]. The collected responses were analysed using thematic analysis following Braun and Clarke [11].

The qualitative phase also included expert evaluation of selected items in order to quantify content validity using the Content Validity Index (CVI) [12].

A quantitative pilot analysis was conducted on a sample of $N = 313$ teachers from a regional school district in the Slovak Republic. The pilot sample was sufficient for statistical inference, which allowed for the application of relevant statistical procedures to assess the reliability of the instrument and to identify suspicious items. Since the teachers who completed the survey had previously successfully completed several professional development programs focused on developing didactic-technological competencies – particularly in the area of software applications and digital teaching tools applicable in pedagogical practice – this research sample can be considered adequate given the focus of our measurement tool. The respondents' opinions and evaluations are therefore relevant and provide a reliable picture of the level and needs of their further professional development in the area of didactic-technological competencies.

Quantitative Phase

The quantitative phase focused on assessing reliability based on data collected from $N = 313$ teachers. The sample consisted exclusively of educational staff in the categories of teacher, educator, teaching assistant, and vocational training instructor, with the group of teachers including primary-level teachers (first-grade elementary school teachers), lower secondary school teachers, and upper secondary school teachers. The research sample was predominantly female and included educational staff with varying lengths of teaching experience. A purposive sampling strategy was used to ensure representation of different levels of education and diversity in the length of teaching experience.

3.4 Data Analysis Procedure

Content validity was evaluated using:

- I-CVI (Item-Level Content Validity Index),
- S-CVI/Ave (Scale-Level Content Validity Index),

where ratings of 3 and 4 on a four-point scale were considered relevant [12].

The reliability of individual dimensions was assessed using Cronbach's alpha [5]. The analysis also included item-total correlations and the calculation of Cronbach's alpha if item deleted, which represents a standard procedure in instrument refinement [4, 6].

Descriptive statistics (mean and standard deviation) were used to identify variability in responses and potential extreme items. In the main study, exploratory factor analysis

will be conducted to examine the underlying structure of the instrument using appropriate extraction and rotation methods.

3.5 Ethical Considerations

Participants were informed about the purpose of the study and the anonymity of data processing. Participation was voluntary, and all data were used exclusively for research purposes.

4 Results of the Pilot Instrument Verification

The presented results refer to the pilot analysis of selected dimensions of the instrument (Dimensions *B* and *D*), which were identified as methodologically prioritised for reliability assessment. The pilot verification was conducted in two complementary phases: a qualitative phase (focused on content validity and item clarity) and a quantitative phase (focused on reliability and item structure).

4.1 Qualitative Pilot Testing and Content Validity

The qualitative phase was conducted with a sample of five female teachers using cognitive interviewing techniques [8, 9]. The analysis of responses primarily identified terminological ambiguities (e.g., differing interpretations of the concept of "didactic-technological competencies") and minor semantic variability in certain item formulations. However, no substantive content deficiencies were detected.

Content validity was quantified using the Content Validity Index (CVI) [12]. The Item-Level Content Validity Index (I-CVI) was calculated using the following formula (1):

$$I - CVI = \frac{n_e}{N} \tag{1}$$

where:

- n_e = number of experts rating the item as 3 or 4
- N = total number of experts

The I-CVI was calculated as the proportion of experts rating an item as relevant (3 or 4). All evaluated items achieved an I-CVI of 1.00 (Table 1).

The scale-level index (S-CVI/Ave) was calculated as the average of the I-CVI values using formula (2).

$$S - CVI/Ave = \frac{\sum \boldsymbol{I - CVI}}{k} \tag{2}$$

Since all I-CVI values were equal to 1.00, the resulting S-CVI/Ave was 1.00.

With five experts, a value of 1.00 exceeds the recommended threshold of 0.78 [12], indicating a very high level of content validity.

Table 1. Expert Ratings and I-CVI Calculation

Item	E1	E2	E3	E4	E5	Number of Relevant Ratings	I-CVI
A2	4	4	4	4	4	5	1.00
A3	4	4	4	4	4	5	1.00
B1	4	4	4	4	4	5	1.00
B3	3	3	4	3	4	5	1.00
B5	4	4	4	4	4	5	1.00
B7	4	4	4	4	4	5	1.00
B8	3	3	3	4	4	5	1.00

The identified comments resulted in minor wording refinements of selected items rather than their elimination.

Although the obtained values of I-CVI = 1.00 and S-CVI/Ave = 1.00 indicate maximum agreement among experts regarding item relevance, the relatively small number of evaluators must be taken into account. In accordance with methodological recommendations [12], the interpretation of CVI results is sensitive to the size of the expert panel. Therefore, the pilot phase should be considered a preliminary assessment of content validity. Further verification of construct validity will be conducted within the main study using a larger sample of respondents.

4.2 Reliability and Item Analysis

The quantitative verification was conducted on a sample of $N = 313$ respondents.

Dimension *B*: Use of Digital Teaching Aids
Cronbach's alpha reached $\alpha_B = 0.808$. This value indicates good internal consistency [5, 6]. The average inter-item correlation was r = 0.260.

Item analysis identified two items with low item-total correlations:

- B2_e (r = 0.036)
- B4_e (r = 0.011)

The removal of item B4_e would increase Cronbach's alpha to 0.818. However, the item was retained due to its content relevance (Table 2).

Table 2. Item Analysis (Item–Total Correlations)

Variable	Mean if deleted	Std.Dev. if deleted	Itm-totl correlation	Alpha if deleted
B2_a	21.52279	3.797793	0.537755	0.785515
B2_b	21.60673	3.777231	0.563538	0.782579
B2_c	21.89322	3.660132	0.558008	0.782818
B2_d	21.96147	3.877781	0.399605	0.800326
B2_e	21.50808	4.176483	0.035999	0.819931
B4_a	22.54637	3.604179	0.615437	0.774719
B4_b	22.49808	3.583518	0.667454	0.767150
B4_c	22.66384	3.668180	0.575036	0.780226
B4_d	22.86503	3.747046	0.557250	0.782697
B4_e	23.75712	4.193345	0.010959	0.818199

Dimension *D*: Self-Assessment of Digital Competencies

Cronbach's alpha reached $\alpha_D = 0.886$. This value indicates a high level of reliability [5]. The average inter-item correlation was $r = 0.315$.

Item–total correlations ranged from 0.395 to 0.662. The removal of any item did not result in a substantial increase in Cronbach's alpha (Table 3).

Table 3. Item–Total Statistics for Dimension *D*

Item	Mean if deleted	Std.Dev. if deleted	Itm-totl correlation	Alpha if deleted
D1	40.27191	8.958088	0.501786	0.880856
D2	40.33323	8.923450	0.563828	0.878493
D3	40.50638	8.992394	0.476659	0.881742
D4	40.29287	8.938460	0.499962	0.881042
D5	40.41755	8.942505	0.528925	0.879800
D6	39.93321	8.858931	0.580859	0.877832
D7	39.92409	8.840987	0.566805	0.878483
D8	40.58987	8.951900	0.568996	0.878383
D9	40.23645	8.870901	0.542151	0.879487
D10	38.97861	9.131210	0.440498	0.882671
D11	40.82340	9.053228	0.579604	0.878829
D12	40.71073	9.087982	0.479461	0.881493
D13	40.80867	9.000753	0.661753	0.876565
D14	40.10286	8.951199	0.509222	0.880573
D15	39.60368	8.947601	0.528859	0.879793
D16	38.88993	9.157866	0.474423	0.881961
D17	38.82235	9.230066	0.395002	0.884034
D18	39.24548	9.063003	0.467678	0.881841

4.3 Descriptive Statistics

Mean values in Dimension *B* ranged from M = 1.00 to M = 3.25. In Dimension *D*, mean values ranged from M = 1.56 to M = 3.56 (Tables 4 and 5).

Table 4. Descriptive Statistics for Dimension *B*

Means and Standard Deviations		
Item	Mean	Std.Dev.
B2_a	3.235294	0.666761
B2_b	3.151351	0.674067
B2_c	2.864865	0.850380
B2_d	2.796610	0.676240
B2_e	3.250000	0.256330
B4_a	2.211712	0.867965
B4_b	2.260000	0.844439
B4_c	2.094241	0.821006
B4_d	1.893048	0.724704
B4_e	1.000958	0.012623

Table 5. Descriptive Statistics for Dimension *D*

	Means and Standard Deviations	
Item	Mean	Std.Dev.
D1	2.109929	0.989678
D2	2.048611	0.952930
D3	1.875458	0.970338
D4	2.088968	1.026532
D5	1.964286	0.973434
D6	2.448630	1.028799
D7	2.457746	1.077885
D8	1.791971	0.900752
D9	2.145390	1.069746
D10	3.403226	0.765014
D11	1.558442	0.727132
D12	1.671111	0.791911
D13	1.573171	0.720983
D14	2.278986	0.989558
D15	2.778157	0.965063
D16	3.491909	0.668019
D17	3.559486	0.621850
D18	3.136364	0.855362

Item B4_e exhibited minimal variability (SD = 0.01), which may indicate a specific interpretation by respondents or its distinct positioning within the construct.

4.4 Integration of Qualitative and Quantitative Findings

The low item–total correlation of item B4_e (r = 0.011) corresponds with findings from the qualitative pilot phase, in which terminological and interpretative variability was identified in selected formulations. At the same time, the minimal variability of responses (SD = 0.01) indicates a highly homogeneous pattern of responses, which may limit the item's contribution to the differentiation of the construct.

From a methodological perspective, the low correlation does not necessarily indicate content irrelevance but may reflect the item's specific function or limited discriminative power within the dimension. This type of finding is consistent with recommendations

for multi-stage instrument validation combining qualitative and quantitative approaches [4, 7].

5 Discussion

The aim of the pilot verification was to assess the psychometric properties of the instrument designed to examine the use of digital teaching aids in school education. The results indicate that the instrument demonstrates satisfactory to high reliability and very high content validity.

The values of Cronbach's alpha (0.808 for Dimension *B* and 0.886 for Dimension *D*) are consistent with recommended criteria for research instruments [5, 6]. Comparable reliability coefficients have been reported in validation studies measuring teachers' digital competencies, such as the DigCompEdu Check-in instrument [2] and instruments based on the TPACK model [3]. This suggests that the analysed instrument achieves a level of internal consistency comparable to internationally used measurement tools.

The higher alpha value observed in Dimension *D* compared to Dimension *B* may be related to the nature of the measured construct. Self-assessment of digital competencies likely represents a more cohesive latent construct, whereas the declared use of digital teaching aids may be more strongly influenced by contextual factors, such as school type, subject specialisation, or material and organisational conditions. This difference reflects varying levels of homogeneity across the dimensions and aligns with the multidimensional nature of teachers' digital competencies described in the literature [2, 3].

The item B4_e exhibited very low item–total correlation and minimal variability, suggesting limited discriminative ability. Its retention is justified by its conceptual relevance, while its performance will be further evaluated in the main study. The qualitative pilot phase also identified potential interpretative variability in selected formulations. The combination of these findings suggests that the weaker association of the item with the overall construct may result from its specific meaning or limited discriminative power rather than from content irrelevance. This interpretative approach is consistent with recommendations for multi-stage instrument validation that combine qualitative and quantitative procedures [4, 7].

The difference in internal consistency between the dimensions may further reflect the distinct characteristics of the measured phenomena. While self-reflection of digital competencies represents a relatively stable self-perception of professional abilities, the reported use of digital teaching aids may be more strongly determined by contextual variables such as technical infrastructure, subject area, or institutional organisation. Therefore, the lower homogeneity of Dimension *B* should not necessarily be interpreted as a weakness of the instrument, but rather as an indication of greater variability in educational practice.

The very high content validity values (I-CVI = 1.00; S-CVI/Ave = 1.00) indicate maximum agreement among experts regarding item relevance. Although a value of 1.00 reflects strong consensus, the relatively small number of experts must be considered, as this represents a common limitation of pilot validation studies [12]. Nevertheless, the combination of expert evaluation and quantitative reliability analysis provides a more robust assessment of instrument quality than the use of a single validation procedure.

Additional limitations include the purposive sampling strategy applied in the pilot phase and the absence of factor analysis, which would allow a more detailed examination of the construction structure. Factor analysis is planned in the main study using a larger sample.

Another limitation concerns the self-report nature of the instrument, which may be influenced by subjective perceptions of one's own competencies (self-report bias). The pilot phase also did not include test–retest reliability assessment or structural factor analysis. These steps are planned for the main phase of the research, where a more comprehensive evaluation of construct validity and structural stability will be conducted.

Overall, the results of the pilot verification confirm the psychometric readiness of the instrument for application in the main study and support its potential use across different levels of the regional school system.

5.1 Implications

The validated instrument provides a basis for systematic mapping of the level of implementation of digital teaching aids across different levels of the regional school system. Its application may enable the identification of areas requiring support within teachers' professional development and may serve as a foundation for revising the content of pre-service teacher education programmes.

In a broader context, the instrument may contribute to evidence-based formulation of innovation strategies in the field of digital transformation of education.

6 Conclusion

This paper presented the results of the pilot verification of a questionnaire designed to examine the use of digital teaching aids by teachers in the school education process. The instrument was verified through two complementary phases—qualitative and quantitative—with the aim of assessing its content adequacy and internal consistency.

The results of the expert evaluation confirmed a very high level of content validity, while the identified comments led to minor wording refinements of selected items. The quantitative analysis demonstrated satisfactory to high reliability of the analysed dimensions. Item analysis enabled the identification of selected items with weaker associations with the overall construct; however, their retention was methodologically justified and will be further examined in the main study.

The combination of qualitative and quantitative validation procedures provided a comprehensive evaluation of the instrument's psychometric properties and strengthened its methodological grounding. The remaining sections of the instrument will be analysed in the main study using a larger sample. Based on the pilot verification, the questionnaire can be considered ready for application in the main research focusing on the use of digital teaching aids across different levels of the regional school system.

Acknowledgments. This work was supported by the Grant Agency for Culture and Education of the Ministry of Education, Science, Research, and Youth of the Slovak Republic as part of the KEGA project No. 077UK-4/2025: *Edu-Steam laboratory in primary education teaching.*

References

1. European Communities: Key competences for lifelong learning. Official Journal of the European Union (2006)
2. Martín Párraga, L., Llorente Cejudo, C., Barroso Osuna, J.: Validation of the DigCompEdu check-in questionnaire through structural equations. Educ. Sci. **12**, 574 (2022)
3. Aqib, M.A., Ekawati, R., Khabibah, S.: Development of TPACK-LK instrument for mathematic preservice teacher. TEM J. **14**(1), 438–446 (2025)
4. Kalkbrenner, M.T.: A practical guide to instrument development and score validation in the social sciences: the MEASURE approach. Pract. Assess. Res. Eval. **26**(1) (2021)
5. Tavakol, M., Dennick, R.: Making sense of Cronbach's alpha. Int. J. Med. Educ. **2**, 53–55 (2011)
6. Taherdoost, H.: Validity and reliability of the research instrument. Int. J. Acad. Res. Manag. **5**(3), 28–36 (2016)
7. Govindasamy, P., Cumming, T.M., Abdullah, N.: Validity and reliability of a needs analysis questionnaire. J. Res. Spec. Educ. Needs. **24**, 637–652 (2024)
8. Willis, G.B.: Cognitive Interviewing: A Tool for Improving Questionnaire Design. Sage, Thousand Oaks (2005)
9. Beatty, P.C., Willis, G.B.: Research synthesis: the practice of cognitive interviewing. Public Opin. Q. **71**(2), 287–311 (2007)
10. Tourangeau, R., Rips, L.J., Rasinski, K.: The Psychology of Survey Response. Cambridge University Press, Cambridge (2000)
11. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006)
12. Polit, D.F., Beck, C.T.: The content validity index: are you sure you know What's being reported? Res. Nurs. Health. **29**, 489–497 (2006)

Data-Driven Analysis of EV Charging Behavior for Priority-Oriented Load Management Systems

István Szűcs[1(✉)], József Kopják[2], and Gergely Sebestyén[2]

[1] Doctoral School of Applied Informatics and Applied Mathematics, Obuda University, Budapest, Hungary
szucs.istvan@uni-obuda.hu

[2] Kandó Kálmán Faculty of Electrical Engineering, Obuda University, Budapest, Hungary
{kopjak.jozsef,sebestyen.gergely}@kvk.uni-obuda.hu

Abstract. The rapid growth of electric vehicle (EV) adoption increases electricity demand, creating challenges for distribution grids during peak periods. Intelligent load management systems provide a cost-effective alternative to grid reinforcement, but their effectiveness depends on accurate modeling of user charging behavior. This paper presents a data-driven analysis of EV charging patterns based on real-world data collected from an office context in Hungary. The infrastructure consists of 23 AC chargers divided into two independent charger groups. The dataset covers 520 days, including 7268 charging sessions from 70 users. The analysis investigates session count, session duration, energy demand, charger loyalty, and session utilization. Results show a typical workplace routine, however, high intra-user variability and significant differences between the two charger groups were observed, limiting the accuracy of aggregated prediction approaches. The study concludes that user grouping is necessary and should be used as the basis for forecasting and priority-oriented load management strategies in office charging environments.

Keywords: EV charging · Load management · User patterns

1 Introduction

The rapid increase in the number of electric vehicles (EVs) has led to a significant rise in electricity demand on distribution grids. This growing demand poses challenges for grid stability and infrastructure capacity, especially during peak periods. To address these challenges, two primary approaches are commonly considered: reinforcing and expanding the existing power grid infrastructure, or implementing intelligent load management systems. While grid expansion can provide a long-term solution, it often requires substantial financial investment, long planning cycles, and regulatory approval. As an alternative, load management systems offer a more flexible and cost-effective solution [1–3]. These systems rely on intelligent control strategies to dynamically distribute the available electrical power among multiple EV chargers based on predefined criteria such

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 385–394, 2027.
https://doi.org/10.1007/978-3-032-32104-6_33

as charging demand, user priority levels, time of arrival, or grid constraints. The development of effective load management strategies requires a thorough understanding of user behavior and charging usage patterns. Key factors include arrival and departure times, energy demand, charging session count, session utilization, and user loyalty to a specific charger. Analyzing such statistical data is essential for designing control algorithms that can balance user satisfaction with grid limitations. This paper focuses on the analysis of EV charging user patterns based on real-world data collected in an office environment. The facility under study is equipped with 23 alternating current (AC), OCPP-compliant (Open Charge Point Protocol) [4] chargers, each with a nominal charging power of 22 kW. The charging infrastructure is divided into two separate groups, where users are restricted to using only the chargers assigned to their respective group. This configuration allows for a comparative analysis between the two groups in terms of usage intensity, load profiles, and temporal charging behavior. The primary objective of this study is to identify and characterize charging patterns within the office context, including daily, weekly trends, and charger utilization rates. The insights gained from this analysis are intended to support the design and validation of intelligent load management systems that can optimize power distribution, prevent grid overloads, and improve the overall efficiency of EV charging infrastructure in similar commercial settings.

2 Related Works

Prior research has shown that EV charging behavior differs substantially across functional locations (e.g., residential vs. public) and across EVSE (Electric Vehicle Supply Equipment) types (e.g., Level 2 vs. DC fast charging), leading to distinct utilization patterns over hours, weekdays, and seasons [5–7].

Authors of [8] analyze 1 year of charging behavior from 31 EVs, categorizing charging by location (Home, Work, Public, Other) and user type. The results show that charging is strongly location-dependent, with the largest peak occurring at work between 09:00 and 10:00, while home charging peaks in the evening. The authors also report that only a small fraction of charging occurs during off-peak hours (00.00–06.00), motivating demand-shifting.

The study [9] examines a dataset and applies multiple clustering techniques (k-means, Ward linkage, and an intersection-based hierarchical method) to group users based on connection time, session duration, and energy delivered. The results consistently identify four user groups with distinct charging patterns. A key behavioral result is that peak energy delivery occurs in the early afternoon, with long session durations and a strong relationship between session duration and delivered energy. The authors also show that afternoon charging users split into two groups: one with high energy demand and long sessions and another with lower but still significant demand.

A campus case study [10] at the University of Georgia analyzes charging behavior using data from three charge point (level 2 charging stations) and 3204 charging events collected between 2014 and 2017. The authors report that charging predominantly occurs during weekday office hours, with 65.2% of charging events taking place between 08:00 and 18:00 (Monday–Friday), additionally, 84.5% of charging events took place on weekdays. In terms of charging duration, they find that charging times are typically short:

77.4% of charging events lasted 3 h or less. They distinguished three most popular time periods for charging: 11:00 a.m. to 12:00 p.m.; 12:00 p.m. to 1:00 p.m.; and 5:00 p.m. to 6:00 p.m., which the authors suggest may be explained by the fact that these spikes "coincide with typical lunch and dinner hours".

3 Methodology

Charging session data were collected in an office environment equipped with 23 OCPP-J1.6-compatible AC chargers, each rated at 22 kW and operating without load balancing or power limitations, allowing chargers to reach their nominal output when supported by the connected vehicle. The data comprise two independent datasets, referred to as Group 1 and Group 2, corresponding to two separate charging entities within the building. The dataset spans 520 days for Group 1 and 449 days for Group 2, and includes charging activity from 70 registered users. Group 1 includes 10 chargers serving 30 users, with a total session count of 4404, while Group 2 includes 13 chargers serving 41 users and 2864 sessions. One user appears in both datasets, therefore, the total number of unique users across the two datasets remains 70. The total session count is 7268. User identification is based on RFID tags; therefore, a single user may be associated with multiple tags over time (e.g., due to loss and replacement).

Communication between the chargers and the central system was carried out using the Open Charge Point Protocol (OCPP), enabling standardized data exchange and session monitoring. The collected data consist of charging session timestamps, energy meter readings, user identifiers (RFID Tags), and charge point identifiers, forming the basis for subsequent analysis. Further details on the data collection methodology are provided in a previous publication [7].

The collected dataset was analyzed using the Python Pandas library, which provides efficient data structures and tools for processing large-scale tabular and time-series data. Pandas was used for data cleaning, filtering, and aggregation [11]. The dataset includes the start and stop timestamps of each charging session, corresponding energy meter readings at the beginning and end of the session, as well as user ID, session ID, charger ID and a group to which the user belongs. In addition, the actual charging power was monitored during active sessions with a one-minute sampling interval, providing power measurements for each charging event. The dataset contains anomalies and records from invalid sessions. A session is considered invalid if the corresponding energy consumption is zero. These anomalies mainly originate from user errors, such as starting multiple charging sessions due to inexperience. However, from the perspective of this study, these events do not carry useful information. Other anomalies were also observed, such as charging sessions with extraordinarily high energy consumption or session duration for a single user. These outliers were handled by applying Winsorization to the dataset. Winsorization is a statistical technique for handling extreme values by replacing them with less extreme values instead of removing them [12]. In this case, the resulting distortion is acceptable, as EVs have a limited battery capacity, making it physically impossible to draw such large amounts of energy (exceeding 100 kWh, likely due to charger issues). Regarding session duration, it was observed that some users left their vehicles connected to the charging point, resulting in session durations of approximately 3–4 days. This phenomenon was rare (4–5 sessions) and was therefore also winsorized.

In numerical terms, zero energy consumption was recorded in 203 sessions, representing 2.8% of the total sessions (these were removed from the dataset). Winsorization was applied to 122 sessions (energy range: 0.4–77 kWh; session duration range: 0.1–22.7 h). As a result, the total number of valid sessions after removing zero-energy sessions and applying winsorization is 7065.

This study analyzes the dataset within the framework of priority-based charger scheduling and load management, with the aim of identifying user behavior and usage patterns that can guide the design of optimal management strategies.

4 Results

In the following section, the results of the data analysis are presented, providing an overview of user charging patterns within the given context. Figure 1 illustrates the energy consumption of both user groups over the data collection period. A delay is observed for Group 2, whose users began charging approximately two and a half months after the start of data collection. The figure further suggests that vehicles in Group 2 are associated with higher battery capacities, as reflected by their higher energy consumption levels. In addition, the plot reveals seasonal usage patterns, with identifiable periods where charging behavior deviates from typical trends, such as during the winter holiday season. Analyzing the dataset on a yearly basis enables the investigation of how variations in temperature and seasonal effects influence charging behavior and energy consumption over time. Previous studies have shown that ambient temperature has a significant effect on electric vehicle charging behavior, as reported in [7].

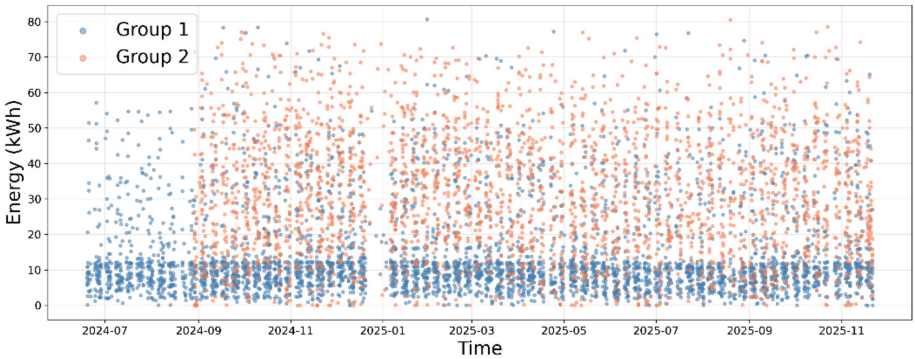


Fig. 1. Energy demand by all users over time during the studied period

4.1 User Patterns

Three primary measures can be examined: session count, session duration, and energy consumption. The average weekly session count is a particularly important metric, as it provides a high-level overview of user activity levels. The corresponding averages

are shown in Fig. 2. The plot indicates that, within each group, the mean values are mostly similar across users, with only a few notable exceptions. However, the relatively large standard deviations around the means—driven, for example, by many users having weeks with zero sessions—indicate substantial intra-user variability.

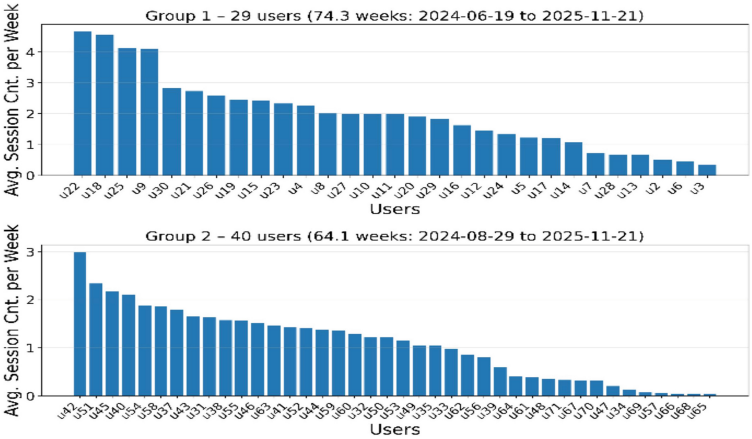


Fig. 2. Average weekly charging sessions per user

Due to the intra-user variability, the session count parameter requires further examination. Figure 3 shows substantial variation in session counts across users in both groups. In Group 1, the average weekly session count of 0–6 sessions is most common; however, significantly higher values occur more frequently than in Group 2. In contrast, users in Group 2 predominantly exhibit 0–4 charging sessions per week, with fewer occurrences of higher session counts.

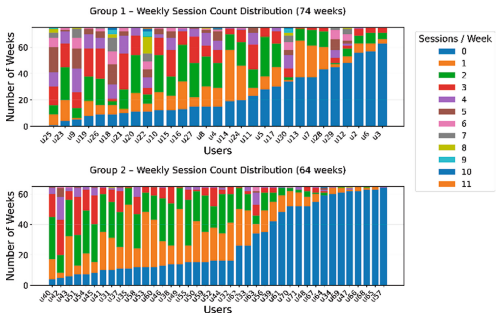


Fig. 3. Session count distribution of users

Weekly patterns are also important to study because they provide crucial information for applications such as load forecasting. The two groups are also analyzed separately in

this case. In Fig. 4, the user patterns are shown by the day of week. The average session-count pattern is essentially the same for both groups. Three distinct days (Tuesday–Thursday) show increased session counts, likely due to the possibility of working from home. In contrast, only a few sessions occur on weekends, which is consistent with the context of the charging locations. The error bars indicate the standard deviation. The middle two plots show the average energy consumption by weekday. The average demand is broadly similar across weekdays for both groups; however, Group 2 shows increased demand on weekends. This may be explained by users leaving their vehicles at the charging point over the weekend (increased duration on the last plot can also be observed). The standard deviation is higher for Group 1, which—based on Fig. 1—can be attributed to a larger share of EVs with lower battery capacities. Session duration is relatively constant across weekdays, except on weekends. For Group 2, the increased weekend energy demand can be explained by the longer session duration: They have more time to fully charge their EVs.

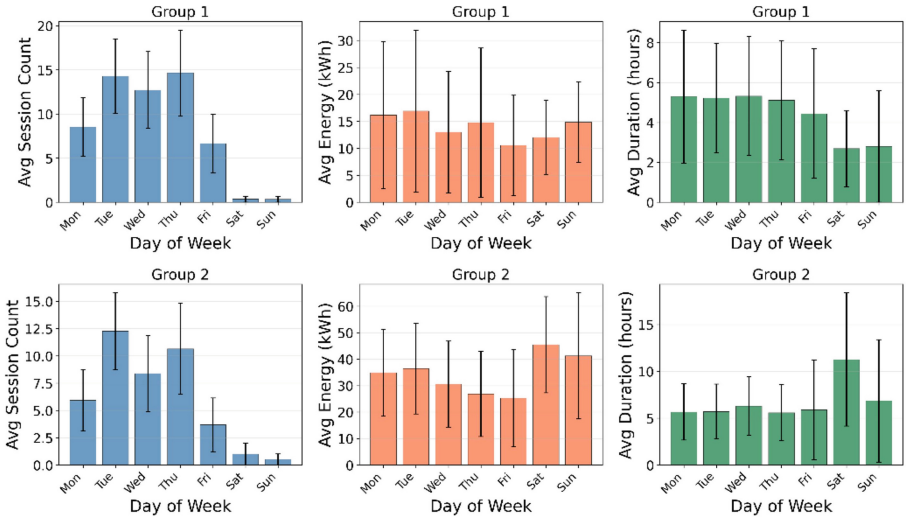


Fig. 4. User patterns based on weekdays

In terms of load forecasting, daily patterns are also important, as they provide information about the expected session count, charging duration and the amount of energy the users require depending on the start time of the charging. Since the number of charging sessions on weekends is negligible, only workdays will be used to calculate the averages and standard deviations. Figure 5 shows the relationship between these metrics and the start hour. As expected, the session count exhibits a morning peak (7 a.m.–8 a.m.), consistent with an office-related context (start of the work). In addition, a secondary peak can be observed between 12:00 and 13:00 that coincide with lunch time. Regarding energy demand, sessions starting in the morning require more energy for both groups, while late-night sessions are energy-intensive only for Group 2. Session duration differs between the two groups. Group 1 members tend to stay at the charge point for shorter

periods during the night, whereas Group 2 sessions are substantially longer, indicating that users in this group often leave their vehicles at the charging point overnight, or it can be a nightshift. The standard deviation is high in all cases, suggesting substantial variability in user behavior.

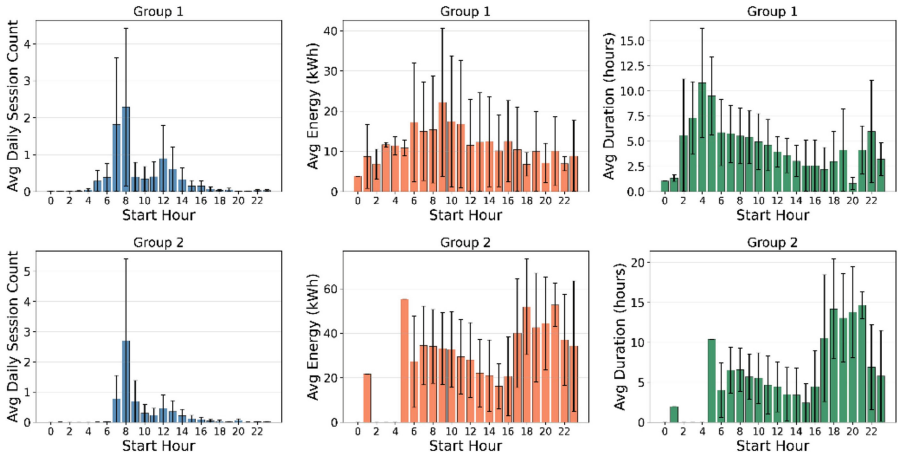


Fig. 5. User patterns vs. start time

As the standard deviations were relatively high in the previous cases, Fig. 6 shows the previous metrics at user level.

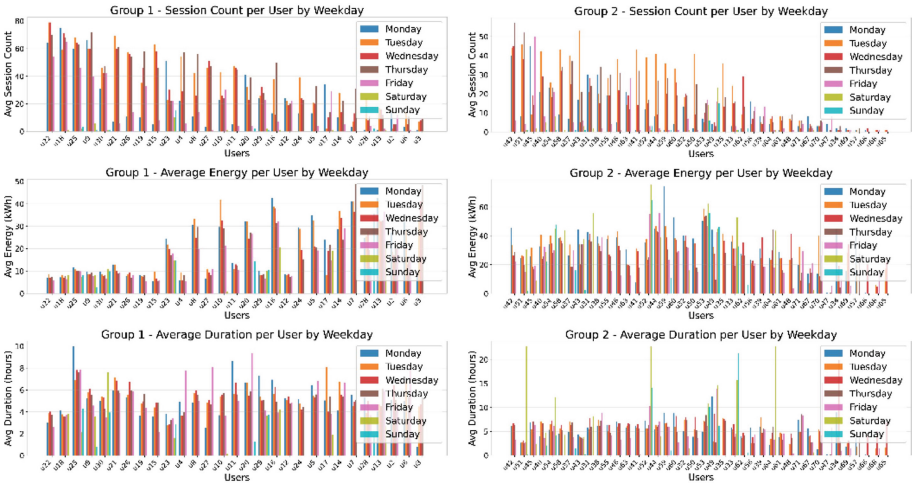


Fig. 6. User patterns on weekdays

4.2 Charger Patterns

User loyalty to a specific charge point is important because it determines how predictable the load is at the charger level. If users repeatedly return to the same charge point, demand becomes more stable and consistent, which enables a static prioritization strategy for chargers. In this case, chargers can be assigned fixed priorities based on expected usage patterns, and static load management may be sufficient. In contrast, when loyalty is low and users frequently switch between charge points, charger utilization becomes less predictable, making dynamic prioritization and dynamic load management more appropriate. The loyalty matrix can be seen in Fig. 7. In Group 1 each user used about 9 (8.8) different chargers on average at least once, 48% of their sessions occurred on their dominant charger on average. Group 2 is a bit more consistent in this aspect, the average unique chargers are 5.1, and 73% is the average dominant charger usage. Based on the results, users tend to use multiple chargers rather than relying exclusively on a single one, particularly in Group 1.

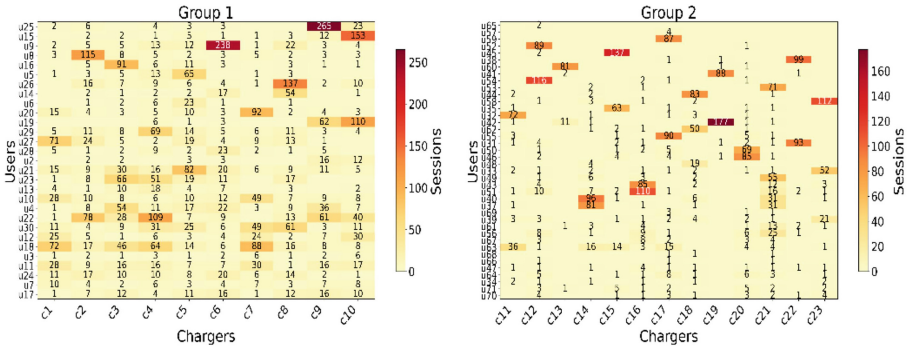


Fig. 7. Users loyalty to a specific charger

The session utilization (actual charging time/session duration) of a session is an important indicator because it reflects how much temporal flexibility exists for scheduling and controlling charging. If a significant share of the session consists of non-charging (idle) time, the charging process can potentially be shifted or scheduled (e.g., using load management) without increasing the overall plug-in duration. In contrast, sessions with a high utilization offer less flexibility, since most of the connected time is already required for energy delivery. Figure 8 shows the session utilization of the two groups. In numerical terms, Group 1 has an average utilization of 0.611 with an average STD of 0.104. Group 2 has an average utilization of 0.598 with average STD of 0.137.

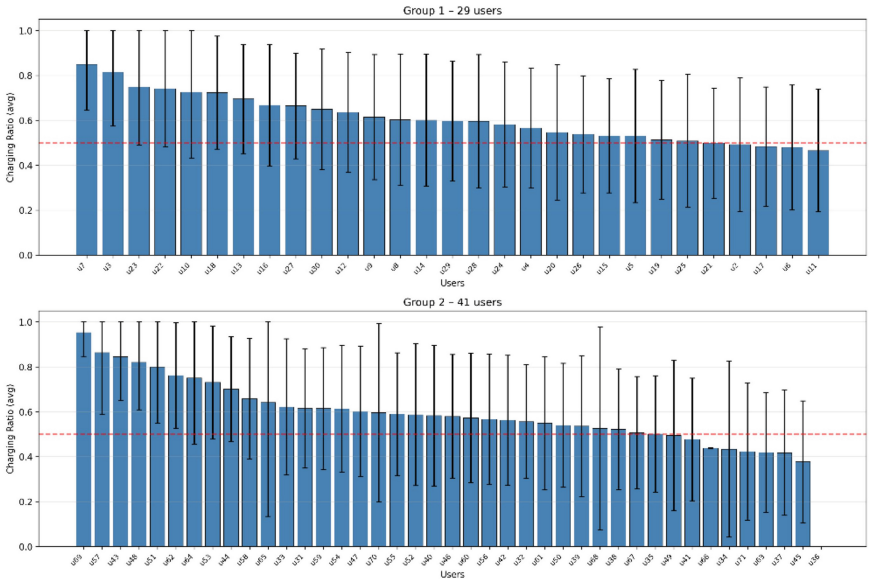


Fig. 8. Session utilization

5 Discussion and Conclusion

This paper analyzed EV charging behavior in an office environment using long-term session data from AC chargers divided into two independent user groups. The results show clear workplace-related temporal patterns: charging activity is concentrated on weekdays, with a dominant morning peak and a smaller midday peak, while weekend session counts remain low. Differences between the two groups were also identified, particularly in energy demand, session duration, and charger loyalty, indicating that even within the same facility, charging behavior can vary significantly. Although global averages provide a useful overview for infrastructure planning, the high standard deviations observed across session count, energy demand, and session duration indicate substantial heterogeneity and intra-user variability. This makes purely aggregated (group-level) prediction unreliable for priority-based scheduling, therefore, the study suggests that user grouping is necessary [9, 13], and then these groups should be used as the basis for forecasting and priority-oriented load management strategies. These findings are consistent with earlier workplace and campus studies, which report weekday-dominant charging behavior and daily peaks linked to office routines. In addition, the observed variability aligns with findings from clustering-based studies, indicating that EV users fall into distinct behavioral groups rather than forming a homogeneous population. Compared to prior work, this study strengthens these results using long observation periods and a multi-charger office setting. Charger loyalty examination is essential for obtaining an overview of charger utilization. It provides crucial information for determining the required number of chargers (e.g., whether more or fewer chargers are needed). Overall, the results highlight the importance of incorporating user-level data into workplace

charging models to support more accurate forecasting and more efficient, priority-based infrastructure and load management.

Future work will apply clustering methods to identify distinct user groups based on charging behavior. These groups can be used to improve system stability and support more effective, priority-based load management in real-world systems.

References

1. de Quevedo, P.M., Muñoz-Delgado, G., Contreras, J.: Impact of electric vehicles on the expansion planning of distribution systems considering renewable energy, storage, and charging stations. IEEE Trans. Smart Grid. **10**(1), 794–804 (2017)
2. Van den Berg, M.A., Lampropoulos, I., AlSkaif, T.A.: Impact of electric vehicles charging demand on distribution transformers in an office area and determination of flexibility potential. Sustain. Energy Grids Netw. **26**, 100452 (2021)
3. Acharige, S.S.G., et al.: Grid integration of electric vehicles—impact assessment and remedial measures. J. Power Sources. **650**, 236697 (2025). https://doi.org/10.1016/j.jpowsour.2025.236697
4. Open Charge Alliance: Open charge point protocol. https://openchargealliance.org/protocols/open-charge-point-protocol/. Last accessed 8 Feb 2026
5. Jonas, T., Daniels, N., Macht, G.: Electric vehicle user behavior: an analysis of charging station utilization in Canada. Energies. **16**(4), 1592 (2023)
6. Szucs, I., Kopják, J., Sebestyén, G., Wendler, M.: Analyzing EV users' charging patterns for estimating session parameters and optimizing load management. In: 2025 IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 521–526 (2025)
7. Szucs, I., Takacs, M., Kopjak, J., Sebestyen, G.: Impact of weather conditions on electric vehicle charging behavior in an office context in Hungary. Acta Polytech. Hung. **22**(12), 159–170 (2025)
8. Robinson, A.P., Blythe, P.T., Bell, M.C., Hübner, Y., Hill, G.A.: Analysis of electric vehicle driver charging behaviour and use of charging infrastructure. In: 18th World Congress on Intelligent Transport Systems (2012)
9. Nazari, Z., Musilek, P.: Uncovering patterns in electric vehicle charging: a data-driven approach. In: 2024 IEEE Power & Energy Society General Meeting (PESGM), pp. 1–5 (2024). https://doi.org/10.1109/PESGM51994.2024.10688611
10. Hovet, S., Farley, B., Perry, J., Kirsche, K., Jerue, M., Tse, Z.T.H.: Introduction of electric vehicle charging stations to university campuses: a case study for the University of Georgia from 2014 to 2017. Batteries. **4**(2), 27 (2018). https://doi.org/10.3390/batteries4020027
11. Pandas: Python data analysis library. https://pandas.pydata.org/about/index.html. Last accessed 4 Feb 2025
12. Martinoz, C.F., Haziza, D., Beaumont, J.-F.: A method of determining the winsorization threshold, with an application to domain estimation. Surv. Methodol. **41**(1), 57–78 (2015)
13. Khaleghikarahrodi, M., Macht, G.A.: Patterns, no patterns, that is the question: quantifying users' electric vehicle charging. Transp. Policy. **141**, 291–304 (2023). https://doi.org/10.1016/j.tranpol.2023.07.020

Digital Green Campus Interventions: Conditions Under Which AI and IoT Generate Educational Value Alongside Operational Sustainability

Imre Tóbel[1,2](✉)

[1] University of Dunaújváros, Institute of Informatics, Dunaújváros, Hungary
tobeli@uniduna.hu
[2] Eszterházy Károly Catholic University, Doctoral School of Education, Eger, Hungary

Abstract. Digital Green Campus initiatives increasingly use artificial intelligence and the Internet of Things to reduce energy demand and emissions in higher education. Operational impacts are frequently reported, but educational impacts are inconsistently measured and often inferred rather than evidenced. The study asks when digital Green Campus interventions produce dual outcomes: measurable sustainability performance improvements and demonstrable learning outcomes such as sustainability competencies, data literacy, systems thinking, and student agency. An evidence informed synthesis of recent peer reviewed journal articles, book chapters, and conference proceedings applies mechanism focused pattern matching. Studies are coded by student role, data interpretability and accessibility, implementation maturity, governance and readiness conditions, and the explicitness of educational measurement. Convergent patterns indicate that educational value is most likely when infrastructure is embedded in living lab and project based learning, co creation links students with faculty and campus operations, and governance enables ethical data access, cross unit collaboration, and continuity beyond pilot phases. Eco feedback shows short term effects, but persistence is rarely tested. The synthesis proposes a dual outcomes evaluation frame and identifies priorities for longitudinal and more comparable learning measurement.

Keywords: green campus · smart campus · Internet of Things · artificial intelligence · sustainability education · living lab · dual outcomes

1 Introduction

Universities are increasingly engaging with sustainable campus operations, with a reliance on digital infrastructures, smart metering, building automation, analysis, and decision support systems, in achieving decarbonisation and resource efficiency. Systematic reviews of campus operations literature show that there is a strong focus on operational sustainability themes, with leadership, governance, stakeholder engagement, and technological innovation playing a key role, with evidence of a lack of consistent reporting [1].

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 395–404, 2027.
https://doi.org/10.1007/978-3-032-32104-6_34

At the same time, there is an educational mandate in higher education institutions to develop competencies that align with sustainable development, including systems thinking, anticipatory thinking, normative reasoning, collaboration, and integrated problem-solving. Digital Green Campus systems have clear potential in achieving these competencies, but there is no guarantee of educational value. There is potential for campus automation systems to improve campus performance, making sustainability processes less accessible to learners, with data inaccessible or uninterpretable. Research on campus energy management systems, which links educational performance, has shown that educational value is dependent on how it is integrated with learning processes and how it is measured [6]. Figure 1 summarizes the dual outcomes logic and the three enabling conditions required for educational value in Digital Green Campus interventions.

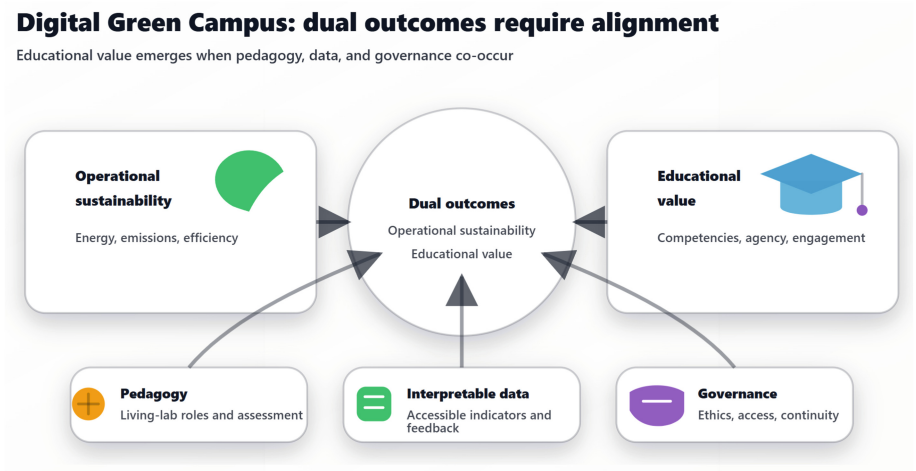


Fig. 1. Two recurring mechanism chains that convert smart campus infrastructure into learning outcomes.

A new stream of literature is also emerging on the relationship between artificial intelligence and institutional Sustainable Development Goals, with a focus on higher education, highlighting that it is implementation, rather than technology, which impacts sustainability [2]. The issue, then, is not that AI and IoT do not have potential benefits for campus operations, but how they can be used as an educational tool as well.

Research aim. The aim is to synthesize the conditions and mechanisms by which AI- and IoT-based Green Campus interventions achieve dual benefits of operational sustainability performance and educational value.

2 Conceptual Framing: Dual Outcomes and Learning Mechanisms

2.1 Dual Outcomes in Digital Green Campus Interventions

Dual outcomes describe the simultaneous change of two dimensions:

a) Operational sustainability performance, such as fluctuations in energy consumption, peak demand, emissions, or documented efficiency improvements as reported via campus systems [4, 6].
b) Educational value, such as fluctuations in competencies, engagement, agency, or learning-related processes like reflection, evidence-based decision-making, or design-oriented problem-solving [3, 19–21].

This dual outcomes model has benefits because many pilots are optimized for operational performance while treating learning as an afterthought. Similarly, educational pilots may be highly effective on a small scale but are often vulnerable without institutional support, integration, and sustainability. This has repeatedly been noted in the discussion of sustainable campus operations and smart campus strategy [1, 9].

2.2 Mechanism Based View

A theoretical perspective grounded in the mechanisms approach is used to explain the reasons why similar technologies lead to dissimilar education outcomes in a variety of contexts. The underlying theory is premised on the idea that the outcomes of the technology are a product of a chain of mechanisms that are dependent on the enabling conditions. Evidence from the campus living labs reveals that the governance structure and organizational contexts influence the sustainability of experimentation and learning in the realms of education, operation, and engagement [16].

Two recurring mechanism chains can be described along with the context under which they operate. In the case of mechanism-chain visualizations, the boxes show each step in a process, while the arrows show an expected progression from one step to another, while the context shows the design requirements required to keep the mechanism in operation. When students face authentic campus issues in an environment where there is defined responsibility, iterative improvements, and decision-making based on evidence, then it is likely to promote systems thinking, data literacy, and sustainable decision making. In practice, this mechanism would manifest itself when students face issues related to campus monitoring, energy management or data design as part of a living lab or project work and have to develop and refine solutions. The requirements for such a mechanism include availability of data to interpret, responsibility boundaries set, alignment of assessments and actions needed, and operational partnership connecting pedagogy with campus management. On the other hand, the mechanism of data visibility, reflection, and actions involves making feedback data visible and interpreting it with sensemaking processes linked directly to feasible actions. This is often associated with enhanced awareness, creation of social norms, behaviors, and efficiency practices. In practice, this mechanism occurs when there is some kind of dashboard, classroom display or ecological feedback system available and meaningful because of benchmarking and interpretation of data with feasible actions.

The first chain of mechanisms is typical of the living labs and project-based approaches, as they are of the integrated campus sustainability initiatives [4, 18]. The second chain of mechanisms is typical of the eco-feedback and dashboard approaches in smart classrooms and eco-feedback contexts, with the important caveat of the absence of long-term persistence measures [6, 10].

3 Evidence Identification and Synthesis Approach

3.1 Evidence Scope

The present study focuses on scholarly articles published from 2018 until 2025 dealing with the intersections between sustainability on college campuses, smart campus technologies, and educational processes and results in higher education institutions. The inclusion criteria were based on publications that covered any aspect related to the implementation and impact of campus-based AI applications, IoT implementations, smart campuses, and living labs; and, moreover, provided insights into operational, educational, governance-related, and/or implementation results. The exclusion criteria included sources not related to higher education campuses, non-academic papers, as well as purely technical articles without any implications for learning, governance, or implementation. The total number of selected sources was 14. Among others, the evidence base includes integrated living-lab programs at college campuses, energy management IoT case studies, effectiveness analysis of campus energy management and educational achievement, energy efficient classrooms, implementation readiness models, and living lab university governance [1, 4–6, 14, 16].

3.2 Coding Dimensions

Each source was coded using a consistent analytic framework with the following dimensions:

a) student role (occupant, data interpreter, analyst, designer, co-creator);
b) data properties (availability, interpretability, timeliness, granularity);
c) implementation maturity (pilot, course-embedded, campus-embedded, scaled);
d) governance and readiness (leadership, funding stability, data governance, cross-unit collaboration);
e) educational measurement quality (explicit outcomes and instruments vs. implicit claims).

3.3 Synthesis Logic

The use of pattern-matching logic was applied in order to establish the presence or absence of certain combinations of student role, data properties, and governance with respect to educational impact. In this regard, the focus of the analysis was on recurrent patterns rather than on effect sizes; for each study source analyzed, the researcher determined whether or not the intervention used triggered one or two mechanisms, which conditions enabled its operation, and whether or not learning impact was assessed directly or indirectly. It is critical to realize that such an analysis cannot be considered as meta-analysis, since a comparison between different studies would be hard to undertake.

4 Findings: When Do AI and IoT Deliver Educational Value?

4.1 Condition 1: Pedagogical Activation Through Living Lab and Project Based Designs

Across these integrated campus programmes, educational value is seen as being most clearly instantiated when intelligent infrastructure is used as a genuine problem space, and when students are able to contribute to that space as part of any form of analysis, design, or implementation activity [18]. An integrated concept of living lab, focusing on renewable energy, electric vehicles, monitoring, and demand response, makes specific links between campus as experimentation platform, systems, and learning activities [4].

Discipline-specific examples also illustrate this general pattern. In terms of architectural education, AI-enabled green campus development research is seen as being educationally valuable when integrated as part of any design activity, as opposed to being seen as passive exposure to automation [13].

The main implication of these examples is that educational value is seen as being dependent on epistemic participation and responsibility, as opposed to any level of sophistication in terms of automation.

4.2 Condition 2: Interpretable and Accessible Data as a Pedagogical Resource

The case studies of IoT-based energy management in IoT environments suggest that data availability and intelligibility play a role in the level of student engagement in evidence-based inquiry and decision-making. When data is available and intelligible, it can support both management and educational activities, especially when it is relevant to course work [5, 23].

The link between ICT-based campus energy management systems and educational performance highlights the need to measure and make intelligible these systems, as enhanced performance does not necessarily imply educational benefit [6, 22].

Interventions in classroom energy efficiency similarly highlight the need to measure and make intelligible energy efficiency, as educational benefit is difficult to infer without directly measuring it [7].

4.3 Condition 3: Governance, Readiness, and Continuity Beyond Pilots

Readiness and governance conditions are found to be critical moderators of potential in terms of creating and sustaining educational value. The process of smart campus vision development reveals opportunities and challenges that are associated with integration, interoperability, and institutional conditions, which are critical factors that may affect whether infrastructure can be considered as an educational asset [9].

Adoption modeling in terms of higher education energy management contexts reveals that organizational and perceptual factors are critical, suggesting that educational integration will not be achieved without organizational commitment and structures. This also supports the conclusion that educational value requires organizational design, not pedagogical design [14].

Findings from governance-related studies on university living labs suggest that institutional structures, cultures, and practices can facilitate or hinder interdisciplinary experimentation in terms of research, education, operations, and engagement, further supporting the conclusion that educational value requires organizational design, not pedagogical design [16]. Thus, transferability will be subject to variation due to differences in the size of the institution, the availability of resources, data-governance capabilities, and regulation. The larger and resourceful universities will offer greater opportunities for cross-departmental coordination and analytics capacity, while smaller institutions will rely on smaller pilot studies, simpler dashboards, and course partnerships.

4.4 Eco Feedback and Smart Classroom Systems: Promising, But Limited Evidence on Durability

IoT-based smart classroom systems are often seen as playing a role in sustainability on campus, mainly in terms of automation and monitoring features. However, the learning potential in terms of sustainability literacy depends on whether feedback can be translated into reflective thinking, decision-making, and action [19, 21]. There is little evidence on long-term effects of cue removal in campus sustainability technology, as many studies use short evaluation windows. Suitable longitudinal approaches would include repeated-measures follow-up across semesters, delayed post-tests after dashboard or cue withdrawal, cohort comparisons between exposed and non-exposed groups, and triangulation between metering logs and reflective or project-based learning artefacts.

5 Discussion: Design, Evaluation, and Policy Implications

5.1 Treating AI and IoT as Educational Infrastructure

The underlying suggestion is that artificial intelligence (AI) and the Internet of Things (IoT) ought to be considered as educational infrastructure in cases where dual results are desired. This line of argument implies that decisions on dashboards, explanations of anomalies, data levels of detail, and rights to access are also considered to be educational decisions [23]. Research on AI-based sustainability education also emphasizes the point that the value depends on purposeful integration rather than on general use [17].

5.2 Data Governance and Ethics as Enabling Conditions

The opening of operational data in education evokes concerns about privacy, security, and fairness. Ethical and responsible considerations in AI-enabled higher education are key, as they will determine the viability of student access to data [15]. In terms of operational aspects, it is important to consider what data can be used, how it can be used, and who is responsible [16]. Figure 2 provides a compact evaluation dashboard for reporting dual outcomes with minimum viable indicators. For AI and IoT campus projects, illustrative metrics may include HVAC-related energy reduction, peak-load avoidance, anomaly-detection response time, dashboard consultation frequency, quality of student interpretations, and the share of student proposals taken up by campus operations.

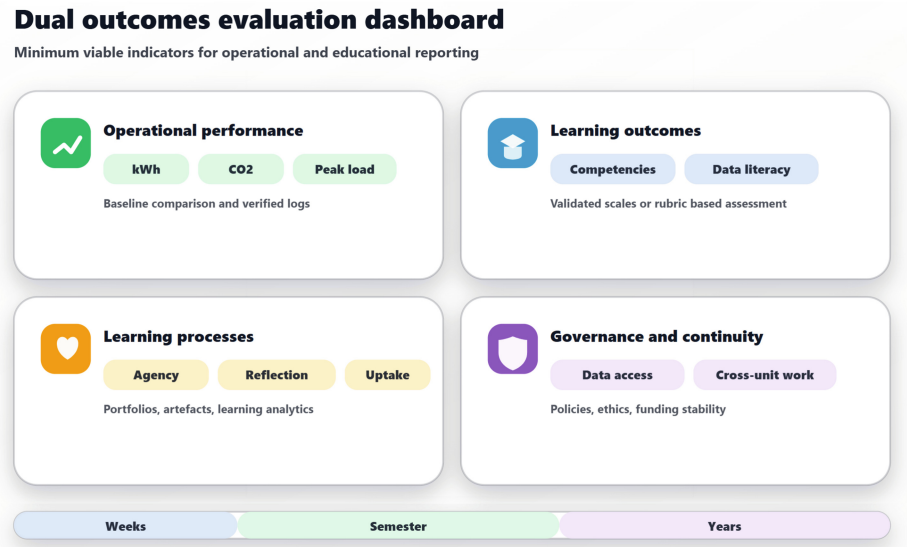


Fig. 2. Dual outcomes evaluation dashboard with minimum viable indicators and time horizon.

5.3 A Dual Outcomes Evaluation Framework

To avoid making assumptions, dual results should be measured by using parallel indicators with corresponding time horizons. Figure 3 visualizes the two dominant mechanism chains observed across the evidence base: the nodes represent successive learning-generating stages, the arrows represent the expected progression between them, and the surrounding design conditions indicate the institutional and pedagogical requirements needed for the mechanism to remain active (Table 1).

This approach is in line with the governance concept, which states that the success of living lab initiatives will depend on the institutional context in terms of sustainability beyond single-case applications. It also responds to reporting gaps as identified in a broader analysis of campus sustainability.

Table 1. Mechanism chains and design conditions

Evaluation domain	Minimum indicators	Recommended measurement approach	Typical time horizon
Operational sustainability performance	Energy use or peak reduction, emissions proxy, efficiency indicators, system uptime	Smart metering and analytics logs, verification against baselines	Semesters to multi year
Learning outcomes	Sustainability competency gains, data literacy, systems thinking, decision quality	Validated scales aligned to sustainability competencies, rubric based assessment of projects	Semesters with follow up
Learning processes	Student agency, participation intensity, reflection quality	Learning analytics plus qualitative artefacts, reflective prompts, portfolio evidence	Within course and follow up
Governance and continuity	Cross unit collaboration, data access protocols, funding continuity	Document analysis, stakeholder interviews, process indicators	Multi year

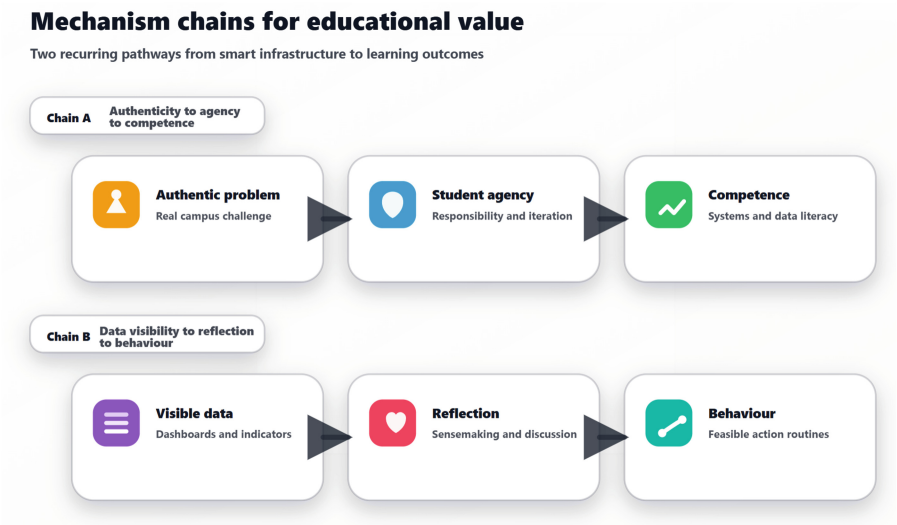


Fig. 3. Two recurring mechanism chains that convert smart campus infrastructure into learning outcomes.

6 Limitations and Research Gaps

Evidence is limited by heterogeneous study design, inconsistent learning outcomes measurement, and the use of short-term evaluation windows. There is also the issue of geographic concentration and clustering of disciplines, which is similar to the broader findings from campus sustainability research. Longitudinal study design with follow-up measurements, a set of competency indicators, and alignment with sustainability learning objectives, and a comparative multi-campus approach with tests of governance and readiness mechanisms, will be important for future studies.

7 Conclusion

Digital Green Campus interventions do not have inherent educational value. Educational value is most likely to be created when three factors converge: pedagogical activation through the use of living labs and projects, data that is interpretable and easily used to inform decisions, and governance that can facilitate ethical data use and collaboration. The strategic implication is that the educational return on investment in AI and IoT is just as much a function of governance and pedagogy as it is of technology. A forward-looking agenda is therefore to move beyond pilot logics toward institutionally embedded implementations, more comparable dual-outcome reporting, and longitudinal evaluation capable of testing whether changes in reflection, agency, and competence persist beyond the immediate intervention period.

References

1. Oliveira, M.C., Proença, J.: Sustainable campus operations in higher education institutions: a systematic literature review. Sustainability. **17**, 607 (2025)
2. Filho, W.L., et al.: Using artificial intelligence to implement the UN sustainable development goals at higher education institutions. Int. J. Sustain. Dev. World Ecol. **31**(6), 726–745 (2024)
3. Rieckmann, M.: Education for Sustainable Development Goals: Learning Objectives. UNESCO (2017)
4. da Silva, L.C., et al.: Sustainable campus model at the University of Campinas Brazil: an integrated living lab for renewable generation, electric mobility, energy efficiency, monitoring and energy demand management. In: Towards Green Campus Operations: Energy, Climate and Sustainable Development Initiatives at Universities, pp. 457–472 (2018)
5. Yasuoka, J., et al.: IoT solution for energy management and efficiency on a Brazilian university campus–a case study. Int. J. Sustain. High. Educ. **24**(2), 426–448 (2023)
6. Zhai, X., et al.: Efficiency of an ICT-based campus energy management system on educational performance: a perspective of a fuzzy DEA approach. Int. J. Environ. Pollut. **70**(3–4), 111–129 (2021)
7. Patil, G.N., Tanavade, S.S.: Eco-friendly energy efficient classrooms and sustainable campus strategies: a case study on energy management and carbon footprint reduction. Int. J. Energy Econ. Policy. **14**(3), 188–197 (2024)
8. Laiolo, P., et al.: The University of Genoa Savona Campus sustainability projects. In: 6th International Forum on Research and Technology for Society and Industry, pp. 115–120 (2021)

9. Eltawil, A., Mostafa, N.A., Matsushita, Y.: Toward a smart and sustainable campus: future vision, opportunities, and challenges. In: Resilient and Responsible Smart Cities, vol. 1, pp. 233–242 (2021)
10. Revathi, R., Suganya, M., Gladiss Merlin, N.R.: IoT based cloud integrated smart classroom for smart and a sustainable campus. Procedia Comput. Sci. **172**, 77–81 (2020)
11. Mahmood, S., et al.: Raspberry PI and role of IoT in education. In: 4th MEC International Conference on Big Data and Smart City, pp. 1–6 (2019)
12. Kesuma, M.E.K., Kesuma, G.C., Taher, A.: Internet of things to realize education in Industry 4.0 based on sustainability environment. E3S Web Conf. **482**, 5008 (2024)
13. Cudzik, J., Nyka, L., Szczepański, J.: Artificial intelligence in architectural education-green campus development research. Glob. J. Eng. Educ. **26**, 20–25 (2024)
14. Ayyash, M.M., Salah, O.H.: AI adoption in higher education: advancing sustainable energy management in Palestinian universities. J. Open Innov.: Technol. Mark. Complex. **11**(2), 100534 (2025)
15. Khan, S., et al.: Harnessing AI for sustainable higher education: ethical considerations, operational efficiency, and future directions. Discov. Sustain. **6**(1), 23 (2025)
16. Hadfield, P., et al.: Governing University Living Labs for sustainability transformations: insights from 18 international case studies. Sustain. Sci. **20**, 1753–1774 (2025)
17. Filho, W.L., et al.: Using artificial intelligence in sustainability teaching and learning. Environ. Sci. Eur. **37**(1), 124 (2025)
18. Kovari, A.: Project-based teaching supported by cost-effective IT tools. In: New Ways of Learning, pp. 273–284. HERA (2017)
19. Gogh, E.: Metacognition and lifelong learning: a survey of secondary school students. In: 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 271–276 (2018)
20. Gogh, E.: Examining the relationship between lifelong learning and language learning in a vocational training institution. J. Appl. Tech. Educ. Sci. **8**(1), 52–69 (2018)
21. Kovari, A.: Experiences of self-regulated learning in a vocational secondary school. J. Appl. Tech. Educ. Sci. **9**(2), 72–86 (2019)
22. Balogh, Z., et al.: Development of available IoT data collection devices to evaluate basic emotions. Acta Polytech. Hung. **19**(11), 165–184 (2022)
23. Farkas, I., et al.: Wireless sensor network protocol developed for microcontroller based Wireless Sensor units, and data processing with visualization by LabVIEW. In: 12th International Symposium on Applied Machine Intelligence and Informatics, pp. 95–98 (2014)

AI-Generated Single-Use Educational Applications: The Process of Creation, Publication and Quality Verification

Petr Voborník(✉)

Department of Technics and Informatics, Faculty of Education, University of Hradec Králové, Hradec Králové, Czech Republic
petr.vobornik@uhk.cz

Abstract. Generative artificial intelligence has lowered the barriers to creating small educational web applications focused on a single learning objective. This article addresses how such applications can be created, published, and reviewed before classroom deployment. The study is framed as design science research, in which the main contribution is a lightweight pre-publication verification procedure supplemented by illustrative artefacts. The article outlines an AI-supported workflow for generating one-off applications and proposes a repeatable process involving specification, generation, local verification, and publication. A checklist synthesised from software quality, accessibility, and security frameworks is presented and demonstrated on four selected applications created in different tools and with varying deployment requirements. The contribution lies not in proposing a new learning theory, but in linking established didactic principles of short-term practice with AI-assisted development workflows and a practical minimum of technical quality control. The article shows that even simple AI-generated applications can be deployed more safely and reliably if they are systematically reviewed for functional correctness, usability, accessibility, privacy, security, and operational suitability.

Keywords: Generative artificial intelligence · single-use teaching applications · design science research · quality verification · web applications · accessibility · security · vibe coding

1 Introduction

Digital teaching aids have long taken the form of presentations, worksheets, videos, or specialised teaching programmes. However, with the development of generative artificial intelligence, the availability of one specific category is changing: small interactive applications that can be created quickly, often without a full-fledged development background, and used for a single specific purpose. Typically, these involve practising unit conversions, vocabulary, factual knowledge, short calculations or simple algorithmic procedures. In this article, we refer to these applications as single-use teaching applications. They are not intended to compete with large LMS or e-learning courses, but rather

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 405–419, 2027.
https://doi.org/10.1007/978-3-032-32104-6_35

to serve as a supplement that teachers can prepare ad hoc according to the current needs of the class. Similarly, in the Czech environment, the impact of specific technologies on teaching has been monitored for a long time, for example, computer measurement systems in physics [1] or the use of robotic kits and 3D printing in primary schools [2].

At the same time, however, there are risks. AI-generated code may contain errors, inappropriate UX, accessibility weaknesses, or security flaws. In a school environment, there is also the issue of personal data protection and content appropriateness. Therefore, it makes sense to describe not only how to quickly create applications, but also how to systematically verify them before publication. The research goal of this article is therefore to combine the practical creation of these applications with a simple, repeatable quality verification methodology that can be used as a pre-publication check.

The contribution of this article does not lie in proposing a new learning theory. The theoretical principles underlying short practice sessions, such as the testing effect, spaced learning, and microlearning, are already well established. The novelty of this work lies in linking these didactic principles with the current ability to rapidly generate small educational applications using generative AI, and in formulating a lightweight pre-publication verification procedure applicable in everyday educational practice. The research objective is therefore twofold: to describe a repeatable process for creating and publishing such applications and to demonstrate, using selected examples, how their technical quality can be verified prior to deployment.

2 Theoretical Basis

2.1 Generative AI in Education

Large language models are used in education in a number of roles, from supporting teachers in preparing materials to assisting pupils in explaining the curriculum. An overview of the opportunities and risks is summarised, among other things, in works focusing on the impact of language models in education, including the requirement for critical verification of outputs and ethical use [3]. Recommendations from international institutions are also relevant for school practice. For example, UNESCO emphasises the need for transparency, data protection and the development of digital competences when introducing generative AI into teaching [4]. The European context is complemented by ethical guidelines for educators, which highlight the risks of bias, lack of transparency and inappropriate data handling [5]. These principles are reflected in the article in the section on quality verification and safe publication of artefacts.

2.2 Microlearning and Practice

One-off educational applications often support short, repeated practice. This is consistent with findings from cognitive psychology. The testing effect shows that actively retrieving information from memory leads to better long-term retention than simply rereading [6]. At the same time, spaced learning, where practice episodes are spread out over time, is important, as confirmed by an extensive meta-analysis of the spacing effect [7, 8]. This is also related to the concept of microlearning, where the curriculum is divided into short

modules and practice can be frequent and situationally available, with systematic reviews showing positive effects on motivation and performance, especially in combination with digital technologies [9]. For small applications, this means that even a simple mechanism of random tasks, immediate feedback and continuous progress saving can be didactically beneficial, especially if the selection of tasks is controlled to minimise duplication and promote variability in practice [10, 11].

2.3 The Quality of Software Artefacts in a School Context

Although these are small applications, it is advisable to consider the quality of the artefact systematically. A general framework is provided, for example, by the ISO IEC 25010 quality model, which includes functional suitability, usability, reliability, security, portability and other characteristics [12]. For web applications, accessibility is also key, which is specified in the international context by the WCAG 2.2 standard [13]. The minimum security requirements can be related to OWASP Top 10, which summarises the most common risks of web applications, such as injection, poor access control, or configuration errors [14]. For quick and comparable checks, it is advisable to use automated audits in *Lighthouse*[1] [15] and accessibility checks using *WAVE Web Accessibility Evaluation Tools* [16], or supplement them with simple automated runs in *Playwright* [17]. The article does not cover these frameworks in their entirety, but uses them as a source of structure for a practical quality verification checklist.

3 AI Tools That Can Be Used to Create One-Off Applications

In practice, there are two main approaches to quickly creating small educational applications. The first is direct code generation in a conversational interface, where the model creates HTML, CSS, and JavaScript, or variants in Blazor, React, or another framework. The second approach is AI-assisted development in an IDE, where the model continuously assists with design, completion and refactoring, while the developer has standard control over the project and versioning.

The most commonly used AI tools include *ChatGPT*, *Gemini* and *Claude*, including their specialised modes and models designed for working with code, i.e. variants focused on generating, editing and debugging source code. Recently, the concept of an interactive workspace with live preview has been gaining ground, where the generated code or document is displayed alongside the conversation as a separate output that can be continuously edited, run, and immediately verified both visually and functionally.

In the *Claude* environment, this approach is implemented by the Artifacts feature, which also supports easy publishing and sharing of the resulting applications [18]. *Canva* offers similar functionality, where interactive output can be published directly as a link, without the need for hosting. *Claude* also allows you to create so-called AI-powered Artifacts, where the resulting application uses *Claude's* artificial intelligence during use at no cost to the author. AI consumption is deducted from the user's quota (in the Free version, it is renewed after approximately 5 h).

[1] Lighthouse: https://pagespeed.web.dev.

GitHub Copilot, integrated into *Visual Studio* or *Visual Studio Code*, is important for developer-oriented workflows. Currently, *Copilot* allows users to choose from multiple models from various providers, including specialised Codex variants designed for working with source code [19]. This reduces the gap between what a conversational tool can do and what can be comfortably developed in an IDE.

In addition to these main tools, agent editors and application generation environments such as *Antigravity*, *Cursor* and *Codex* are rapidly gaining ground. These are based on the idea that AI not only complements code, but can also plan steps, edit multiple files, and, to a certain extent, act as a development agent. For non-programmers, services such as *Replit Agent*, *bolt*, and *v0* are also interesting, as they can create an application prototype from a description and deploy it in a running environment. These tools are mentioned here more for context, because they show that one-off applications are becoming an accessible category for an ever-wider group of teachers, and at the same time, the need for a simple verification process is growing.

In practice, this method of working is often referred to as *vibe coding*. It is an approach where the intent of the application is formulated in natural language and most of the implementation is generated by an AI tool. The developer controls development primarily through prompting and verifying behaviour, rather than writing detailed code [20].

4 Methodology, Artefact and Research Framework

The article is guided by the methodology of *design science research*. In this paradigm, the main result of the research is an artefact, such as a method, procedure, model or software, and its verification in relation to the problem being solved is key [21]. In the information systems (IS) tradition, design science is understood as a complement to natural science because, in addition to describing reality, it provides proposals and evaluations of effective solutions [22]. The DSRM methodology summarises the practical application of design science, which includes problem identification, solution goal design, design and development, demonstration, evaluation, and communication of results [23]. Recommendations on how to emphasise the contribution of an artefact and its evaluation are also relevant for the publication structure of design science articles [24]. For a broader methodological context, Wieringa's approach, which links the design cycle with empirical verification in the context of use, can also be mentioned [25].

In this article, the term "*artefact*" refers both to selected sample applications and, above all, to a repeatable process for verifying their quality prior to deployment, whereas "*Artifacts*" refers specifically to *Claude's* product feature for creating and displaying standalone outputs. Evaluation is understood here as a technical pre-publication check, not as an experiment involving respondents. The goal is to ensure that the published application is functionally correct, usable, and sufficiently secure given its intended educational purpose. This approach is suitable in situations where it is necessary to responsibly publish a practice-oriented output, even if there is no opportunity for user studies prior to its release. The four applications presented below were deliberately selected to cover various AI tools and different deployment scenarios. They serve as illustrative examples of the proposed procedure, not as a representative sample or a statistically grounded comparison of AI tools.

5 The Process of Creating and Publishing One-Off Applications

The creation of a one-off application can be structured into several steps that are repeated across the AI tools used.

First, it is advisable to define the learning objective and scope. With one-off applications, it is crucial not to succumb to the temptation to add secondary functions. A single sentence works well, for example, practise converting units of length from millimetres to kilometres, or practise irregular verbs using flashcards.

This is followed by a proposal for minimal interaction. For most similar applications, a simple cycle of task, response, check, next task is sufficient. If the application saves progress, it is advisable to use local storage so that a server is not needed. This also reduces security risks and operational demands.

The third step is the actual generation. In conversational tools, it is advisable to specify the desired output as a single-file page, i.e. a single HTML file with embedded CSS and JavaScript, as this facilitates both publishing and auditing. In the IDE workflow, it is worth defining the project structure and requesting small, controllable changes. In both cases, it is useful to explicitly state accessibility requirements in the prompt, such as keyboard control, element labels, and contrast.

The fourth step is local verification and editing, which is the core of this article. Only then does publishing follow. In practice, there are several ways to do this for one-off applications. Some platforms allow direct hosting, for example for Artifacts in *Claude* or publishing from *Canva*, which is fast and requires no infrastructure of your own. An alternative is static hosting, such as your own school or institutional website, or *GitHub Pages*, where you can publish static pages directly from the repository [26]. In a LMS environment, it is often possible to upload an HTML file as study material, e.g. in *Moodle* in the form of a file or page, which allows students to run the application within the course without external accounts [27], or it can be integrated via an external web tool [28]. The choice of hosting is reflected in data protection and availability requirements.

6 Quality Verification Prior to Publication

The main practical goal of the article is to propose a verification methodology that can be applied to any AI-generated one-off application before its publication. The procedure is designed to be feasible for a teacher or developer without special measurement research, while reflecting basic quality and security frameworks.

6.1 Functional Correctness

Functional verification should be performed using a combination of manual scenarios and simple test data. For applications such as unit conversions, it is necessary to verify boundary values and randomly generated tasks. For cards, it is important to verify data import and the subsequent behaviour of the application during normal use, for example, whether items are randomly selected correctly, whether the sides of the cards are alternated correctly, whether the number of remaining cards is calculated correctly, and

whether the function for discarding already learned items actually affects further selection. At the same time, it is advisable to test empty, incomplete, or invalid input so that the application does not crash and clearly guides the user to complete the data.

It is recommended to write a set of at least ten test scenarios covering typical and error states, and to always go through these scenarios after each major modification or generation of a new version.

6.2 Usability and Didactic Suitability

Simplicity is key for one-off applications. A practical criterion is whether the user understands what to do without reading the instructions. From a didactic point of view, immediate feedback, the possibility of repetition and, ideally, an element of motivation, such as a success counter or a simple goal, are important. This is supported in particular by the concept of the testing effect [6] and spaced learning [7]. In practice, it has proven useful when the application allows for short practice sessions and at the same time provides the option to continue later, for example by saving the status in local storage.

6.3 Accessibility

In a school environment, accessibility should be treated as a basic requirement, as the application may also be used by users with specific needs. The minimum set of checks is based on WCAG 2.2 [13]. In practice, this means verifying keyboard control or, conversely, touchscreen-only control, responsiveness, visible focus, form element labels, reasonable tab order, sufficient contrast, and understandable error messages. If possible, it is advisable to use an automated audit in the browser or an extension that alerts you to common errors, followed by a manual check, ideally on several different devices.

6.4 Security and Data Protection

Single-use applications often do not require a server and can also work offline. This is advantageous because it can minimise the processing of personal data. If the application does not send any data to a server, this behaviour should be maintained and verified, and the application should be checked for unexpected external scripts or tracking elements. If an external API is used, such as an AI service, keys, limits, and data processing must be explicitly addressed. Minimum security checks can be structured according to the OWASP Top 10 [14], in practice, this means, in particular, careful handling of user input, avoiding dangerous constructs such as *eval*, and not transmitting sensitive data in URLs. If client-side storage is sufficient, browser storage can reduce unnecessary data transfer, but only non-sensitive data should be stored there, and such storage should not be considered completely secure.

6.5 Operational Suitability and Publishing

For static applications, it is important to verify that they run without errors in common browsers, do not require installation, and load correctly on mobile devices. It is also

advisable to check the size of files and dependencies, as connection speeds may be limited at school. When publishing via *GitHub Pages*, it is necessary to verify that the repository does not contain sensitive data, as the pages and source codes are public, and that the licence and author are clearly stated [26]. When publishing in LMS, it is advisable to verify how the system handles scripts and whether the page runs in a secure context [27]. If the application is hosted directly by an artefact-type platform, it is necessary to take into account the terms of service and long-term availability.

6.6 Quality Verification Methodology

The criteria in Table 1 are based on the quality, accessibility, and security frameworks in Sect. 2.3. The table was not taken from a single source, but was compiled by the author as a synthesis of recommendations for software quality (ISO/IEC 25010) [12], web accessibility (WCAG 2.2) [13], and minimum security requirements for web applications (OWASP Top 10) [14], supplemented by readily available measurement and audit methods (*Lighthouse* [15], *WAVE Web Accessibility Evaluation Tools* [16], and, where applicable, *Playwright* [17]).

The Table 1 thus serves as a practical pre-publication quality verification procedure. To ensure comparable results, it is advisable to perform measurements in the cleanest possible environment, ideally in an anonymous window without extensions, because *Lighthouse*, for example, is sensitive to device configuration, extensions and stored data in the browser, and results from different machines cannot be directly compared [15].

Tables 1 and 2 use the following result markings: ✓ fulfilled without major reservations, ~ partially fulfilled (application is usable, minor adjustments recommended), ✗ not fulfilled (correction recommended before publication),? not yet verified.

6.7 Measurement Conditions and Tools

To ensure transparency and reproducibility, the tests underlying Tables 1 and 2 were conducted under the following conditions. *Lighthouse*-based scores (Performance, Accessibility and Best Practices) were obtained via *PageSpeed Insights* in desktop mode across three runs, and the median value was recorded. Where the difference between desktop and mobile was significant, both values are reported. Manual browser-based checks were performed in *Microsoft Edge 146* and *Google Chrome 146* on *Windows 11*, and in their mobile versions on *Android 16*, using InPrivate/incognito mode without extensions. Criterion A3 (form element labels and related accessibility markup) was assessed using the *WAVE Web Accessibility Evaluation Tools*, supplemented by manual inspection. Manual accessibility checks also included keyboard-only walkthroughs, focus checks, responsive layout verification and touch controls on the device used. Runtime errors and network requests were monitored in the browser's *DevTools* on the desktop. For applications using an external AI API, only observable client-side behaviour and request patterns were assessed. A full-scale security audit of the server-side was beyond the scope of this article.

Table 1. Table for quality verification of typical AI-generated applications.

ID	Area	Criterion	How to verify	Target
F1	Functionality	Basic scenario, task completion	Manual walkthrough of the main flow	✓
F2	Functionality	Boundary inputs, error states	Enter empty, extreme, invalid values	✓ or ~
F3	Functionality	State persistence	Restoration after reload, browser closure	✓
R1	Reliability	Console without errors	DevTools Console, no errors during runtime	✓
R2	Reliability	Recovery after error without loss of work	Trigger error, return without reset	✓ or ~
U1	Usability	Clarity of instructions and controls	Heuristic check, brief instructions in the UI	✓
U2	Usability	Feedback, status, score	Visible status, history, counters	✓ or ~
U3	Usability	Responsiveness mobile, tablet	Width change, orientation, touch	✓ or ~
A1	Accessibility	Keyboard only control	Tab, Enter, arrows, task completion	✓ or ~
A2	Accessibility	Visible focus, logical order	Tab order, focus does not disappear and is not overlaid	✓ or ~
A3	Accessibility	Form element labels	Label, aria label, screen reader, WAVE, manual inspection	✓ or ~
A4	Accessibility	Contrast, target size, touch	Lighthouse Accessibility, manual check	≥ 90 or ~
P1	Performance	Lighthouse Performance	PageSpeed Insights (Lighthouse), Desktop, 3 runs, median	≥ 80
P2	Website quality	Lighthouse Best Practices	PageSpeed Insights (Lighthouse), no critical warnings	≥ 90
S1	Privacy	Network communication, data leakage	Network, no unexpected requests	✓
S2	Security	Inputs and injections, especially in SQL	Negative tests, OWASP thinking	✓ or ~
D1	Deployment	Hosting portability	GitHub Pages, school hosting, LMS file	✓

7 Typical Examples of Applications

This section briefly describes four typical applications that represent different creation tools and different didactic scenarios. All of the applications listed are designed as single-file HTML, CSS, and JavaScript applications, with the exception of the *SQL trainer* application, which is based on a framework solution.

7.1 Flashcards, Canva AI

The *Flashcards* application[2] allows the user to insert or import data in CSV format or enter it manually (see Fig. 1). The application then randomly selects a row and a random column within it, displays the selected value, and the user recalls the remaining values of the same row in their mind. They then display the solution and can continue or discard items they have already mastered. Practice data is stored in local storage, so it is possible to return to the topic without a server account. The application was created in the *Canva AI* environment and then hosted as static page on our own server.

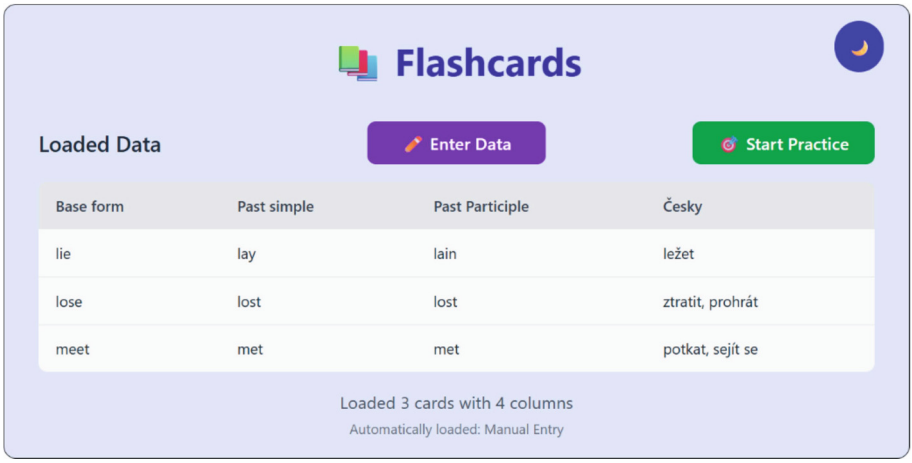


Fig. 1. Sample of the *Flashcards* application generated in *Canva AI*.

7.2 Length Unit Conversions, ChatGPT

The application for *length unit conversions*[3] generates a random input value and requests a conversion between units in the range of millimetres to kilometres. The user enters the result, the application evaluates it and, in case of an error, provides graded help. The aim is to practise routine work with orders and units, a skill commonly used in primary school. The application was created through iterative generation and editing in *ChatGPT* and was published as a static page.

[2] Flashcards: https://programko.net/app/zs/flash-cards.html?l=en.
[3] Unit conversions: https://programko.net/app/zs/prevody-jednotek-delky.html.

7.3 SQL Trainer, Claude AI Artifacts and AI via API

The *SQL trainer* application[4] is designed for practising queries on a modified and shortened *AdventureWorksLT* database[5] (a training database from Microsoft[6]) and focuses on topics common in introductory database courses, such as SELECT, WHERE, JOIN, aggregation, and sorting. The user enters a query, the application runs it on the local database and evaluates whether the resulting table meets expectations. It also includes AI assistance, which provides explanations of parts of the query or points out shortcomings in the solution. This example represents a situation where the application is not purely static but integrates AI via API, which places higher demands on security, key management and operating costs. The prototype was created in *Claude* in *Artifacts* mode, which allows for rapid iteration over interactive output [18]. Deployment can be implemented either directly within the platform or by exporting and then hosting with your own server layer for API calls.

7.4 Schedule Planning, GitHub Copilot in VS Code

The application for *schedule planning*[7] is visually similar to the internal university (UHK) scheduling system and is used to design schedules so that they can be shared without being stored on a server. It was created in *Visual Studio Code* with *GitHub Copilot* support, using a selected model (in this case, *GPT 5.2*) focused on code. It corresponds to a scenario where the application is more domain-specific, involves more complex interactions, and requires more precise control over structure and versioning, which is facilitated by the IDE workflow. *Copilot's* current capabilities allow you to choose from multiple AI models, which can affect the quality of the generated code and the speed of iterations.

7.5 Illustrative Verification Results for the Selected Applications

Table 2 summarises the results of applying the proposed checklist to the four selected applications presented in this section, in accordance with the methodology described in Sect. 6.6 and under the conditions specified in Sect. 6.7. This is not a rigorous comparative evaluation of AI tools, rather, the goal is to demonstrate how this approach can be applied in practice across various development workflows and deployment scenarios. The applications were deliberately selected to represent diverse development environments and various types of educational tasks. The table should therefore be read as an illustrative overview of technical verification, identifying areas that were satisfactory (✓), acceptable with minor reservations (~), or required correction (✗) prior to publication. After deployment to static hosting (e.g. *GitHub Pages* [26]) or within an LMS as

[4] SQL trainer: https://programko.net/databaze/awlt/sql-trenager.

[5] Modified AdventureWorksLT database: https://github.com/PetrVobornik/prednasky/tree/master/Databaze/db/awlt.

[6] Original AdventureWorksLT database: https://learn.microsoft.com/sql/samples/adventureworks-install-configure.

[7] Schedule planning: https://programko.net/app/uhk/rozvrhy/planovani.html.

a file resource [29], it is advisable to repeat the key checks, particularly regarding file handling, browser storage, focus management, and externally loaded resources.

Table 2. Results of quality verification of AI-generated applications.

ID	Criterion	Flashcards	Conversions	SQL training	Timetables
F1	Basic scenario, task completion	✓	✓	✓	✓
F2	Boundary inputs, error states	✓	✓	✓	✓
F3	State persistence	✓	✗	~	✓
R1	Console without errors	✗[a]	✓	✓	✓
R2	Recovery after error without loss of work	✓	✗	~	✓
U1	Clarity of instructions and controls	✓	✓	✓	✓
U2	Feedback, status, score	~	✓	~	~
U3	Responsiveness mobile, tablet	✓	✓	✗	✗
A1	Keyboard only control	✓	✓	✓	✗
A2	Visible focus, logical order	✓	✓	✓	✓
A3	Form element labels	✓	✗[b]	✗[b]	✗[b]
A4	Contrast, target size, touch	100	88	100	98
P1	Lighthouse Performance	92/100[c]	98/100[c]	32/51[c]	100
P2	Lighthouse Best Practices	96	96	73	100
S1	Network communication, data leakage	✓	✓	~	✓
S2	Inputs and injections, especially in SQL	✓	✓	✓	✓
D1	Hosting portability	✓	✓	✗	✓

Note: The selected applications were used as examples to demonstrate the proposed method. The results depend on the environment and should not be interpreted as a benchmark of the AI tools used.

[a] *Canva* adds links to internal scripts to function within its environment. However, it does not remove them during export.

[b] The "aria-label" attributes are missing, which the AI used does not add by default and must be specifically requested, as in the case of the *Flashcards* application.

[c] Results for mobile/computer, if different.

7.6 Other Examples in the Blazor Framework

In addition to single-file applications, a similar approach can be used for applications in the *Blazor WebAssembly* framework [30], which can also be hosted as a static website. In practice, this approach has been used to create educational games such as *memory*

game[8] (similar to the mobile version [31]), *jigsaw puzzles,*[9] *crossword*[10] and *quizzes.*[11] An interesting feature is the versatility, where the content of these activities is defined via a JSON file and the application itself can be reused for different scenarios. These applications are listed here as a supplement, showing that a *one-off* teaching activity does not always have to be reprogrammed, but can be created using a single universal tool that is simply filled with different data for a specific topic.

8 Discussion and Summary of Contributions

The examples presented show that one-off educational applications can be created through a variety of AI-assisted workflows, ranging from conversational generation and design-oriented environments to development in AI-supported IDEs. Across these workflows, the obvious advantages are rapid iteration and low barriers to publication. At the same time, however, these advantages increase the risk that technically weak applications may be published prematurely. Therefore, it is useful to have a simple verification process that covers functional correctness, usability, accessibility, security, and operational suitability. This article presents a practical bridge between established didactic principles of short exercises and a structured pre-publication verification process suitable for small AI-generated applications.

The methodology proposed in this article and the checklist used should be viewed as a minimum starting point, not as a comprehensive evaluation framework. It does not replace systematic software testing, unit tests [32], specialized security assessments, or empirical pedagogical evaluation involving students. Similarly, the four applications in Sect. 7 were deliberately selected to cover various tools and deployment scenarios. They serve as illustrative examples, not as a representative sample of all AI-generated educational applications. The results should therefore be interpreted as technically informative, case-based verification outputs, not as statistically generalisable evidence. However, even without this, it is already possible to responsibly publish functional and safe micro-tools for teaching.

Another limitation is that some results depend on the environment in which the measurements are taken. Lighthouse scores, accessibility findings, runtime behaviour, and even the quality of the generated code may vary depending on browser versions, devices, hosting conditions, and subsequent updates to the AI tools used. For applications with a server component or third-party APIs, the proposed procedure should be supplemented with checks of server configuration, key management, logging, and, where applicable, penetration testing. Future work may therefore expand this baseline minimum to include an evaluation of the impact on learning directly in the classroom, comparative studies of prompting and tool selection, and more automated regression testing of generated applications. This direction is also relevant in light of existing research on technology-enhanced instruction, such as a study focused on project-based

[8] Memory game for one player: https://ff.uhk.cz/hk800/aktivity/pexeso/bila-vez/0, to play with the computer, change the number at the end to /1, and for two players, change it to /2.

[9] Jigsaw puzzle: https://ff.uhk.cz/hk800/aktivity/puzzle/kolej.

[10] Crossword: https://ff.uhk.cz/hk800/aktivity/krizovka/kostel-nanebevzeti-panny-marie.

[11] Quiz: https://ff.uhk.cz/hk800/aktivity/karticky-kviz/jezuitska-kolej.

STEM education using educational robotics, which demonstrated a positive impact on the development of student competencies, particularly problem-solving skills [33].

9 Conclusion

This article introduced the concept of AI-generated one-off educational applications that can be quickly created and deployed as static websites or lightweight hosted artefacts, including Claude Artifacts in relevant cases. It described AI-supported workflows for their creation, proposed a repeatable process involving design, generation, local verification, and publication, and demonstrated a practical pre-publication checklist on four selected applications. The results suggest that even very small AI-generated tools can be published more responsibly if they are explicitly checked for functional correctness, usability, accessibility, privacy, security, and operational suitability. The proposed procedure is intentionally lightweight but provides a useful minimal baseline solution for authors who need to deploy educational applications quickly without sacrificing basic quality control.

Acknowledgements. The article was funded by the Faculty of Education, University of Hradec Králové.

References

1. Němec, R., Hubálovská, M.: Computer measuring systems and their impact on the popularity of individual parts of the physics subject in primary schools in the Czech Republic. In: 13th International Scientific Conference on Distance Learning in Applied Informatics (DIVAI 2020), pp. 443–455. Wolters Kluwer, Štúrovo (2020)
2. Němec, R., Voborník, P.: Using robotic kits and 3D printers at primary (lower secondary) schools in the Czech Republic. Int. J. Educ. Inf. Technol. **11**, 68–73 (2017)
3. Kasneci, E., et al.: ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Individ. Differ. **103**, 102274 (2023). https://doi.org/10.1016/j.lindif.2023.102274
4. UNESCO: Guidance for generative AI in education and research. UNESCO (2023)
5. European Commission: Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators. European Commission (2022)
6. Roediger, H.L., Karpicke, J.D.: Test enhanced learning, taking memory tests improves long term retention. Psychol. Sci. **17**(3), 249–255 (2006). https://doi.org/10.1111/j.1467-9280.2006.01693.x
7. Cepeda, N.J., et al.: Distributed practice in verbal recall tasks: a review and quantitative synthesis. Psychol. Bull. **132**(3), 354–380 (2006). https://doi.org/10.1037/0033-2909.132.3.354
8. Voborník, P.: Universal testing environment. Ph.D. thesis, University of Hradec Králové (2012)
9. Ferreira, A., et al.: Contribution of microlearning in basic education, a systematic review. Educ. Sci. **15**(3), 302 (2025). https://doi.org/10.3390/educsci15030302

10. Voborník, P.: Optimization of the method for selection of questions into the repeated electronic test for the purpose of teaching and evaluated testing. In: The 11th International Scientific Conference on Distance Learning in Applied Informatics (DIVAI 2016), pp. 445–455. Wolters Kluwer, Štúrovo (2016)
11. Voborník, P., Němec, R.: Innovative algorithm optimized for multiple rounds and stateless assigning of tasks non-duplicate for the same subjects. AD ALTA J. Interdiscip. Res. **9**(2), 404–408 (2019). https://doi.org/10.33543/0902
12. ISO: ISO/IEC 25010:2023. Systems and software engineering, Systems and software Quality Requirements and Evaluation (SQuaRE), Product Quality Model. Edition 2. ISO (2023)
13. W3C: Web content accessibility guidelines, WCAG 2.2. W3C Recommendation (2024). https://www.w3.org/TR/WCAG22/. Last accessed 22 Jan 2026
14. OWASP: OWASP Top 10:2021. OWASP Foundation (2021). https://owasp.org/Top10/2021. Last accessed 25 Jan 2026
15. Google: Lighthouse. Chrome for developers. Google (2025). https://developer.chrome.com/docs/lighthouse/overview. Last accessed 27 Jan 2026
16. WebAIM: WAVE web accessibility evaluation tools. WebAIM, Utah State University. https://wave.webaim.org. Last accessed 8 Apr 2026
17. Microsoft: Playwright (framework for web testing and automation). GitHub Repository (2026). https://github.com/microsoft/playwright. Last accessed 26 Jan 2026
18. Anthropic: Artifacts are now generally available. Anthropic News (2024). https://claude.com/blog/artifacts. Last accessed 26 Jan 2026
19. GitHub: Supported AI models in GitHub Copilot. GitHub Docs, online reference (2026). https://docs.github.com/en/copilot/reference/ai-models/supported-models. Last accessed 24 Jan 2026
20. Koren, M., Békés, G., Hinz, J., Lohmann, A.: Vibe coding kills open source, arXiv preprint https://arxiv.org/abs/2601.15494, (2026)
21. Hevner, A.R., et al.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004). https://doi.org/10.2307/25148625
22. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decis. Support Syst. **15**(4), 251–266 (1995). https://doi.org/10.1016/0167-9236(94)00041-2
23. Peffers, K., et al.: A design science research methodology for information systems research. J. Manage. Inf. Syst. **24**(3), 45–77 (2007). https://doi.org/10.2753/MIS0742-1222240302
24. Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. MIS Q. **37**(2), 337–356 (2013). https://doi.org/10.25300/MISQ/2013/37.2.01
25. Wieringa, R.J.: Design Science Methodology for Information Systems and Software Engineering. Springer (2014). https://doi.org/10.1007/978-3-662-43839-8
26. GitHub Docs: Configuring a publishing source for your GitHub pages site. GitHub Documentation (2026). https://docs.github.com/en/pages/getting-started-with-github-pages/configuring-a-publishing-source-for-your-github-pages-site. Last accessed 18 Jan 2026
27. Moodle: Working with files. Moodle Documentation (2024). https://docs.moodle.org/501/en/Working_with_files. Last accessed 26 Jan 2026
28. Voborník, P.: Universal testing environment as an external tool of Moodle. In: 10th International Scientific Conference on Distance Learning in Applied Informatics (DIVAI 2014), pp. 215–225. Wolters Kluwer, Štúrovo (2014)
29. Moodle: File resource settings. Moodle Documentation (2024). https://docs.moodle.org/501/en/File_resource_settings. Last accessed 26 Jan 2026
30. Mahdhi, M.: Blazor WebAssembly: full-stack web development. In: Mastering ASP.NET Core 10, pp. 581–611. Apress, Berkeley (2025). https://doi.org/10.1007/979-8-8688-1892-9_20

31. Voborník, P.: New modern application for gamified memorizing paired facts by more fun form of memory game with the possibility of own content development. In: The 13th international scientific conference on Distance Learning in Applied Informatics (DIVAI 2020), pp. 217–231. Wolters Kluwer, Štúrovo (2020)
32. Williams, L.A., Kudrjavets, G., Nagappan, N.: On the effectiveness of unit test automation at Microsoft. In: 2009 20th International Symposium on Software Reliability Engineering (ISSRE), pp. 81–89. IEEE (2009). https://doi.org/10.1109/ISSRE.2009.32
33. Coufal, P.: Project-based STEM learning using educational robotics as the development of student problem-solving competence. Mathematics. **10**(23), 4618 (2022). https://doi.org/10.3390/math10234618

Teaching Art History with the Help of a Three-Dimensional Virtual Model

Jan Vopršal(✉) and Marie Hubálovská

Faculty of Education, University of Hradec Kralove, Hradec Kralove, Czech Republic
{jan.voprsal,marie.hubalovska}@uhk.cz

Abstract. This article deals with the implementation of three-dimensional models in art education at art-oriented secondary schools. The theoretical part is based on an overview of the current state of virtual reality use in art. Based on a review of specialized research studies, it summarizes key trends in the use of virtual reality and its application in teaching, describes the current state of the teaching process with the support of a three-dimensional virtual model, its benefits, limitations, and the results of research conducted. The main part of the article is devoted to a pedagogical experiment involving two teaching groups in the subject of art history. An architectural monument from the Czech environment was selected for the experiment. Two groups of students were compared, the control group being taught using traditional methods with the projection of art examples. The experimental group had access to a three-dimensional virtual model on which individual architectural features were described. A test was used to compare the results of both groups, with the aim of verifying the impact of using a three-dimensional virtual model on understanding the subject matter and the level of knowledge acquired in the context of innovative approaches in contemporary secondary education.

Keywords: virtual reality · three-dimensional model · art history teaching · secondary school

1 Introduction

The teaching of fine arts in secondary schools aims to familiarize students with the historical context and successive stages of development in the fields of architecture, painting, sculpture, and applied arts. In art-focused secondary schools, art history is part of the curriculum. During their studies, students prepare for a final exam in this subject. In art-focused secondary schools without a final exam, this subject develops awareness of the development of fine arts.

A number of art-focused professional publications describe teaching in connection with virtual or augmented reality (Cabero-Almenara et al., 2022). Teaching with the support of 3D models is closely related to virtual three-dimensional models.

Modern technologies such as virtual reality or augmented reality allow for a better understanding of a given place. An example from Spain included a more detailed introduction to a building (a church). It was explored, and respondents positively evaluated the connection to art history (Cabero-Almenara et al., 2022).

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 420–431, 2027.
https://doi.org/10.1007/978-3-032-32104-6_36

The study by Hui et al. (2022) focuses on verifying the benefits of virtual reality in art education at the primary school level. The authors compared VR-supported teaching with the traditional approach and evaluated the impact on several areas – in particular, students' creativity, the degree of experiential "immersion" (flow), learning outcomes, and active participation in teaching. The results show statistically significant differences in favor of the VR variant: students achieved higher scores in creative outputs and showed a higher level of engagement and better learning performance. Qualitative feedback from teachers also showed that the immersive and vivid VR environment helps students better understand the processes and details of creation, as they can "experience" the subject matter and visualize complex phenomena directly in space. The authors conclude that meaningfully integrated VR can improve the quality and effectiveness of art education and enhance active learning (Hui et al., 2022).

An interesting topic in the field of fine arts is the digitization of sculptures. We can imagine repairing or restoring a sculpture. Today's technology allows us to create a three-dimensional scan. The missing parts of the sculpture are then added. An intermediate stage in the form of an existing model offered itself as a suitable source of data for further work with the virtual model (Hutson and Olsen, 2022).

Zhang (2024) presents an algorithm for creating a VR scene that combines visual-perceptual criteria with a method for optimizing a 3D network. He tests the approach on three works of art and shows improvements in alignment/scenic processing and evaluation of the resulting visual quality.

Mu et al. (2022) examine the attention and behavior of viewers when "encountering" a work of art in virtual reality—that is, how visitors move around the space, where they view the work from, and which elements they focus their gaze on. The authors created an abstract VR "painting" (Tilt Brush) and an experimental system in Unity that allows detailed recording of the viewer's interaction with the work (especially eye-tracking, head orientation, and movement in space) as well as voice comments from participants. In the study, 35 participants freely walked through and viewed the work in a defined physical space, while extensive data on their gaze and movement was collected.

Each work contains elements that attract attention. The results also include an analysis of which visual elements attract attention (e.g., colors) and how preferences may differ between groups—the study describes how many colors participants "encountered" during viewing and indicates differences in preferences by gender (Mu et al., 2022).

A very similar experiment, which is the subject of this article, was conducted in the field of technical imaging. Two groups of students were used for the research. The first group was taught in the traditional way, and the students modeled the output on paper. The second group had everything available in virtual form. The result of the research was that the group with virtual demonstrations performed better (Liao et al., 2025).

An experiment in the form of a quiz was conducted by Cecotti et al. (2024). It was based on a survey of the existence of "virtual galleries and museums." The number of movements around the space was used to collect data. The result is to set up the model so that the user "teleports" as little as possible and so that the controls are easy for all users.

Another article focuses on the use of virtual reality in teaching. Once again, the contribution is interesting. The part of the world where it was created—Indonesia—plays

a significant role. The author describes virtual reality as a technology that is currently hampered in its development in teaching by limited financial resources. On the contrary, in science-based subjects, it helps to improve understanding of the subject matter thanks to its vividness and degree of immersion (Dhanil and Mufit, 2024).

Cultural heritage is an important part of culture. From the perspective of material monuments, it helps to preserve historical periods of society's development. Through virtual, augmented, or mixed reality, we convey this culture and raise awareness in society about the existence of these monuments (Chatsiopoulou and Michailidis, 2025).

A museum space focused on mediating artistic artifacts has entered augmented reality. The main advantage was the ability to see the entire object from all sides. It was necessary to have a smartphone, tablet, or other device with you in the museum (Liu and Liu, 2025).

In terms of teaching, virtual reality helped improve educational outcomes by up to 20%. It helped in solving construction details or originality and correcting suitable colors for a specific user. On the other hand, from the teacher's point of view, it increases the time needed for preparation by up to a quarter (Mao and Ling, 2025).

An interesting factor in the use of virtual reality is the source device. The authors of the article confirmed the assumption that headsets are the most widely used. Computers rank second (Han et al., 2025).

Zekeik et al. (2025) describe virtual reality in their article as a constantly evolving technology. Over the past 10 years, they identify the release of the first experiments carried out 10 years ago, followed by a decline in price, as key milestones in its development. In 2018, immersive simulations came onto the market. The year 2020 saw the emergence of virtual classrooms in response to the coronavirus pandemic. The next step was hybrid teaching. Part of the teaching takes place in an immersive and interactive environment. The latest development came with the advent of generative artificial intelligence, which made it possible to create materials relatively quickly and efficiently.

Angel-Urdinola et al. (2021) mention virtual reality as a tool for preparing simulated situations in their research article. In their research experiment, they concluded that simulations can be repeated several times in a row. It is also possible to apply an individual pace. The result of the experiment was the participants' preparedness for a given situation. Compared to traditional exercises, participants prepared with the support of virtual reality reacted to the situation more quickly and managed to solve it in less time.

2 Research Part

The research part focuses on comparing two groups of students and their academic results in art history. I focus on teaching with the support of a three-dimensional virtual model of a specific building. The virtual model is intended for teaching new material. I am entering the research with the assumption that it will improve students' understanding and consolidation of new material. The teaching of new material in the control group was supported by images. An image was presented, on which all the characteristic features were described and shown. A virtual three-dimensional model will be projected for the experimental group. The description of the characteristic features will be supplemented by the projection of the model. A key change will be the ability to rotate the model and show the characteristic features of the selected structure from all sides.

Based on the results of the research, a project supporting virtual reality in teaching will be implemented in the future. In the project, I plan to change the presentation of the three-dimensional virtual model. The model will now be loaded into a headset (glasses). The projection of the model will work according to the student's movements.

The research aims to answer the research question:

"Will the experimental group of students with a demonstration of a virtual three-dimensional model achieve better results than the control group of students taught only with the support of visual material?"

2.1 Research Sample

A sample of students from the secondary school where I work was selected to carry out the research. These are students in graduating classes who have the subject of art history included in their study plan. Specifically, they are studying construction with a focus on building construction and architecture. The subject is taught in the third and fourth years, and its content is designed as long-term preparation for the oral graduation exam, with preparation taking place over a period of approximately 2 years.

Two classes participated in the experiment. This is a selective sample, where students choose their own focus. Their interest in the arts is their interest.

I chose the topic for comparison between the two groups with regard to the subject's thematic plan. The current topic was the early Middle Ages. Romanesque and Gothic art in Bohemia. An opportunity arose to find freely available three-dimensional virtual models needed for teaching. The first of these was the Rotunda of St. George on Mount Říp in Central Bohemia. It is a circular building situated in a prominent and symbolically significant location that has fundamental historical and cultural significance. The model is nicely integrated into the landscape. It offers a view of the interior space, allowing the entire building to be presented with all the features of the artistic style.

The experimental group consisted of 20 third-year students. These students were just getting acquainted with Romanesque and Gothic art. The lesson focused on teaching new material. The lesson included a demonstration of the features on a specific model and a demonstration of the location. Photographs were deliberately not shown, only a three-dimensional virtual model.

The control group consisted of 14 fourth-year students. These students had already covered both topics in the previous school year, so the lessons served as a review of the material in preparation for their final exams. The material was taught in a traditional way using visual aids, with the architectural features of individual buildings being explained and the students taking written notes.

2.2 Verification of the Results of the Experimental and Control Groups

A test was used to compare the two groups. The test was the same for both groups. It consisted of 20 questions. The questions focused on the rotunda and the characteristics of Romanesque art. I divided the questions into four parts. The first part consisted of questions with one correct answer out of three options marked with letters (a, b, c). The second part of the test consisted of five closed dichotomous questions with the options

true or false. The task was to choose whether the statement was true or false. The third part of the test contained questions that required a short answer. The answer contained a technical term. There were again five questions in this part. The fourth part focused on describing a picture. The schematic picture depicted a construction site. Five elements were marked in the picture. The names of the elements were listed below the picture. The elements were marked with letters. The task was to match the letters of the marked elements in the picture with the correct names.

Data Measurability
The entire test consists of four parts. The wording of the answers to the questions is unambiguous. A correct answer means one point is awarded. An incorrect answer means zero points. A maximum of 20 points can be obtained in the entire test. The overall accuracy of the test is evaluated according to the success rate in each part. For each part, the correct answers can be expressed as a percentage.

2.3 Research Survey Results

The first part consisted of five questions with a single possible answer. The first question concerned the floor plan of the rotunda. The choices were cross, oval or circle. The second question focused on the function of the rotunda. The choices were market, church and chapel or fortress. The third question concerned the shape of the rotunda's roof. The options were onion dome, dome or pyramid. In the fourth question, I asked about the part of the rotunda where the altar is located. The options were nave, apse or tower. The fifth question focused on the period of construction. The choices were 9th to 10th century, 10th to 12th century or 11th to 12th century.

The second part contained five questions. The task was to determine whether the statements were true or false. The first question was whether rotundas are among the oldest types of stone churches in our country. This statement is true. The second question asked whether rotundas could have more than one tower. This statement is false. Rotundas built in the Czech Republic have a maximum of one tower. The third question contained the statement that rotundas were built mainly for large numbers of people. Again, this statement is false. On the contrary, they were buildings for small numbers of people. The fourth question asked whether rotundas are often located in significant places, such as on hills. The statement is true. The last, fifth question focused on whether there are frescoes on the walls of the interior of rotundas. The statement is true.

The third part contained open-ended questions. The questions were asked with clear, short answers. The first question was what a rotunda is. The answer is a circular building or a church building or a predecessor of a church. The second question focused on the building material used for the construction of rotundas. Here, the answer is stone. The third question asked about the furnishings of a church located in the apse. The answer was an altar. The fourth question aimed to identify the material used for roofing. The answer here was wood or fired tiles. The last, fifth question focused on a simple element used in the construction of windows and vaults. The answer here was a semicircular arch.

The fourth part focused on describing a picture. It was a simple schematic picture of a rotunda. Five elements were marked in the picture. Five terms were listed below

the picture. Each term was marked with a letter (a, b, c, d, e). To complete the task, the selected letter had to be added to the reference line. The terms had to be assigned in the correct order. In this part, it was possible that the correct answers could be mixed up when guessing.

The image in Fig. 1 below shows the terms that both groups assigned. The terms are marked with letters in the following order:

a) main ship
b) double window
c) tympanum
d) cornice
e) apse

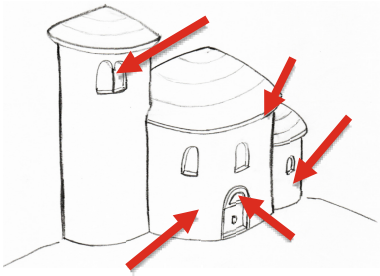


Fig. 1. Picture of rotunda to assign concepts.

2.4 Evaluation of the Data Obtained

The data for both groups was obtained from completed tests. These were printed tests that both groups completed. The completed answers were then analyzed and the data was obtained for final processing.

The tests were completed by 14 students from the control group and 20 students from the experimental group. Table 1 shows the number of correct and incorrect answers. In the first part, I base my analysis on the correct and incorrect answers.

Table 1. Number of answers.

	Control group		Experimental group	
Question	Correct	Wrong	Correct	Wrong
1	12	2	20	0
2	14	0	20	0
3	11	3	15	5
4	4	10	10	10
5	5	9	7	13
6	10	4	17	3
7	7	7	8	12
8	7	7	9	11
9	8	6	13	7
10	6	8	10	10
11	14	0	20	0
12	13	1	19	1
13	5	9	7	13
14	5	9	7	13
15	9	5	15	5
16	14	0	20	0
17	14	0	20	0
18	11	3	16	4
19	10	4	16	4
20	11	3	16	4

Part One

The questions for the first part are listed below.

1. What is the floor plan of a rotunda?
   a) cross-shaped; b) oval; c) circular
2. What were rotundas most commonly used for?
   a) marketplaces; b) churches or chapels; c) fortresses
3. What shape is the roof of a rotunda?
   a) onion-shaped; b) domed; c) pyramidal
4. Which part of the rotunda is intended for the altar?
   a) nave; b) apse; c) tower
5. From which period do most rotundas in Czechia originate?
   a) 9th–10th century; b) 10th–12th century; c) 11th–12th century

In the first part, students chose one of three options. According to the graph, the experimental group performed better on two questions. The result for the fifth question

was close. Both groups answered the second question correctly which can be interpreted as evidence that the subject matter was well mastered across all students, regardless of group placement.

The first part of the test allowed students to choose one of three possible answers. This allowed students to think about the possibilities in connection with the interpretation of the subject matter. Tasks designed in this way support not only memorization of facts, but also the ability to apply knowledge and critical thinking (Fig. 2).

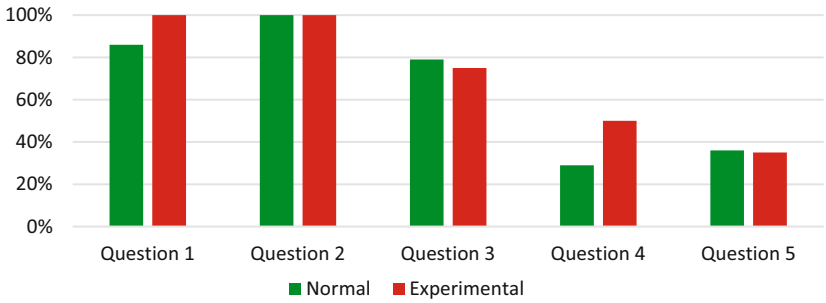


Fig. 2. Number of correct answers in the first part.

Second Part

In the second part, both groups chose between true and false.

6. Rotundas are among the oldest types of stone churches in our country. - true
7. Rotundas can have more than one tower. - false
8. Rotundas were built mainly for large numbers of people. - false
9. Rotundas are often located in prominent places, such as on hills. - true
10. Frescoes can be found on the walls of the interior of rotundas. - true

According to the correct answers (Fig. 3), the experimental group performed better on three questions. On question seven, the difference in correct answers between the groups was 10%. On question eight, the difference in correct answers was 5%. This result indicates a higher level of understanding and ability to apply knowledge among students in the experimental group compared to the control group.

Overall, the experimental group performed better in the second part, which supports the assumption that the applied teaching strategy was more effective than the traditional approach used in the control group.

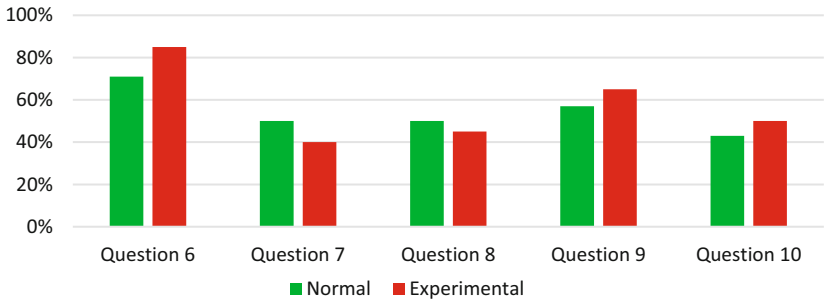


Fig. 3. Number of correct answers in the second part.

Third Part

In the third part, concepts were added, and the questions were designed as open-ended questions requiring a short and simple answer. These were the following questions:

11. What is a rotunda?
12. What building material was used to construct rotundas?
13. What church furnishings were located in the apse?
14. What material was most commonly used for roofing?
15. Which simple element was used in the construction of windows and vaults?

It should be noted that the answers received in this section were legible and identifiable which made their evaluation easier. The students tried to formulate their answers briefly and factually (Fig. 4). It should also be noted that the answers to questions 11 and 12 were evenly distributed. The normal group performed better in their answers on questions 13 and 14, demonstrating a higher level of knowledge in these specific areas. It should be noted here that, in terms of the curriculum covered, these students have a greater knowledge of the material covered. They have a broader and more comprehensive overview of the topic which was probably reflected in their better results on these questions.

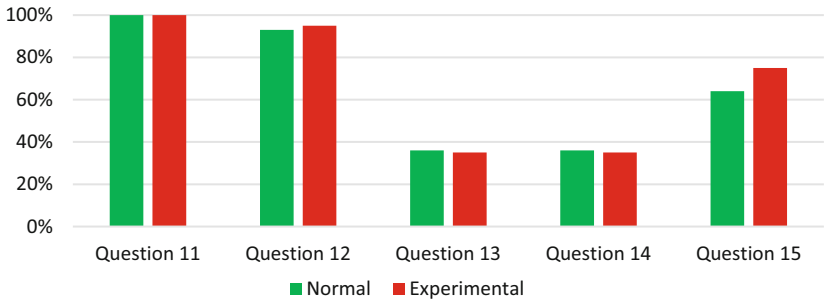


Fig. 4. Number of correct answers in the third part.

Part Four

I have marked the question numbers in the figure (Fig. 5). The correct solution is based on the following terms:

a) main nave - 16
b) double window - 17
c) tympanum - 18
d) cornice - 19
e) apse - 20

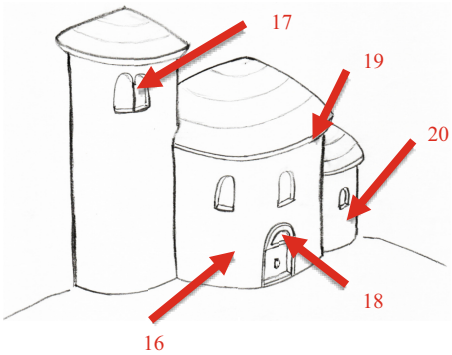


Fig. 5. Correct answers in the picture of rotunda

In the fourth part, the answers to questions 16 and 17 were identical. Both groups achieved 100% accuracy. It can be assumed that these were topics that were sufficiently practiced or clearly explained during the course of the lesson. The experimental group performed better on question 19. Questions 18 and 20 were again answered more successfully by the experimental group. The overall concept of the questions involved filling in the correct term from a list of options. If one of the terms was changed, the next answer to the question could not be correct. This format allows for a better understanding of the subject matter, as it requires systematic thinking and the correct connection of individual pieces of knowledge (Fig. 6).

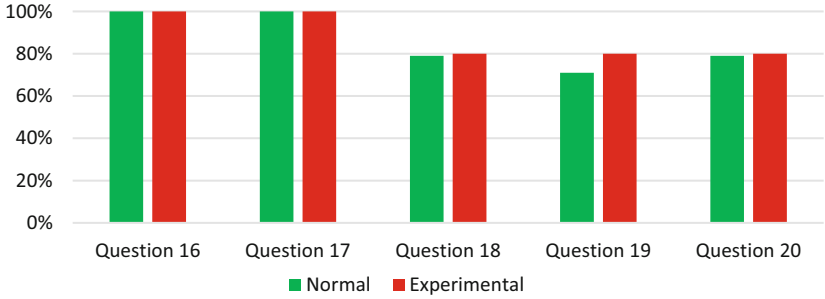


Fig. 6. Number of correct answers in the fourth part.

2.5 Overall Comparison

In the overall comparison, the experimental group performed better (Fig. 7). The result is quite close. The difference in overall success is 5%. The experimental group always performed better in each section. The experimental group achieved the best results in the fourth section. This fact indicates higher success across different types of tasks, whether they were multiple choice, open-ended questions or concept completion. It can therefore be concluded that students from the experimental group were able to work more effectively with different forms of assignments and better apply their knowledge in different contexts.

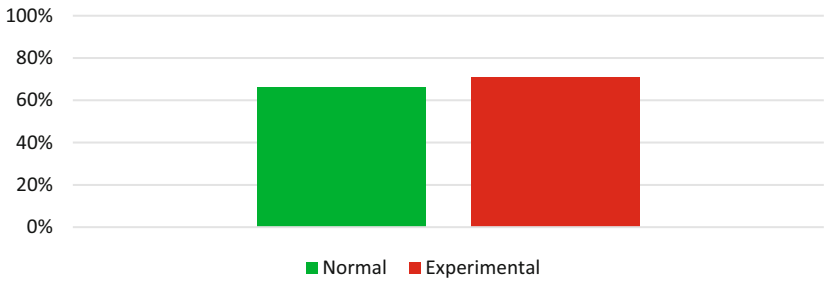


Fig. 7. Overall success rate.

3 Conclusion

The research explored the possibilities of using a virtual 3D model in art history education and showed that spatial visualization of architecture can support the understanding of concepts and the recognition of stylistic features. In the overall comparison, the experimental group achieved higher success in the test (a difference of 5%), most notably in tasks involving orientation in the building. Although the groups were not entirely equivalent, the ratio of group sizes corresponded approximately to a two-thirds to one-third ratio, which was also reflected in the interpretation of the results. Nevertheless, the final data show differences in the number of correct answers, which are evident in the graphical comparison between the groups.

From the perspective of the research question, it can therefore be stated that teaching supported by a virtual three-dimensional model led to better results in the given setting than teaching based solely on visual material. The benefit of the 3D model was probably most evident in the ability to rotate and "walk around" the structure, which made it easier for students to connect terminology with the specific location and function of elements in space.

3.1 Limitations of the Research

At the same time, the results must be interpreted with caution. The research sample was relatively small and the groups were not entirely equivalent (3rd grade – new material

vs. 4th grade – review). The results may have been influenced by differences in prior knowledge and motivation, as well as the novelty effect of the technology. In addition, the test related to a single monument and a single thematic area.

Further research should verify the findings on a larger sample and in a design that captures the initial level (pre-test–post-test) and retention over time.

Acknowledgment. The article was supported by Specific research project No 2115/2025, Faculty of Education, University of Hradec Kralove.

References

Cabero-Almenara, J., Llorente-Cejudo, C., Martinez-Roig, R.: The use of mixed, augmented and virtual reality in history of art teaching: a case study. Appl. Syst. Innov. **5**(3), 44 (2022). https://doi.org/10.3390/asi5030044

Cecotti, H., Huisinga, L., Peláez, L.G.: Fully immersive learning with virtual reality for assessing students in art history. Virtual Real. **28**, 33 (2024). https://doi.org/10.1007/s10055-023-00920-x

Dhanil, M., Mufit, F.: How virtual reality impacts science learning? A meta-analysis. Int. J. Interact. Mob. Technol. **18**(22) (2024). https://doi.org/10.3991/ijim.v18i22.49989

Han, X., Luo, H., Wang, Z., Zhang, D.: Using virtual reality for teacher education: a systematic review and meta-analysis of literature from 2014 to 2024. Front. Virtual Real. **6**, 1620905 (2025). https://doi.org/10.3389/frvir.2025.1620905

Hui, J., Zhou, Y., Oubibi, M., Di, W., Zhang, L., Zhang, S.: Research on art teaching practice supported by Virtual Reality (VR) technology in primary schools. Sustainability. **14**(3), 1246 (2022). https://doi.org/10.3390/su14031246

Hutson, J., Olsen, T.: Virtual reality and art history: a case study of digital humanities and immersive learning environments. J. High. Educ. Theory Pract. **22**(2), 50–65 (2022). https://doi.org/10.33423/jhetp.v22i2.5036

Chatsiopoulou, A., Michailidis, D.P.: Augmented reality in cultural heritage: a narrative review of design, development and evaluation approaches. Heritage. **8**(10), 421 (2025). https://doi.org/10.3390/heritage8100421

Liao, C., et al.: Integrating virtual reality into art education: enhancing public art and environmental literacy among technical high school students. Appl. Sci. **15**(6), 3094 (2025). https://doi.org/10.3390/app15063094

Liu, K., Liu, J.: A study of the educational applications of augmented reality in museums. In: ICBDIE '25: Proceedings of the 2025 International Conference on Big Data and Informatization Education, pp. 209–214. Association for Computing Machinery, New York (2025). https://doi.org/10.1145/3729605.3729642

Mao, Y., Ling, J.: Innovative applications of virtual reality in art and design education. J. Cases Inf. Technol. **27**(1), 19 (2025). https://doi.org/10.4018/JCIT.382567

Mu, M., Dohan, M., Goodyear, A., Hill, G., Johns, C., Mauthe, A.: User attention and behavior in virtual reality art encounter. Multimed. Tools Appl. **83**, 46595–46624 (2022). https://doi.org/10.1007/s11042-022-13365-2

Zekeik, H., Chahbi, M., Sefian, M.L., Bakkali, I.: Augmented reality and virtual reality in education: a systematic narrative review on benefits, challenges, and applications. EURASIA J. Math. Sci. Technol. Educ. **21**(9) (2025). https://doi.org/10.29333/ejmste/16830

Zhang, N.: Research on virtual reality art creation mode and visual perception under the integration of art and technology. Appl. Math. Nonlinear Sci. **9**(1) (2024). https://doi.org/10.2478/amns-2024-3342

Voice-Controlled Game Interaction

Patrik Voštinár(✉) and Radoslav Proft

Matej Bel University, Tajovského 40, 974 01 Banská Bystrica, Slovakia
patrik.vostinar@umb.sk

Abstract. This article investigates the use of automatic speech recognition (ASR) for voice-controlled interaction in computer games. It presents the design and implementation of two simple games (Tic-Tac-Toe and 8-Puzzle) developed in Unity and controlled using voice commands via the Whisper model. The study combines quantitative evaluation of recognition accuracy with qualitative analysis of user experience based on semi-structured interviews with 20 participants. The results show that modern ASR systems can achieve sufficient accuracy (approximately 84%) in controlled scenarios with a limited command vocabulary. At the same time, the findings highlight key usability challenges related to feedback, transparency, and user confidence. The study also confirms that voice interaction is most effective as a complementary modality rather than a full replacement for traditional input methods. The paper provides an empirical evaluation of voice-controlled interaction in simple gaming environments and identifies important design considerations for developing accessible and user-friendly voice interfaces.

Keywords: voice control · automatic speech recognition · games · Unity · Whisper · Hugging Face

1 Introduction

Video game control has evolved from simple joysticks and buttons to modern touchscreens, motion sensors, and specialized accessibility controllers (Brown et al., 2010; Strzałko, 2021). In recent years, voice control has emerged as an additional input modality, allowing players to control games through spoken commands. Voice control can provide a more enjoyable and intuitive interaction method, improve accessibility for players with movement limitations, and open new types of gaming experiences (Belgacem et al., 2011).

Despite advances in automatic speech recognition (ASR), voice control has not yet become mainstream in gaming. The primary obstacles are technical limitations (background noise, diverse accents, latency) and design challenges in creating appropriate voice interactions (Bohouta et al., 2017).

Recent research (2023–2024) further highlights the growing importance of Voice User Interfaces (VUI) within the broader field of Human–Computer Interaction (HCI), particularly in the context of multimodal and accessible interaction systems. Advances in automatic speech recognition, driven by large-scale models such as Whisper (Radford

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 432–444, 2027.
https://doi.org/10.1007/978-3-032-32104-6_37

et al., 2023), have significantly improved robustness to noise and speaker variability, enabling more reliable real-world deployment. At the same time, HCI research emphasizes that effective voice interaction design extends beyond recognition accuracy, focusing on usability, feedback mechanisms, and contextual awareness (Seaborn et al., 2024; Dubiel et al., 2024). Studies of voice assistants in everyday environments show that users often encounter challenges related to ambiguity, lack of transparency, and limited system feedback, which directly impact user experience (Porcheron et al., 2023). Moreover, VUI has been identified as a key enabler of accessibility, supporting users with motor impairments, older adults, and novice users by providing a natural and low-barrier interaction modality (Pradhan et al., 2023). In parallel, the increasing interest in multimodal interaction frameworks (Oviatt, 2023; Turk, 2023) suggests that voice is most effective when combined with traditional input methods rather than used in isolation. These findings provide important context for evaluating voice-controlled interaction in games and underline the need for careful interaction design alongside technical implementation.

The purpose of this article is to demonstrate how available ASR technologies can be utilized in developing simple voice-controlled games and to share practical user experiences with this type of control.

This study addresses the following research questions:

RQ1: How accurately can ASR systems recognize simple voice commands in controlled gaming scenarios?
RQ2: What usability challenges arise when using voice as the primary input modality in games?
RQ3: How do users perceive voice control compared to traditional interaction methods?

The main contribution of this paper is an exploratory evaluation of voice-controlled interaction using a constrained command vocabulary combined with user experience analysis.

2 Voice Control in Games

In the early era of video games, simple controllers dominated—directional joystick levers and a few buttons sufficed for arcade games of that time (Allison et al. 2017). As game complexity increased, keyboard and mouse input became prevalent, particularly in genres such as strategy games and first-person shooters, where high precision and reaction speed are essential. Game consoles established the gamepad standard with analog sticks, which represented a compromise between simplicity and functionality.

Touchscreens on smartphones and tablets introduced another significant shift, enabling direct interaction with game interfaces and simplifying input for casual players. Motion sensors (Nintendo Wii Remote, Xbox Kinect) connected physical player movement to on-screen gameplay. Concurrently, experimentation with voice control began, not to replace traditional inputs, but rather to expand the spectrum of how a game responds to the player (Kent, 2001).

Early attempts at voice control in games date back to the 1980s, but the technology was severely limited and largely unreliable. The game Seaman (1998) for the Sega Dreamcast, sold with a microphone, allowed interaction with a virtual “pet,” yet the

system frequently failed to understand even simple commands (Allison et al. 2017). Similarly, Tom Clancy's EndWar (2008) experimented with voice control of military units, but ASR systems at that time suffered from noise sensitivity, accent variation, and the requirement for strictly defined phrases (Zewoudie, 2020).

With the advent of machine learning and deep neural networks, speech recognition accuracy improved dramatically (Bohouta et al., 2017). Commercial systems such as Google Speech-to-Text, Microsoft Azure Speech, and OpenAI Whisper now support multiple languages, diverse accents, and noisy environments. This has enabled games where voice plays a more prominent role—for example, the VR title Starship Commander: Arcade, controlled by natural language in real time, or the tactical game There Came an Echo (Belgacem, 2011).

At the same time, the potential for accessibility and inclusion has become evident. Research emphasizes that voice control can make action and logic games accessible to players with limited mobility and provide an alternative to complex keyboard or controller schemes (Lekshmi et al., 2016).

3 Technological Foundations and Automatic Speech Recognition Tools

An automatic speech recognition system converts acoustic signals into text through several steps. Sound waves are captured by a microphone and analog-to-digital converter, after which the signal is segmented into short frames and acoustic features are extracted (frequency spectrum, energy, etc.) (Gladia, 2024). These features are then mapped to phonemes and words using statistical models (e.g., Hidden Markov Models—HMM) or, in modern systems, deep neural networks (RNN, CNN, and their combinations) (Radford et al., 2023).

Recognition accuracy is influenced by multiple factors: microphone quality, background noise, speaker accent and articulation, speech rate, and the size and character of the "vocabulary" of possible inputs (Bohouta et al., 2017). In gaming environments, latency is also critical—the delay between uttering a command and the game's response. Fast-paced action games require low latency, while turn-based or logic games can tolerate slightly higher delays (Radford et al., 2023).

This section provides only a brief overview of ASR technologies, as the main focus of this paper is on practical implementation and evaluation.

3.1 Cloud-Based ASR Services

Cloud-based ASR services are popular due to their high accuracy and relatively straightforward API integration. The most widely used include:

- **Google Speech to Text** – high accuracy, low latency, support for over 100 languages, features including timestamps, speaker identification, and automatic punctuation. Suitable for games requiring fast response in relatively quiet environments (Emms, 2023).

- **Microsoft Azure Speech** – good integration with Microsoft tools (Unity, Visual Studio), ability to train custom models with domain-specific vocabulary (character names, fictional terms), support for translation and speech synthesis. Useful for multilingual games and games with specialized lexicons (Ebert, 2023).
- **Amazon Transcribe** – optimized for noisy environments and scalability, supports streaming, custom vocabularies, and timestamps, though accuracy with less common accents tends to be lower than Google and Azure solutions (Lugar, 2016).
- **OpenAI Whisper in the cloud** – a model trained on approximately 680,000 h of multilingual data, highly robust against accents and noise. For gaming purposes, it is more suitable for slower-paced or prototype applications, as standard configurations are not optimized for ultra-low real-time latency.

Comparisons of these services in terms of accuracy, response time, customization options, and typical use cases are provided by resources such as (Bohouta et al., 2017).

3.2 Offline and Open-Source Solutions

For games where offline capability, privacy protection, or extremely low latency is critical, local ASR systems are viable options. A classic example is CMU Sphinx, which supports simple command vocabularies and runs directly on the device, but offers significantly lower accuracy compared to newer models (Welder, 2021). A modern open-source solution is Vosk, providing multilingual models capable of real-time recognition even on lower-powered devices, with multiplatform support (Python, Java, C#).

The Whisper model can also be run locally. However, this requires greater computational power and more implementation knowledge, though in return the developer gains excellent accuracy and control over data processing (Radford et al., 2023).

4 Design and Implementation of Voice-Controlled Games in Unity

This implementation serves as an experimental platform for evaluating the usability and performance of voice-controlled interaction in a controlled gaming context.

In selecting a game engine, Unity, Unreal Engine, and Godot were primarily compared. Unity offers:

- multiplatform deployment (more than 25 platforms),
- scripting in C#,
- built-in audio support (microphone, recording),
- classes for HTTP requests (UnityWebRequest) and JSON processing

These features significantly simplify integration with external ASR APIs (Hugging Face). Unreal Engine, while offering superior graphics performance, typically requires C++ work and/or third-party modules for microphone integration and HTTP communication, increasing complexity (Seaborn et al., 2024). Godot is attractive as an open-source 2D/3D engine (GDScript, C#, C++), but has weaker ecosystem support for ASR and fewer readily available tutorials.

For this project, Unity was selected, allowing focus on interaction design rather than low-level technical obstacles.

4.1 Game Design: Tic-Tac-Toe and 8-Puzzle

Both implemented games utilize a 3 × 3 grid and interact with the player exclusively through voice commands—numbers from 1 to 9 (in English). This design has several advantages:

- small and unambiguous command vocabulary,
- straightforward mapping of voice to grid positions,
- rapid testing of recognition accuracy.

Tic-Tac-Toe
The player controls the "X" symbol, and the computer controls "O". Each number 1–9 represents one cell in the grid. Upon recognizing a number, the system checks whether the cell is empty; if so, it places the player's symbol and then executes the AI move. Game logic includes checking for winning combinations and draws. Figure 1 shows a sample game.

Fig. 1. The game Tic-Tac-Toe.

8-Puzzle
A classic sliding puzzle with eight tile pieces and one empty space. The image is divided into 9 sections at game start, shuffled, and placed in the grid. The player speaks a tile number to move it—if the selected tile is adjacent to the empty space (horizontally or vertically), the tile moves. The game ends when all tiles match the original image.

In terms of implementation, 8-Puzzle is more demanding for animations and move validation, while Tic-Tac-Toe allows a simpler prototype for voice interaction.

Figure 2 shows the 8-Puzzle game.

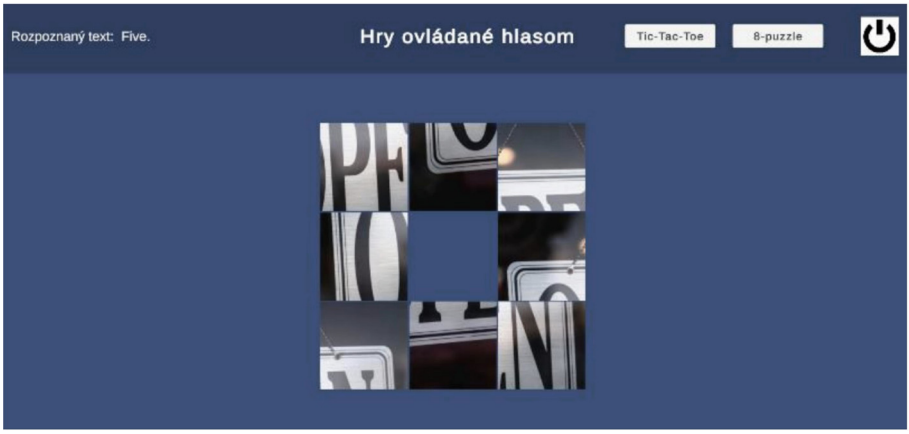


Fig. 2. The game 8-Puzzle.

4.2 ASR Integration via Hugging Face API (Whisper)

```
private void StartRecording()
    {
        if (recording) return;
        Debug.Log("StartRecording");
        clip = Microphone.Start(null, false, 10, 44100);
        recording = true;
    }
```

Fig. 3. Source code to start recording

```
private void StopRecording()
    {
        if (!recording) return;
        var position = Microphone.GetPosition(null);
        Microphone.End(null);
        var samples = new float[position * clip.channels];
        clip.GetData(samples, 0);
        bytes = EncodeAsWAV(samples, clip.frequency, clip.channels);
        recording = false;
        StartCoroutine(SendToWhisperAPI(bytes));
    }
```

Fig. 4. Source code for ending recording.

The core of the solution is the SpeechRecognition script in Unity, which:

```
private IEnumerator SendToWhisperAPI(byte[] wav)
    {
        using (UnityWebRequest request = new
UnityWebRequest("https://router.huggingface.co/hf-
inference/models/openai/whisper-large-v3-turbo", "POST"))
        {
            request.SetRequestHeader("Authorization", $"Bearer
hf_TOKEN");
            request.SetRequestHeader("Content-Type", "audio/wav");

            request.uploadHandler = new UploadHandlerRaw(wav);
            request.downloadHandler = new DownloadHandlerBuffer();

            yield return request.SendWebRequest();
            if (request.result == UnityWebRequest.Result.Success) {
                WhisperResponse response =
JsonUtility.FromJson<WhisperResponse>(request.downloadHandler.text);

                text.color = Color.white;
                text.text = "Rozpoznaný text: " + response.text;

                string trimmedResponse = response.text.ToLower().Trim();
                int number = 0;

                if (!int.TryParse(trimmedResponse, out number)) {
                    if (Regex.IsMatch(trimmedResponse, @"\bone\b"))
number = 1;
                    else if (Regex.IsMatch(trimmedResponse, @"\btwo\b"))
number = 2;
                    else if (Regex.IsMatch(trimmedResponse,
@"\bthree\b")) number = 3;
                    else if (Regex.IsMatch(trimmedResponse, @"\bfour\b"))
number = 4;
                    else if (Regex.IsMatch(trimmedResponse, @"\bfive\b"))
number = 5;
                    else if (Regex.IsMatch(trimmedResponse, @"\bsix\b"))
number = 6;
                    else if (Regex.IsMatch(trimmedResponse,
@"\bseven\b")) number = 7;
                    else if (Regex.IsMatch(trimmedResponse,
@"\beight\b")) number = 8;
                    else if (Regex.IsMatch(trimmedResponse, @"\bnine\b"))
number = 9;
                }
```

Fig. 5. Source code for sending data via the voice recognition API.

1. **Records audio** – when key “V” is pressed, Microphone.Start creates an AudioClip (max. 10 s, 44.1 kHz).

2. **Stops recording and prepares WAV** – upon key release, actual recorded samples are extracted from the clip and converted to WAV format (16-bit PCM) using a custom method (inspired by Hugging Face documentation) (AbuZeina and Elshafei, 2011)
3. **Sends data to API** – via UnityWebRequest, the WAV is sent to the ASR model endpoint openai/whisper-large-v3-turbo hosted on Hugging Face Inference API, with authentication via Bearer token (Emms, 2023).
4. **Processes text result** – the JSON response is deserialized, recognized text is displayed, and a number 1–9 is extracted from the string (as a digit or word form, e.g., "one", "two"…).
5. **Applies command in game** – depending on the active scene, it calls either TicTacToe.PlayerMove(number) or EightPuzzle.MoveSpecificTile(number - 1) and simultaneously tracks successful and failed recognitions for later analysis.

This architecture allows switching ASR models or providers without modifying game logic—only the API endpoint and request/response formatting need adjustment.

One of the most important parts of the whole game was to ensure that voice commands were recognized correctly. Without this, our games wouldn't work. For this reason, we decided to use an API provided by Hugging Face, which uses OpenAI's Whisper model. This service allowed us to convert spoken words to text without having to build our own recognition system from scratch. The basis of this part is a Unity C# script called "SpeechRecognition". Figure 3 shows the script as it starts recording audio.

Figure 4 shows the script when voice recognition is finished.

The "StopRecording" method sends us data (recordings) via the API, this functionality is implemented in the "SendToWhisperApi" function. The source code is shown in Fig. 5.

5 Methodology and Results

To assess the usability of voice control, qualitative evaluation through semi-structured interviews was employed, followed by quantitative analysis of recognition accuracy. This approach aligns with recommendations from game user research, where combining questionnaires, interviews, and observation reveals both subjective and objective dimensions of experience.

Twenty volunteers aged approximately 20–29 years, with varying levels of gaming and technology experience, participated—from regular gamers and IT students (e.g., Marek) to individuals who rarely play games (e.g., Jana, Silvia). Participants played games on their own devices in real-world conditions (various microphone types, varying background noise levels).

Participants were recruited from a convenience sample of university students and young adults. All participants performed the experiment under comparable conditions (indoor environment, standard microphone, moderate background noise). Each participant performed multiple command attempts during gameplay.

After gameplay, a semi-structured interview followed using a set of open-ended questions focused on:

- ease of understanding the controls,
- subjective perception of voice control (comfort, "awkwardness", naturalness),
- reactions to recognition errors,
- comparison of voice vs. traditional controls,
- suggestions for improvement and potential applications in other games.

Interview data were analyzed using a simple thematic coding approach. Responses were grouped into categories such as usability, recognition errors, and perceived advantages. The following themes emerged:

- **Easy control comprehension.** All participants quickly understood that pressing "V", speaking a number 1–9, and waiting for the game response was the interaction model. The small and clear command vocabulary significantly reduced cognitive load.
- **Initial hesitation, rapid adaptation.** Many participants, particularly those less "tech-savvy", initially felt awkward speaking commands to a computer; this sensation quickly faded and was replaced by curiosity and enjoyment.
- **Missing feedback.** A recurring comment was that players were uncertain whether the game was currently "listening" or had processed a command. Participants suggested adding visual or audio indicators of recording and processing status.
- **Game preferences**. Some participants preferred Tic-Tac-Toe for its speed and familiarity, others favored 8-Puzzle for its slower, more deliberate pace and visual interest. In both cases, voice control was perceived as an element that enhanced enjoyment of otherwise simple games.
- **Perceived broader applicability.** Several participants spontaneously noted the potential of voice control for children, elderly persons, or individuals with motor disabilities, and for educational and therapeutic applications, aligning with findings from other research on educational voice games.

Quantitatively, for each participant, recognition accuracy was recorded (correct vs. incorrect commands). Recognition accuracy ranged between 80–90% across participants. Table 1 summarizes detailed results including number of attempts, correct recognitions, and Word Error Rate (WER).

Word Error Rate (WER) was calculated as the ratio of incorrectly recognized commands to the total number of attempts. The average recognition accuracy across all participants was approximately 84.1%, with an average WER of 15.9%.

The response latency was typically around 1 s (approximately 900–1200 ms), depending on network conditions, as the Whisper model was accessed via a cloud API.

RQ1: How accurately can ASR systems recognize simple voice commands in controlled gaming scenarios?

The results show that ASR systems, specifically the Whisper model used in this study, achieve relatively high accuracy in recognizing simple voice commands. As presented in Table 1, recognition accuracy across participants ranged from approximately 80% to 90%, with an average accuracy of 84.1% and an average Word Error Rate (WER)

Table 1. Recognition results per participant.

Participant	Attempts	Correct	Errors	Accuracy (%)	WER (%)
Marek	42	38	4	90.48	9.52
Laura	33	27	6	81.82	18.18
Tomáš	41	33	8	80.49	19.51
Jana	36	29	7	80.56	19.44
Róbert	50	41	9	82.00	18.00
Silvia	35	30	5	85.71	14.29
Michal	40	34	6	85.00	15.00
Peter	38	32	6	84.21	15.79
Lucia	34	29	5	85.29	14.71
Martin	45	38	7	84.44	15.56
Zuzana	37	31	6	83.78	16.22
Jozef	39	33	6	84.62	15.38
Anna	31	26	5	83.87	16.13
Daniel	44	37	7	84.09	15.91
Veronika	36	30	6	83.33	16.67
Lukáš	47	40	7	85.11	14.89
Katarína	35	29	6	82.86	17.14
Andrej	43	36	7	83.72	16.28
Monika	32	27	5	84.38	15.62
Filip	46	39	7	84.78	15.22

of 15.9%. These findings indicate that for constrained vocabularies (e.g., numeric commands 1–9), modern ASR systems provide sufficient reliability for practical use in simple gaming scenarios.

RQ2: What usability challenges arise when using voice as the primary input modality in games?

The analysis indicates several recurring usability challenges. Participants frequently reported uncertainty about whether the system was actively listening or had successfully processed a command, highlighting the importance of clear feedback mechanisms. Recognition errors, although relatively infrequent, occasionally disrupted gameplay and caused frustration. Additionally, some users initially experienced discomfort when interacting via voice, particularly those less familiar with voice-based technologies. These findings suggest that usability issues are not primarily caused by recognition accuracy alone, but also by interaction design factors such as feedback, transparency, and user confidence.

RQ3: How do users perceive voice control compared to traditional interaction methods?

The results indicate that users generally perceive voice control as an interesting and engaging alternative to traditional input methods, particularly in simple or slower-paced games. Participants appreciated the novelty and accessibility of voice interaction, especially in contexts where traditional controls may be less convenient. However, voice control was not seen as a full replacement for conventional input devices, particularly in fast-paced or precision-demanding scenarios. Instead, users perceived voice interaction as a complementary modality that can enhance gameplay experience when appropriately integrated.

6 Discussion

The results of this study can be interpreted in the context of recent research on Voice User Interfaces (VUI) and Human–Computer Interaction (HCI). The observed recognition accuracy (approximately 84%) is consistent with findings from prior work, which suggest that modern ASR systems achieve reliable performance when operating with constrained vocabularies and controlled interaction scenarios (Radford et al., 2023; Dubiel et al., 2024). This confirms that limiting the command space to simple inputs, such as numbers, significantly improves recognition performance.

At the same time, the qualitative findings highlight that usability challenges are not solely determined by recognition accuracy. Consistent with previous studies (Seaborn et al., 2024; Porcheron et al., 2023), participants reported uncertainty related to system feedback and interaction transparency. This suggests that effective voice interaction design requires not only robust ASR models, but also well-designed feedback mechanisms that inform users about system state and processing.

Furthermore, the results support the view that voice interaction is best suited as a complementary modality rather than a full replacement for traditional input methods (Oviatt, 2023; Turk, 2023). While participants appreciated the accessibility and novelty of voice control, they also indicated limitations in scenarios requiring speed and precision. This aligns with the broader HCI perspective that multimodal interaction—combining voice with traditional controls—provides the most flexible and user-friendly approach.

Overall, the findings reinforce the importance of combining technical performance evaluation with user-centered design considerations when developing voice-controlled interactive systems.

7 Conclusions

This article presented the design and implementation of two simple voice-controlled games using Unity and the Whisper ASR model via the Hugging Face API. The Tic-Tac-Toe and 8-Puzzle games served as a testing environment to evaluate the feasibility and usability of voice interaction in controlled gaming scenarios.

The results demonstrate that voice control can provide an intuitive and engaging interaction method, particularly when using a constrained command vocabulary. Users adapted quickly to this interaction style, and modern ASR systems achieved sufficient accuracy for simple game mechanics. At the same time, the findings highlight

the importance of feedback mechanisms and interaction design in ensuring a positive user experience.

Although voice control is unlikely to replace traditional input devices in all contexts, it represents a valuable complementary modality that can enhance accessibility and expand interaction possibilities in games.

Future work should focus on larger participant samples, comparison of multiple ASR systems, and evaluation of performance differences between local and cloud-based solutions. More complex interaction scenarios, including natural language input and multimodal interaction, also represent promising directions for further research.

Despite its limitations, this study provides an initial empirical insight into the use of voice-controlled interaction in simple gaming environments and highlights key factors influencing its effectiveness.

Acknowledgement. This contribution has been processed as part of grant project no. 010TTU-4/2025 Development of pupils' computational and algorithmic thinking through the effective integration of modern technologies into the teaching of computer science and mathematics.

References

AbuZeina, D., Elshafei, M.: Automatic speech recognition: a survey. In: Cross-Language Speech Processing, pp. 1–14. Springer, Heidelberg (2011)

Allison, M.C., et al.: A history of voice interaction in digital games. In: Proceedings of the Audio Mostly Conference (AM 2017), pp. 1–8. ACM, London (2017)

Bohouta, G., et al.: Comparing speech recognition systems: Microsoft API, Google API, and CMU Sphinx. Int. J. Eng. Res. Appl. **7**(3), 20–24 (2017)

Brown, A., et al.: Beyond the gamepad: HCI and game controller design and evaluation. In: CHI Workshop on Game Input Devices, pp. 1–6. ACM, Atlanta (2010)

Dubiel, M., Halvey, M., Azzopardi, L.: Voice user interfaces: usability and user experience challenges. Foundations and Trends in Human–Computer Interaction (2024)

Ebert, D.: Speech recognition in unity using hugging face API—Tutorial (2023). https://dev.to/huggingface/speech-recognition-in-unity . Last accessed 10 Sep 2023

Emms, S.: Godot—2D/3D cross-platform compatible game engine review. Linux Magazine. **274**, 42–47 (2023)

Følstad, A., Brandtzæg, P.B.: Chatbots and the new world of HCI. Interactions Magazine. **24**, 38–42 (2023)

Gladia: OpenAI Whisper vs Google Speech-to-Text vs Amazon Transcribe: Detailed Comparison (2024). https://www.gladia.io/whisper-vs-google-vs-amazon . Last accessed 18 Jan 2025

Kent, S.L.: The Ultimate History of Video Games, Vol. 1, 1st edn. Crown Trade Paperback, New York (2001)

Lekshmi, K.R., et al.: Automatic speech recognition using different neural network architectures—a survey. Int. J. Comput. Appl. **155**(8), 21–28 (2016)

Lopatovska, I., Williams, H.: User perceptions of voice assistants: a systematic review. Int. J. Hum.-Comput. Stud. (2023)

McTear, M.: Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots. Springer (2024)

Oviatt, S.: Multimodal interfaces. In: The Handbook of Human-Computer Interaction (2023)

Porcheron, M., Fischer, J.E., Sharples, S.: Voice interfaces in the wild: understanding real-world use of voice assistants. ACM Trans. Comput.-Hum. Interact. (2023)

Pradhan, A., Mehta, K., Findlater, L.: Accessibility came by accident: use of voice-controlled intelligent personal assistants by people with disabilities. In: Proceedings of the CHI Conference on Human Factors in Computing Systems (2023)

Radford, A., et al.: Robust speech recognition via large-scale weak supervision. In: Proceedings of the 40th International Conference on Machine Learning (ICML 2023), vol. 202, pp. 28492–28518. PMLR, Honolulu (2023)

Reeves, S., Goulden, M., Dingwall, R.: Human–Computer Interaction in the age of AI: new directions. Interact. Comput. (2023)

Seaborn, J., et al.: Understanding and mitigating challenges of voice interfaces in everyday environments. arXiv preprint https://arxiv.org/abs/2401.12345. (2024)

Strzałko, D.: Voice controlled games—the approach and challenges of implementing speech recognition and voice control in games. Bachelor's thesis, Blekinge Institute of Technology (2021)

Feedback and Assessment as a Comprehensive Learning Support System in Online Teacher Education Courses

Jana Vejvodová and Lucie Rohlíková(✉)

Faculty of Education, University of West Bohemia, Plzeň, Czech Republic
vejvod@kcj.zcu.cz, lrohlik@kvd.zcu.cz

Abstract. The case study examines feedback and assessment as a comprehensive learning support system in online teacher education courses. It is based primarily on experience from two online courses implemented in higher education programmes for teacher education students: Didactics of the Czech Language and Information and Communication Technologies in Czech Language Education. The courses are attended by both full-time and part-time students, including those with prior or ongoing teaching experience.

In these courses, feedback is designed as a systematic, multi-component and long-term element of instruction rather than a mere assessment tool. The paper analyses the integration of automated feedback provided through self-tests, tests and interactive exercises with teacher-provided feedback and assessment on submitted assignments and discussion forum contributions. Assessment is embedded within the feedback system and is used primarily for formative purposes, supporting students' understanding, professional thinking and self-reflection.

Discussion forums are analysed as spaces for professional dialogue and peer learning in an online environment. The paper shows that a well-structured feedback and assessment system in e-learning enhances learner autonomy, continuity of learning and active student engagement, while also modelling effective feedback practices that students can later apply in their own teaching.

Keywords: Feedback · Assessment · Online teacher education · E-learning

1 Introduction

Managing learning processes without feedback is ineffective, particularly in online environments where direct, immediate interaction between teachers and students is limited. In e-learning, the quality of instruction depends to a large extent on the feedback students receive from tutors, as well as on the feedback tutors obtain about students' progress towards intended learning outcomes. In digitally mediated learning, feedback functions as a key mechanism for regulating learning processes, compensating for the absence of continuous face-to-face contact [2, 8]. The necessity of refocusing on teacher presence and meaningful learning activities has become even more critical in the post-pandemic era, as online instruction has moved from an emergency solution to a standard pedagogical model [6].

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 445–456, 2027.
https://doi.org/10.1007/978-3-032-32104-6_38

Students in online courses need information that helps them understand what knowledge they have already acquired, which skills they have developed, and how their prior understanding or attitudes have changed throughout the learning process. Feedback has therefore become one of the central concepts in research on e-learning and online instructional design. In this study, attention is paid specifically to feedback in tutor-supported online courses, where feedback is provided both through digital systems and through direct interaction with the tutor.

In educational research, feedback is commonly defined as information provided to learners about the correctness of outcomes, learning processes or strategies, rather than as a judgement of learners' personal qualities [3, 7]. In e-learning environments, meaningful feedback needs to be embedded not only in instructional content but also in the design of the learning management system, enabling students to use feedback for further learning and self-regulation [5].

Assessment can be understood as a systematic process of collecting, interpreting and evaluating evidence about learners' knowledge, skills and attitudes in relation to predefined learning objectives [1]. In e-learning, assessment typically combines summative assessment, focused on learning outcomes, with formative assessment, which supports the learning process through continuous guidance and adaptation [8]. Rather than serving solely a classificatory function, assessment in online courses increasingly operates as a tool for managing learning and supporting learner autonomy.

The conceptualisation of feedback and assessment in e-learning shows that these processes are not separate but closely interconnected components of online instruction. Their deliberate integration into online course design represents one of the key prerequisites for high-quality e-learning and for effective learning support in digitally mediated teacher education.

This study draws on experience from two tutor-supported online courses implemented in higher education teacher education programmes: Didactics of the Czech Language and Information and Communication Technologies in Czech Language Education. Both courses are designed for full-time and part-time students and provide a suitable context for examining feedback and assessment as a comprehensive learning support system in e-learning.

2 Methodology

This study employs a qualitative research design based on a thematic analysis of instructional practices and feedback mechanisms. The primary objective is to analyze how different forms of feedback function as a comprehensive learning support system in online pre-service teacher education.

The research context consists of two tutor-supported online courses implemented over the duration of one semester each. *The course Information and Communication Technologies in Czech Language Education* involved 43 participants and included 5 individual partial assignments and 1 group task, assigned on a weekly basis. The course *Didactics of the Czech Language* involved 38 participants. The learning activities in this course comprised 7 automated self-tests, 5 discussion tasks (two of which focused on analyzing video recordings of teaching), 1 partial task focused on textbook error

analysis, and a final seminar assignment consisting of either a lesson plan or a video recording of the student's own teaching. The data set includes teacher-provided feedback on these continuous partial tasks, feedback on final seminar assignments, interactions within discussion forums, and feedback the teacher receives regarding students' success rates in automated self-tests.

This study is designed as a descriptive qualitative case study based on the reflection of pedagogical practice. Rather than employing a strict quantitative methodology, the research is grounded in the qualitative review of course artifacts and the authors' direct experiences as instructors. The analysis consisted of a retrospective review of the collected data (teacher feedback, forum posts, automated test logs) to inductively identify repeating patterns and key themes regarding how feedback mechanisms influence students' professional development and engagement. Reliability was ensured through prolonged engagement with the students throughout the semester and the cross-referencing of different types of feedback (automated, peer, and tutor).

3 Results: Analysis of Feedback Mechanisms in Online Courses

3.1 Automated and Peer Feedback

The analysis of the selected courses confirms that feedback plays a crucial role, often functioning as a primary bridge for interaction between students and teachers in the absence of face-to-face contact. It acts as a "driving force" that keeps the student engaged and progressing. For students, feedback represents the primary source of information about whether they have correctly understood the subject matter, whether they are progressing in the intended direction, and how they can further improve their performance.

In this context, automated feedback may take the form of

- self-tests,
- tests,
- interactive exercises.

Our findings indicate that the didactic effectiveness of automated feedback in the analyzed tests was significantly higher when students were provided with step-by-step explanations (as shown in Fig. 1), rather than simple 'correct/incorrect' notifications. It becomes more didactically effective when the feedback provided during the review phase includes an explanation of the correct answer or guidance on how to arrive at the correct solution. Examples from the online course *Didactics of the Czech Language* show Figs. 1 and 2.

Teacher-provided feedback most often takes the form of a written comment on

- a submitted assignment,
- a submitted essay,
- a project,
- a discussion forum contribution,
- a portfolio.

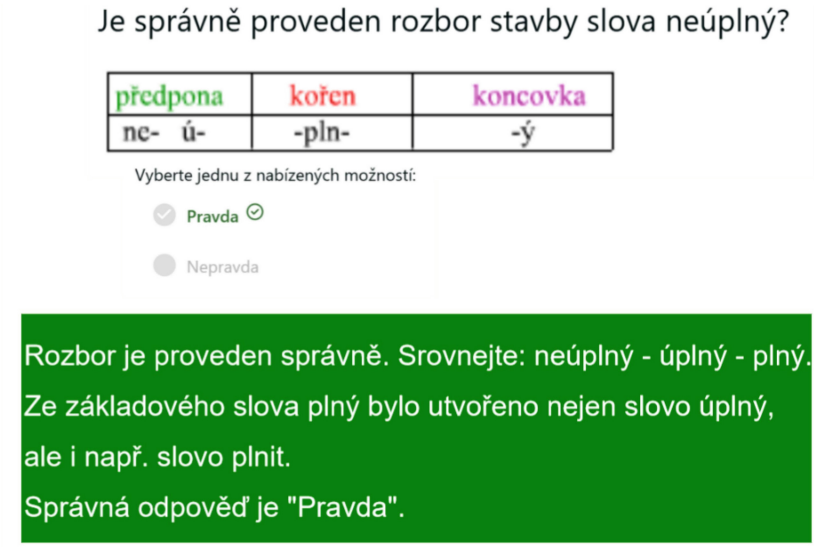


Fig. 1. Test item: *Is the morphemic analysis of the word correct?* The correct answer (*True*) is displayed below the question, accompanied by a green highlighted box explaining the step-by-step word formation process that supports the correct response.

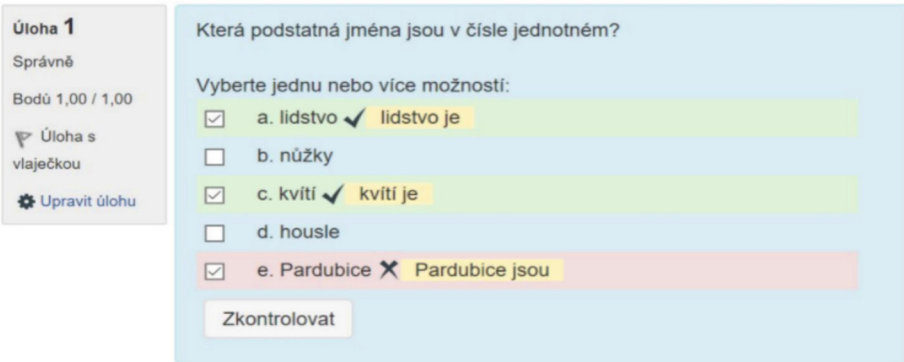


Fig. 2. Test item: *Which nouns are in the singular?* The justification of the correct selection of two nouns and the incorrect selection of one noun is based on subject–verb agreement with the verb *být*. Nouns in the singular combine with the form **"je"**, whereas nouns in the plural combine with the form jsou".

Teacher-provided feedback may be delivered in written form or as an audio or video comment. Its aim is not only to evaluate the degree to which the assignment requirements have been fulfilled, but above all to guide the student towards further improvement, point out errors, highlight strengths, and support understanding of the subject matter.

In self-tests and tests, students receive feedback immediately. When partial assignments are included in an online course, it is beneficial for students to receive feedback

from the tutor as early as possible. Feedback does not evaluate the student's personality but describes performance (e.g. "*At this step you chose the correct procedure, but the calculation is incorrect.*"). Feedback is not only a monologue by the tutor; it also provides information for the teacher about the extent to which the instructional materials in the course are comprehensible. Feedback indicates ways to improve and correct errors. In tutor-supported courses, dialogue or polylogue takes place in discussion forums and represents a collective opportunity for learning.

Automated feedback becomes insufficient particularly in situations where the aim is to develop complex skills, practise argumentation, or foster creativity. In such cases, teacher-provided feedback is indispensable, as it enables individualised and context-sensitive evaluation. The tutor can respond directly to specific student procedures, describe the strengths of the student's work, and offer targeted corrective recommendations. At the same time, space is created for dialogue, which supports deeper learning and strengthens study motivation. In these situations, feedback becomes not only part of the assessment process but also a key means of pedagogical guidance in an online course.

Example of feedback provided to a student on a quiz created in the online course Information and Communication Technologies in Czech Language Education:

> *"One of the advantages of this type of electronic test is that it enables students to receive quick automated feedback. When using quizzes with open-ended questions, however, it is necessary to provide individual feedback to each student, which can be time-consuming. Questions that ask students to list a specific number of characteristic features of the experimental novel or to briefly describe the features of the "Lost Generation" may therefore be more suitable for oral assessment, where immediate and individual feedback can be offered.*
>
> *In an online quiz, it may be beneficial to consider complementing open-ended questions with closed or semi-closed items that allow for automated feedback. This combination helps ensure that students receive timely feedback while still supporting deeper understanding of the subject matter. Even if open-ended questions are reviewed with the whole class after the quiz, it is difficult to address all incorrect or partially appropriate answers individually. Based on the suggestions above, I would kindly ask you to revise the quiz accordingly. Revising the task will allow you to reflect on the feedback provided and further improve the didactic quality of the quiz."*

In tutor-supported courses with a larger number of participants, providing detailed teacher-provided feedback on every individual discussion post may become time-consuming and unsustainable for the tutor. One of the tools of formative assessment in such contexts is structured peer feedback, in which students provide feedback on their peers' contributions based on predefined criteria. In this study, peer feedback is understood as structured, criteria-based feedback provided by students to their peers without assigning grades, and is conceptualised as a tool of formative assessment.

Peer feedback in e-learning is effective when it is not perceived as "substituting the teacher's work" but as an autonomous learning activity in its own right and as a tool of formative assessment. Based on the evaluation of student interactions in the

Moodle Workshop module, three key elements were identified as essential for successful peer feedback: structured scaffolding, automated anonymity, and the teacher's role as a meta-evaluator.

Structure Instead of Openness (Scaffolding)
The tutor does not simply instruct students to "assess a peer's work". Instead, a fixed structure of guiding questions is embedded into the peer assessment module (e.g. the *Workshop* activity in Moodle), which reviewers are required to answer.

Examples of criteria:

- *"Identify one aspect of your peer's work that you find original."*
- *"Point out one element that, in your opinion, does not correspond to task requirement no. 2."*
- *"Suggest one specific resource (e.g. a link) that could help your peer further develop their work."*

Automated Anonymity and Rotation
Digital tools allow the assessment process to be configured so that peer reviews are anonymous and the system automatically assigns, for example, two peers to each student for review. Anonymity in the online environment often leads to greater honesty and reduces social barriers (e.g. *"I do not want to criticise them because I know them"*).

The Tutor's Role as a Meta-Evaluator
A significant shift in the instructional model was observed when the tutor acted as a meta-evaluator, shifting the focus from the final product to the quality of the formative dialogue itself. The tutor awards points for how constructively and thoughtfully a student has advised their peer. In doing so, the tutor sends a clear signal that the process of formative assessment is as important as the final outcome itself.

3.2 Integration of Formative and Summative Assessment

Assessment in e-learning fulfils several interrelated functions, and its form depends on educational objectives, the nature of the course, and the level of student engagement. The two basic types of assessment encountered in e-learning environments are formative assessment and summative assessment. Their meaningful differentiation and deliberate integration represent an important prerequisite for high-quality online instruction.

Formative assessment focuses on the learning process and aims to support students throughout their studies. In e-learning, it is implemented primarily through

- ongoing assignments,
- short tests,
- reflective activities,
- self-assessment tasks,
- discussion contributions.

The essence of formative assessment is not grading but the provision of information that enables students to understand their own progress, identify weaknesses, and adjust their learning strategies. Feedback plays a crucial role in this type of assessment and should be specific, timely, and oriented towards further development. While feedback is information, formative assessment is a process.

In tutor-supported e-learning courses, formative assessment also has a motivational function. Regular contact with students through feedback strengthens their sense of support and reduces the risk of academic failure or premature course withdrawal. Moreover, digital environments enable the use of automated tools that provide immediate responses and contribute to the development of self-regulated learning.

Summative assessment focuses on learning outcomes and serves to provide an overall evaluation of the level of knowledge and skills achieved by students. In e-learning, it is most commonly implemented through

- final tests,
- final assignments,
- project-based work, portfolios,
- a combination of multiple assessment tools.

Its primary purpose is to verify the attainment of predefined learning objectives and it is often linked to grading or certification.

A specific feature of summative assessment in e-learning is the increased demand for transparency, validity, and fairness of assessment criteria. Students must be clearly informed about the rules according to which their performance will be evaluated and about the weight of individual assessment components. An important role is also played by the prevention of unethical behaviour and the selection of assessment formats that support authentic learning and the application of knowledge.

Formative and summative assessment in e-learning should not function as separate processes but as complementary elements of instructional design. Formative assessment prepares students for successful summative assessment and simultaneously provides teachers with valuable information for managing instruction. Their balanced integration contributes to higher learning quality, greater student engagement, and more effective achievement of educational objectives in online environments.

3.3 Case Study: Qualitative Analysis of Teacher-Provided Feedback

In this context, teacher-provided feedback and assessment are understood as complementary components of a unified learning support system rather than as separate instructional processes.

Teacher-provided feedback represents an indispensable component of e-learning, particularly in situations where the aim is to develop higher-order cognitive skills, critical thinking, argumentation, or the ability to reflect on one's own learning. It enables teachers to respond to individual learning processes and to provide targeted, descriptive and corrective feedback. In tutor-supported online courses, teacher-provided feedback and assessment are implemented through a variety of digital tools and instructional formats.

In the Moodle learning management system, the Assignment activity serves as

- a tool for submitting learning tasks,
- a means of teacher-provided assessment,
- a space for providing individual feedback (text-based, audio or video),
- an archive of students' learning outputs.

From a didactic perspective, it represents a comprehensive assessment tool that integrates task assignment, submission, assessment and feedback within a single environment.

In the online course *Didactics of the Czech Language* within the primary teacher education programme, students enrolled in the part-time mode of study were able to choose one of two formats for the final course assignment.

- **Prepare a lesson plan for the language component of the subject taught in Grades 2–5 of primary school.**
  The topic may be chosen freely. Please upload the lesson plan to the assignment repository in the section *Lesson Preparation.*
- **Create a video recording of your own teaching of the language component of the Czech language subject in Grades 2–5.**
  Please submit the video recording via the website **www.uschovna.cz**

Example of feedback on this assignment:

"Dear colleague,

Your pupils are very fortunate to have a teacher like you. The children clearly adore you and would go the extra mile for you. This is due to your positive communication with them, your consistent use of positive feedback, and above all your charisma. The classroom atmosphere is highly positive, friendly, and focused on learning. The children were not afraid to make mistakes, as you consistently used errors as opportunities for further explanation. The pupils remained constantly motivated, as if they were in an amusement park, while at the same time learning.

The entire lesson was dynamic and interactive, built on a varied sequence of methods—from traditional work with the textbook, through writing the practised words, pantomime, and work with cards, to the final creative activity. The introductory motivation through a song was very effective. The pantomime increased pupils' engagement by connecting movement with situations related to the meanings of the target words. You very appropriately reinforced the rules (no speaking during pantomime) and supported pupils who were unsure how to express the meaning of a word through movement.

During the card-based activity on the carpet, you proceeded suitably from simpler to more complex tasks and explained the differences between the paronyms plivat and plýtvat very clearly. You gave all pupils the opportunity to participate and encouraged them to provide logical reasoning rather than mere guessing.

At this point, I have two minor critical remarks. It was evident that the pupils know that i/í is written after neutral consonants outside word endings when the word is neither a listed word nor a related word. However, this logical explanation was several times simplified in a way that made it inaccurate, for example: lískové oříšky, koblížek, blízko – í because the word is not on the list of listed words. (The

correct explanation would be: …because it is neither a listed word nor a related word.)

In the previous seminar, I showed a video in which we demonstrated how an incorrect substitution of the terms "form of a listed word" and "related word" occurred in a lesson. Do you remember? In your lesson, you accepted the explanation that vzlykala is a related word derived from the listed word vzlykat. However, this is not correct—it is a form of the listed word vzlykat.

You manage classroom organisation perfectly. Overall, I consider your lesson to be very successful and I applaud you wholeheartedly."

The qualitative analysis of student responses suggests a strong formative effect of teacher-provided feedback; for instance, the student's reflection demonstrates not only a correction of pedagogical errors but also an increase in study motivation and professional identity:

"I greatly appreciate your feedback. You can probably imagine how much joy it brought me. I would not hesitate to say that you motivated me to continue my studies and remain in the field of education. The shortcomings you mentioned are accurate, even though they make me feel a little disappointed. I remember this topic very well from the lecture and try to avoid such mistakes. I will pay closer attention to this in the future; we will be working with listed words for a long time yet.

Thank you very much for your time and for your evaluation. I look forward to our next meeting on any occasion."

An important role is also played by discussion forums, which serve not only as a means of communication but also as a space for assessing learning activity. Teacher-provided assessment of discussion contributions focuses on factual accuracy, the level of argumentation, the ability to respond to others, and engagement in dialogue. Feedback may take the form of an individual comment or a public response, which supports collective learning and the sharing of different perspectives.

Examples of Feedback on Discussion Forum Contributions

- *"Your contribution presents the teacher's communication very accurately."*
- *"The use of a word in an incorrect meaning does not diminish the excellent quality of your discussion post in any way."*
- *"Your perspective on communication in the classroom is highly professional."*
- *"Your evaluative comments are very stimulating for our discussion."*
- *"Thank you for having the courage to contribute to the discussion as the first participant. Personally, I consider the most important parts of your longer contribution to be those passages in which you draw on your own experience, express your personal opinion, and offer your own evaluation."*

A specific form of teacher-provided assessment is represented by electronic portfolios, in which students collect the results of their work over a longer period of time.

Portfolio assessment is based on predefined criteria and supports reflection on learning, self-assessment and awareness of one's own progress. In this context, the tutor assumes the role of a guide who helps students interpret their achievements and plan further development.

The collected data demonstrate that when teacher-provided feedback is systematic and didactically grounded, it functions as a natural part of the pedagogical dialogue, supporting the development of self-regulated learning. When teacher-provided feedback and assessment are systematic and didactically grounded, they significantly contribute to the quality of online instruction and to the development of students' learning autonomy.

In order for students to engage actively with feedback and not perceive it merely as a "mandatory comment", feedback needs to be linked to subsequent learning activities and embedded meaningfully in the learning process. For example, teachers may require a short feedback reflection in which students briefly describe what they have understood, identify the most important recommendations, and explain how they intend to use them in their further work. In this way, feedback becomes a stimulus for learning rather than a one-off piece of information.

Another effective approach is the sequencing of tasks. When students explicitly apply feedback from a previous assignment in a subsequent task, they learn to perceive assessment as support for improving their performance. At the same time, teachers strengthen the formative function of assessment.

The clarity and specificity of feedback are also essential. Clearly formulated recommendations, possibly supplemented with audio or video comments, can enhance the readability and personal dimension of feedback. In an e-learning course designed in this way, feedback becomes a natural part of the dialogue between student and teacher and supports the development of self-regulated learning.

4 Discussion

The results of this study demonstrate that in online teacher education, feedback and assessment should not be viewed as isolated instructional events but as a comprehensive learning support system. This approach addresses one of the most significant challenges of e-learning: the lack of immediate social presence and the potential for student isolation.

Our analysis confirms that while automated feedback provides necessary immediacy and supports self-regulated learning in factual knowledge, it is insufficient for developing complex pedagogical competences. As shown in the examples from the *Didactics of the Czech Language* course, teacher-provided feedback remains indispensable for guiding students through higher-order cognitive tasks, such as lesson planning or analyzing linguistic nuances. This alignment of automated "speed" and human "depth" creates a scaffolded environment that mirrors the complexity of the teaching profession.

A key finding is the shift in students' perception of assessment. When feedback is descriptive and focused on the process rather than just the final grade (as seen in the student's emotional response to the lesson plan evaluation), it serves a motivational and transformative function. This aligns with the findings of Rapanta et al. [6], who emphasize that in the post-pandemic era, the quality of online instruction is fundamentally determined by the intentional design of teacher presence and the continuous support of

students' learning activities. By providing high-quality feedback, tutors do not only evaluate knowledge; they model professional behavior. Students who experience constructive, criteria-based feedback are more likely to implement similar formative practices in their own future classrooms.

The implementation of structured peer feedback via the Moodle Workshop module proved to be an effective tool for increasing student engagement. By shifting the tutor's role from a primary evaluator to a meta-evaluator (assessing the quality of the feedback itself), students are forced to think critically about pedagogical criteria. Recent research by Morris et al. [4] further supports this approach, identifying that combining self-directed learning with structured peer feedback is essential for maintaining student engagement and interaction in online teacher education programs. This transforms discussion forums from simple communication channels into spaces for professional dialogue and collaborative knowledge construction.

Our findings align with the principles of good feedback practice defined by Nicol and Macfarlane-Dick [5], particularly in supporting self-assessment and encouraging teacher-student dialogue. Furthermore, the emphasis on immediate, corrective feedback in self-tests supports the findings of Shute [8] regarding the power of formative feedback in digitally mediated environments. Similarly, Zhai and Ma [9] demonstrate that automated feedback mechanisms significantly contribute to the improvement of students' writing skills and self-efficacy, provided they are effectively integrated into the instructional design.

4.1 Limitations of the Study

Despite the positive findings, this study has several limitations that should be acknowledged. The analysis is based on two specific online courses within a single university. Therefore, the results may be influenced by the specific subject matter (Czech language didactics) and the institutional culture, which may limit the generalizability of the findings to other academic disciplines or contexts. The study relies primarily on qualitative analysis of feedback examples and student reflections. While this provides deep insight into the quality of the feedback process, it does not provide statistically significant data on the impact of this system on overall academic grades or long-term teaching performance. The feedback system is closely tied to the features of the Moodle learning management system (e.g., the Workshop and Assignment modules). Implementing a similar system in different digital environments might yield different results depending on the tools' functionality. The evidence of student motivation is partly based on voluntary student responses to feedback. These may not represent the views of all students, particularly those who are less engaged or less vocal in the online environment.

5 Conclusion

Within the specific context of the analyzed courses, the findings of this case study provide evidence that a systematically designed feedback system integrating automated, teacher-led, and peer-based elements-effectively supports both the cognitive development and

professional motivation of teacher education students. The integration of automated feedback, teacher-provided feedback and structured peer feedback created a coherent learning support system that addressed both cognitive and motivational aspects of learning.

Automated feedback embedded in self-tests, tests and interactive exercises proved particularly effective in supporting immediate orientation in learning and fostering self-regulated learning. When automated feedback included explanations of correct answers or guidance towards appropriate solution strategies, students were better able to understand their errors and adjust their learning approaches.

Teacher-provided feedback was especially important in tasks aimed at developing higher-order cognitive skills, such as lesson planning, reflection on teaching practice and argumentative discussion contributions. Qualitative analysis of feedback examples shows that descriptive and corrective comments helped students identify strengths and weaknesses in their work and supported deeper understanding of subject matter. The use of audio and video feedback further enhanced the personal dimension of feedback and strengthened students' perception of teacher presence in the online environment.

Structured peer feedback, implemented as a tool of formative assessment, contributed to increased student engagement and reflection. When peer feedback was guided by clearly defined criteria and supported by the learning management system, students provided more substantive and constructive comments and moved beyond superficial expressions of agreement. At the same time, evaluating the quality of peer feedback rather than the peer-assessed product reinforced students' responsibility for the learning process.

Overall, the results from these two courses suggest that feedback and assessment, when deliberately integrated and aligned with instructional goals, function as an effective learning support system in online teacher education.

References

1. Black, P., William, D.: Assessment and classroom learning. Assess. Educ.: Princ. Policy Pract. **5**(1), 7–74 (1998)
2. Garrison, D.R., Vaughan, N.D.: Blended Learning in Higher Education: Framework, Principles, and Guidelines. Jossey-Bass (2008)
3. Hattie, J., Timperley, H.: The power of feedback. Rev. Educ. Res. **77**(1), 81–112 (2007)
4. Morris, T.H., Schön, M., Drayson, M.C.: Reimagining online teacher education: combining self-directed learning with peer feedback for interaction and engagement. J. Res. Innov. Teach. Learn. **18**(2), 212–231 (2025)
5. Nicol, D.J., Macfarlane-Dick, D.: Formative assessment and self-regulated learning: a model and sever principles of good feedback practice. Stud. High. Educ. **31**(2), 199–218 (2006)
6. Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L., Koole, M.: Online university teaching during and after the Covid-19 crisis: refocusing teacher presence and learning activity. Postdigital Sci. Educ. **2**, 923–945 (2020)
7. Sadler, D.R.: Formative assessment and the design of instructional systems. Instr. Sci. **18**(2), 119–144 (1989)
8. Shute, V.J.: Focus on formative feedback. Rev. Educ. Res. **78**(1), 153–189 (2008)
9. Zhai, N., Ma, X.: The impact of automated writing evaluation on second language writing skills: a randomized controlled trial. Front. Psychol. **14**, 1249991 (2023)

PLECS for Resource-Constrained and Distance-Friendly Power Electronics Education: An Education-Focused Comparison with MATLAB/Simulink

Mark Wendler(✉), Sebestyén Gergely, and József Kopják

Obuda University, 1034 Budapest Bécsi út 96/b,, Hungary
mark.wendler@phd.uni-obuda.hu

Abstract. This paper evaluates PLECS Standalone as a simulation platform for electrical engineering education and compares it with MATLAB/Simulink from the perspective of distance learning and limited student hardware resources. We assessed tool readiness through startup and model-opening overhead, simulation throughput on representative teaching benchmarks including a buck regulator, a single-phase photovoltaic inverter, and a synchronous motor drive, and qualitative learning equivalence by verifying that key waveforms remain comparable across platforms. In addition, we report results from a pilot student group in which students could choose PLECS instead of MATLAB/Simulink for a brushed DC motor speed-control homework task. The results show that PLECS provides a faster and more lightweight workflow for typical laboratory-style iterations, while MATLAB/Simulink offers higher-fidelity Simscape-based physical component models but with greater startup overhead and significantly slower switching-system simulations in the tested configurations. Student feedback highlighted rapid iteration and a focused interface, together with an initial adjustment cost for students accustomed to MATLAB block libraries. Based on these findings, we outline next steps for course-level deployment of PLECS and a more structured student-feedback study.

Keywords: Electrical engineering education · Simulation-based learning · PLECS · MATLAB

1 Introduction

Simulation has become the "third lab bench" in modern electrical engineering education: it allows students to explore converter behavior, motor drives, and energy conversion systems before energizing real hardware. This role is even more important in distance or hybrid courses, and in high-voltage power electronics where safety is a major concern, because simulation enables experimentation without the inherent risks of working with live equipment [10–12].

In electrical engineering curricula, especially in power electronics, electrical drives, control systems, and energy-conversion subjects, practical learning is constrained by

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 457–468, 2027.
https://doi.org/10.1007/978-3-032-32104-6_39

laboratory capacity, equipment cost, and safety requirements. These constraints are visible both at Óbuda University [1] and in broader higher-education practice [3, 10]. As a result, simulation plays an increasingly important role as a safe and scalable environment in which students can explore system behavior before working with physical hardware [9].

This paper investigates the use of PLECS as a simulation platform for higher education in areas including:

1. electrical drives and rotating machines,
2. energy conversion in solar power applications,
3. general control theory

This paper presents an education-focused framework for comparing power electronics simulation tools. Beyond runtime, it considers practical teaching factors such as startup overhead, model-opening delay, hardware demands, time to first result, iteration speed, and whether the tools produce qualitatively equivalent learning-relevant waveforms. In this view, simulation software is treated not just as a numerical engine, but as part of the pedagogical infrastructure of modern laboratory education.:

- **RQ1 (Performance):** How do PLECS and MATLAB/Simulink compare in startup time and simulation runtime for representative teaching models?
- **RQ2 (Accessibility):** What hardware profiles are sufficient for completing typical lab tasks on student-owned machines?
- **RQ3 (Time-to-lab):** How do tools compare in total time to complete the lab workflow, not just solver runtime?
- **RQ4 (Student experience):** What qualitative feedback did student report when given a choice of tools?

We ran a pilot in the embedded drives course in which students could choose to use PLECS instead of MATLAB/Simulink for their simulation homework [2].

1.1 Related Work and Motivation

Tool comparison in power electronics simulation has a long history. A frequently cited example is an EE Times study [14] that compared multiple categories of simulators (equation solvers, circuit-equation solvers, and circuit solvers) using a single-phase boost rectifier as a benchmark system. This work demonstrates a key idea that remains relevant in education: tool choice strongly influences feasibility and iteration speed for switching systems [9, 10].

In educational contexts, multiple authors have compared simulation platforms for teaching power electronics circuits. For example, prior studies have compared PSIM and MATLAB/Simulink for modeling common power electronics circuits in teaching settings. Complementary work discusses the broader need for accessible simulation in education, including concerns about availability and cost, especially in resource-limited institutions [6].

What is less commonly addressed is the “distance learning reality”:

- installability on mixed student laptops,

- license friction,
- startup delays,
- and the full workflow time from “download lab materials” to “submit results”.

This paper targets that gap.

2 Tools Under Comparison

MATLAB has been used for many years in BSc and MSc electrical engineering courses at Óbuda University, which makes it a natural baseline for comparison with PLECS in the present study [2]. MATLAB will not be introduced separately, because it is a well known tool used widely in R&D and STEM education. References are available widely [8–10].

2.1 PLECS Introduction

PLECS (Piecewise Linear Electrical Circuit Simulation) is a software tool for system-level simulations of electrical circuits developed by Plexim [3]. It is especially designed for power electronics but can be used for any electrical network. PLECS includes the possibility to model controls and different physical domains (thermal, magnetic and mechanical) [5] besides the electrical system. PLECS exists in two primary forms:

1. **PLECS Standalone:** does not require MATLAB/Simulink; Plexim notes optimized solvers and emphasizes higher simulation speed.
2. **PLECS Blockset:** integrates into MATLAB/Simulink and requires MATLAB and Simulink to run.

In this research PLECS Standalone was in focus.

3 Methodology

3.1 Course Context

The evaluation is grounded in a lab structure used in embedded drive systems course, 2025/26/1 semester, Kando Kalman Electrical Engineering, where students could optionally choose between PLECS and MATLAB/Simulink for simulation tasks. This was a pilot program, first time introducing PLECS at the University. Student feedback was collected from those who selected PLECS. Eight out of fourteen students selected PLECS instead of MATLAB/Simulink. This distribution was notable because PLECS was entirely new to the students. Unlike MATLAB, which students encounter in several courses, PLECS had not previously been used in their curriculum. MATLAB is used in multiple courses where students can learn MATLAB basics [2]. At the particular test course feedback resulted none of the students had used PLECS before.

Because this was the first pilot use of PLECS in the course, student-experience data were collected in an exploratory manner through voluntary free-text feedback rather than a structured questionnaire. The sample was small and self-selected, since students chose

whether to use PLECS or MATLAB/Simulink. Accordingly, the student-feedback results are interpreted as preliminary indicators of usability themes rather than as a statistically robust educational evaluation.

The University at that time did not have agreement with Plexim for campus-wide PLECS license, therefore students had to request trial license or use demo version of PLECS. The trial version is available upon manual request for 90 days. Many of the students were requesting this licenses and they got the trial versions within a working day. Some students were just using the demo version which limitation is that models cannot be saved and data cannot be exported/imported. As the homework was relatively easy on purpose to be able to solve it in few hours, simulation results could be saved via screenshots and hand in to the University e-learning system.

3.2 The Task for Homework

The test homework was the following:

"Simulate a Brushed DC motor speed control. Use a 200W 24V motor with a fan load. Add a speed PI control loop and do step response analyses for 1000RPM and 1500RPM." The task was intentionally formulated in a relatively open-ended manner. This allowed students to apply their own modeling ideas and develop solutions that were not restricted to a single instructor-defined implementation path. Especially the new tool provides the opportunity to prevent the students copying previous semester work. These circumstances provided real variety of models handed in by the students. This experiment also showed PLECS features multiple ways to implement same problem.

3.3 Practical Deployment Constraints

Additional constraints arose from the institutional and technical environment of the university.

- Students using personal laptops with varied specs,
- limited access to department workstations,
- need for repeatable workflows that students can execute off-campus.

Additional constraints are the limited performance of the laboratory PCs and the update/upgrade period of SW system on the laboratory tools.

3.4 Benchmark Models

This paper compares realistic educational workflows rather than identical solver setups. The benchmark circuits were built in PLECS Standalone using its native power-electronics modeling approach, while MATLAB/Simulink models used Simscape Electrical physical blocks, which typically produce DAE-based formulations. As a result, the runtime differences reflect both software efficiency and the modeling methods that students and educators would normally use. This matters in practice because the default workflow in each platform shapes the actual time-to-lab experience more than an artificially matched solver comparison.

Independently of the student homework task, we selected three benchmark models to evaluate the tools on representative electrical engineering education problems. The selected benchmark types are also available in MATLAB/Simulink, or can be implemented using model structures closely aligned with those used in MATLAB-based university coursework, which enables a practically meaningful basis for comparison. The benchmarks comprise (i) a synchronous buck regulator, (ii) a single-phase photovoltaic inverter, and (iii) a synchronous motor drive with speed control. Together, these models provide a comparable foundation for measuring metrics such as startup time, simulation runtime, and workflow effort in both environments. Due to space constraints, this paper presents detailed results for the buck regulator benchmark only, while the remaining models are used to support the overall evaluation.

Test Model: Buck Converter
A simple buck converter stage with output voltage control was implemented in both PLECS and MATLAB/Simulink based on the respective vendor demo models. The synchronous buck examples were then modified to align component values, control structure, and simulation settings as closely as possible, aiming for a near one-to-one comparison between the two environments. The buck regulator represents a typical entry-level power electronics problem and is taught in the BSc electrical engineering curriculum [2] (Figs. 1 and 2).

PLECS Model used:

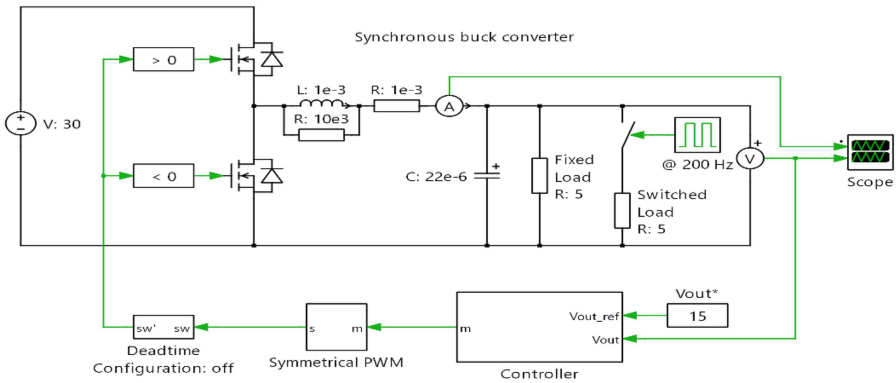


Fig. 1. Modified PLECS simulator synchronous buck converter demo model [4, 5].

MATLAB Model used:

3.5 Metrics

To compare PLECS and MATLAB/Simulink in an education-oriented context, we selected metrics that capture not only numerical solver speed, but also the practical overhead that students experience when completing simulation-based assignments. In introductory and intermediate power-electronics labs, learning progress is strongly influenced by how quickly students can obtain a first correct waveform and how efficiently

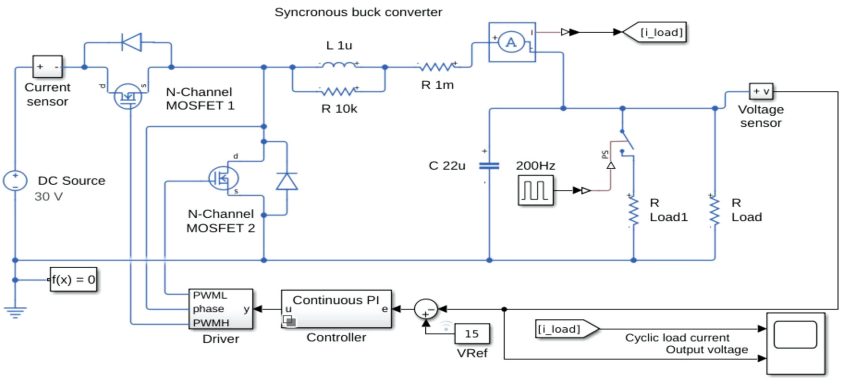


Fig. 2. Modified MATLAB synchronous buck converter demo model [6].

they can iterate on parameters. Therefore, the evaluation combines (i) tool readiness and workflow overhead, (ii) simulation throughput, and (iii) student-reported usability and friction.

(M1) Workflow Overhead and Readiness

These metrics quantify "time-to-first-waveform," which is critical in short lab sessions and distance-learning settings where students work on heterogeneous computers and often have limited time windows to finish tasks. We measure:

- Cold start time - launch application to ready state
- Project open time - open model file to editable state
- First run time - click simulate to first plotted result
- Total time-to-lab (education metric)

  – install/activate tool (if needed)
  – open model
  – run baseline simulation
  – change one parameter (e.g., switching frequency or controller gain)
  – re-run and export plot(s)
  – save results and submit

(M2) Simulation Runtime

These metrics focus on iteration speed once the model is built and help estimate how many parameter sweeps or controller-tuning attempts can realistically fit into a lab session.

- runtime for a fixed simulated duration 1 s and fixed step/solver settings where comparable with the listed models in chapter 3.4 Benchmark models.
- Comparing the differences in simulation results

(M3) Student-Reported Usability
Because simulation tools are ultimately educational instruments, we also collected student impressions of usability, perceived speed, and friction points. In the pilot offering, only eight students selected PLECS, so feedback was collected as free-text comments and evaluated manually. This approach was sufficient to identify recurring themes (e.g., installation friction, ease of iteration, and interface adaptation for students accustomed to MATLAB/Simulink).

4 Results

4.1 Workflow Overhead and Resource Footprint

Tool readiness and resource footprint were measured on two representative machines: a modern student-grade laptop (Intel Core Ultra 7 165H, 32 GB RAM, PCIe SSD, Windows 11) and an older, resource-limited PC (Intel Core i5-9400F, 8 GB RAM, SATA SSD, Windows 10). In both cases, PLECS Standalone reached a "ready-to-model" state faster because the application opens directly into the modelling environment, whereas MATLAB requires launching MATLAB first and then opening Simulink for block-diagram modelling (Table 1).

Table 1. Startup and model opening times (seconds).

Metric (s)	Intel Ultra 7 PC PLECS	Intel Ultra 7 PC MATLAB	Intel i5 PC PLECS	Intel i5 PC MATLAB
Cold start (application ready, no model)	1.9 s	6.75 s	2.4 s	15.3 s
Open model from already opened tool	<1 s	9.5 s	<1 s	28 s
Open model from file explorer (tool not running)	2.1 s	21 s	3.1 s	51 s

Memory usage also differed greatly after one simulation run. MATLAB used about 1400 to 1950 MB, while PLECS used only about 305 to 330 MB. This lower memory footprint is especially important for older or low-cost student laptops, where high memory use can reduce responsiveness. Overall, the results suggest that PLECS lowers non-learning overhead and resource demands without changing the conceptual understanding students gain from the buck regulator simulation.

Installation and Activation of PLECS and MATLAB:
Installation and activation also affected the overall time-to-lab experience. PLECS Standalone uses a single installer of about 371 MB and can be installed without administrator privileges, which makes deployment easier on student laptops and university PCs. In our

pilot course, students could start quickly using either a 90-day trial license or the demo mode [3].

By contrast, MATLAB/Simulink usually requires the MATLAB platform plus additional products such as Simulink and Simscape Electrical, leading to a much larger installation footprint. MathWorks lists about 3.8 GB for MATLAB alone, 4 to 6 GB for a typical installation, and up to 23 GB for all products. As a result, PLECS generally offers lower onboarding friction, while MATLAB/Simulink provides a broader ecosystem at the cost of more complex installation and activation [7].

Institutional Accessibility
From an institutional accessibility perspective, licensing cost is also a practical constraint for scaling simulation-based labs. Óbuda University obtained a campus-wide PLECS license at no cost under our agreement with Plexim, whereas MATLAB/Simulink typically involves a recurring licensing cost for institutions. This difference affects whether all students can be supported with equal access in on-campus and distance-learning scenarios [3].

4.2 Simulation Results

The PLECS and MATLAB/Simulink waveforms for the buck regulator show strong qualitative agreement, with similar steady-state levels and transient responses during step changes. Small differences in peaks and settling are not pedagogically important, since both tools illustrate the same physical mechanisms and parameter-response relationships. For introductory teaching, they provide equivalent conceptual understanding, while the main difference lies in workflow efficiency, such as time-to-first-result and iteration speed (Figs. 3 and 4).

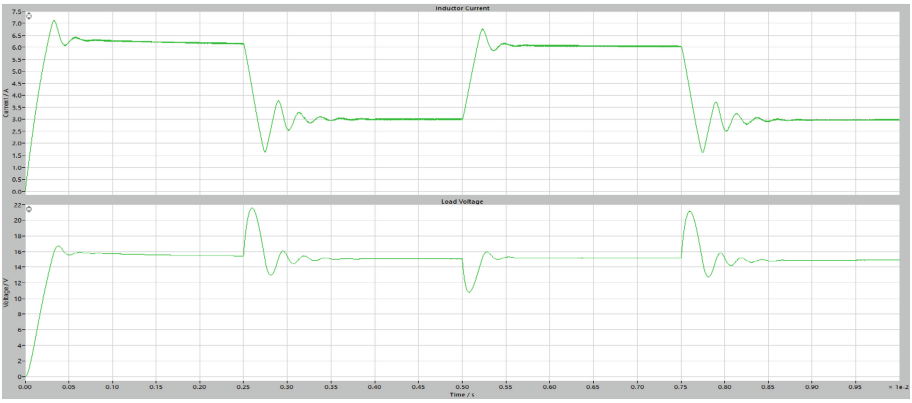


Fig. 3. PLECS simulation result

Simulation Performance:
MATLAB/Simulink was much slower than PLECS Standalone for the 1 s benchmarks:

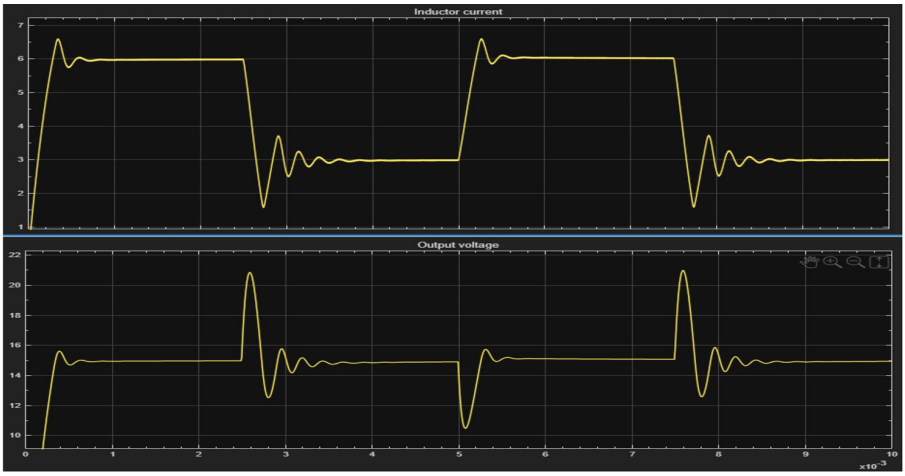


Fig. 4. MATLAB simulation result

buck regulator (2.6 s vs. >60 s), single-phase PV inverter (3.7 s vs. >60 s), and synchronous motor drive (4.9 s vs. >60 s). This is not a strict solver-to-solver comparison, since the MATLAB models used Simscape Electrical physical components, which generate DAE-based systems and increase computational cost, especially in switching circuits [7].

In education, these Simscape models are appealing because they provide high fidelity and ready-made components, but improving runtime often requires extra model simplification and solver tuning that goes beyond introductory lab goals. PLECS, by contrast, offers fast system-level power-electronics models out of the box, enabling quicker iteration and shorter time-to-lab [3, 5].

Recent PLECS versions also add PLECS Spice, which allows SPICE netlist import for device-level detail when needed. This supports selective refinement, so users can keep most of a model fast and system-level while applying detailed SPICE modeling only to the circuit section of interest [5, 6].

4.3 Students Feedback

Eight students selected PLECS for the simulation assignment. Because this was a first pilot deployment and participation in the tool choice was voluntary, the student-feedback dataset should be regarded as a small exploratory sample rather than a controlled comparative study. Feedback was collected in qualitative free-text form and reviewed manually to identify recurring themes related to installation, usability, iteration speed, and perceived focus of the interface.

Three of eight students requested a 90-day trial license from Plexim and received activation within one working day, while five students completed the task using the demo license. Despite the mixed licensing modes, all eight students successfully produced the required simulation results for the brushed DC motor speed-control problem.

The submitted solutions showed substantial diversity: most students built their models from scratch, while others started from existing demo examples and modified them to match the assignment requirements, indicating that the tool supports multiple modelling pathways for the same control problem.

Qualitative feedback was predominantly positive and centered on workflow efficiency and focus, with comments such as: "Faster and up to the point tool," "Tool looks a bit outdated, but super fast," "Easy to install and use," "No distracting features," "I will use it for self projects," and "Fast simulation."

One recurring theme was the initial adjustment cost for students coming from MATLAB/Simulink: "Hard to find blocks after MATLAB experience, but after model is made it is super fast to iterate for finding optimal parameters." Overall, the feedback suggests that once the initial user-interface unfamiliarity is overcome, students perceive PLECS as a streamlined environment that supports rapid iteration, which aligns well with the learning goal of exploring parameter effects and controller tuning in a limited time window.

4.4 Limitations

The study has several limitations. First, the benchmark comparison reflects realistic educational model implementations rather than mathematically identical formulations in the two tools. Second, the student-feedback sample is small and self-selected. Third, detailed quantitative educational outcomes such as completion time, error rate, or concept retention were not measured in the pilot. These limitations motivate a larger follow-up study with more controlled benchmarking and structured student evaluation.

5 Conclusion

This paper evaluated PLECS Standalone versus MATLAB/Simulink from an education-focused viewpoint.

RQ1 (Performance): PLECS reached a ready-to-model state faster on both tested PCs and completed the benchmark simulations much faster for 1 s of simulated time, with PLECS at 2.6 to 4.9 s versus MATLAB at over 60 s for all three models. This gap is partly influenced by modeling choices, since MATLAB used Simscape Electrical physical components, while PLECS used a system-level piecewise-linear approach.

RQ2 (Accessibility): PLECS used much less memory after one run, about 305 to 330 MB compared with about 1400 to 1950 MB for MATLAB/Simulink. This is especially relevant for student-owned computers with limited RAM. Installation and activation were also simpler because of the smaller software footprint and single-tool workflow.

RQ3 (Time-to-lab): In PLECS, faster startup, quicker model opening, lower memory use, and shorter runtimes reduced non-learning overhead and supported rapid iteration, which is important for lab-based learning and distance or hybrid delivery.

RQ4 (Student experience): Eight students used PLECS in the pilot, and all successfully completed the brushed DC motor speed-control task. Feedback highlighted speed and focus, with comments describing it as fast, straightforward, and free of distracting

features. The main drawback was an initial adjustment period for students accustomed to MATLAB's block libraries.

5.1 Next Steps

After the pilot, Óbuda University obtained a campus-wide PLECS license at no cost through its agreement with Plexim, with responsive vendor support during onboarding. In 2025/26/II, the Electronic Converters course will use PLECS as the only simulation tool, and the evaluation will be repeated with a more structured feedback form for a larger student group.

Future assessment will include Likert-scale items and targeted open-ended questions to measure usability, installation effort, learning curve, and iteration efficiency more systematically. We also plan to evaluate recent PLECS Spice features in a workflow that combines fast system-level modeling with selective higher-fidelity SPICE detail.

References

1. Allmeling, J.H., Hammer, W.P.: PLECS-piece-wise linear electrical circuit simulation for Simulink. In: Proceedings of the IEEE 1999 International Conference on Power Electronics and Drive Systems. PEDS'99 (Cat. No.99TH8475), vol. 1, pp. 355–360, Hong Kong (1999). https://doi.org/10.1109/PEDS.1999.794588
2. Curriculum of MSc Electrical Engineering "F" course at Obuda University (2026). [Online]. https://kvk.uni-obuda.hu/villamosmernoki-msc/villamosmernoki-msc-mintatantervek/. Accessed 12 Feb 2026
3. PLEXIM Homepage. https://www.plexim.com/. Last accessed 26 Jan 2026
4. Allmeling, J., Hammer, W., Schönberger, J.: Transient simulation of magnetic circuits using the permeance-capacitance analogy. In: 2012 IEEE 13th Workshop on Control and Modeling for Power Electronics (COMPEL), pp. 1–6, Kyoto (2012). https://doi.org/10.1109/COMPEL.2012.6251786
5. PLECS User Manual Version 4.9. https://www.plexim.com/sites/default/files/plecsmanual.pdf. Last accessed 26 Jan 2026
6. De Doncker, R.W., Pulle, D.W.J., Veltman, A.: Advanced Electrical Drives: Analysis, Modeling, Control, 2nd edn. Power Systems. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-48977-9
7. The MathWorks. MATLAB version: 25.2.0 (R2025b) (2026). https://www.mathworks.com. Accessed 30 Jan 2026
8. Yadav, J.G., Yadav, Y.K., Kumar, N.: Mathematical modelling & simulation of synchronous buck converter and analysis of its simulation results. In: 2023 International Conference on IoT, Communication and Automation Technology (ICICAT), pp. 1–6, Gorakhpur (2023). https://doi.org/10.1109/ICICAT57735.2023.10263672
9. Králiková, R., Lumnitzer, E.: Application of the simulation tools in the educational process. In: 2022 20th International Conference on Emerging eLearning Technologies and Applications (ICETA), pp. 348–351, Stary Smokovec (2022). https://doi.org/10.1109/ICETA57911.2022.9974804
10. Campos, N., Nogal, M., Caliz, C., et al.: Simulation-based education involving online and on-campus models in different European universities. Int. J. Educ. Technol. High. Educ. **17**, 8 (2020). https://doi.org/10.1186/s41239-020-0181-y

11. Horváth, L., Kovács, L.: Software engineering for modeling, simulation, and collaboration in high automation. In: Fujita, H., Hernandez-Matamoros, A., Watanobe, Y. (eds.) Frontiers in Artificial Intelligence and Applications. IOS Press (2025). https://doi.org/10.3233/FAIA250527
12. Kovács, J., Benyó, I., Lipovszki, G.: Simulation tools in control engineering education. In: Simulation in Industry: Proceedings of the 16th European Simulation Symposium San Diego, pp. 335–340. Society for Computer Simulation (2004)
13. Takács, P., Szakál, Z.: Present and future simulation tools, robotics and IT solutions in health care and health science education, new technologies in health, challenges in the field of pedagogy. In: Innovative Digital Curriculum Development in Higher Education: INDEHEP E-HANDBOOK, pp. 369–430. Debreceni Egyetem Egészségtudományi Kar, Debrecen (2024)
14. Rodríguez, J., Weinstein, A., Lezana, P.: A comparison of power-electronics simulation tools. EE Times (2001). [Online]. https://www.eetimes.com/a-comparison-of-power-electronics-simulation-tools/. Accessed 13 Feb 2026

Assessing Cognitive Load in Technology-Enhanced Learning Environments: An Interface-Based Evaluation Framework

Iryna Yaroshchuk[1](✉), Oksana Gomotiyk[1], Oleksandr Kolesnyk[1], Lubomír Pekarcik[2], Tomáš Lengyelfalusy[3], and Filip Makan[4]

[1] West Ukrainian National University, 11 Lvivska Str., Ternopil 46009, Ukraine
i1yaroshchuk@gmail.com

[2] Faculty of Theology, Catholic University in Ruzomberok, 034 01 Ruzomberok, Slovak Republic
lubomir.pekarcik@ku.sk

[3] DTI University, Sládkovičova 533/20, 018 41 Dubnica nad Váhom, Slovak Republic
lengyelfalusy@dti.sk

[4] Faculty of Theology, Theological Institute, The Catholic University in Ružomberok, Spišská Kapitula 12, 053 04 Spišské Podhradie, Slovak Republic

Abstract. The expansion of technology-enhanced learning environments has made digital interfaces central mediators of cognitive activity, positioning interface design as a determinant of learning effectiveness. While cognitive load theory is widely applied to instructional design, interface-induced cognitive load remains insufficiently systematised. This study proposes an interface-based framework for cognitive load assessment in digital learning contexts.

The framework integrates cognitive load theory, human–computer interaction, and educational analytics, conceptualising cognitive load as a multidimensional construct shaped by visual, navigational, and interactional features. Subjective, behavioural, and analytical indicators are consolidated into a normalised, weighted integrated cognitive load index. By holding instructional content constant across interface configurations, the model isolates interface-induced extraneous load.

Results indicate that variations in visual density, navigation complexity, and interaction sequencing significantly affect cognitive load, task efficiency, and cognitive fatigue. The study advances interface-centred cognitive load assessment and provides a scalable tool for comparative analysis and evidence-based design of inclusive digital learning environments.

Keywords: Education · Cognitive Load Assessment · Interface Design · Technology-Enhanced Learning · Human–Computer Interaction · Learning Analytics · Inclusive Digital Learning

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 469–481, 2027.
https://doi.org/10.1007/978-3-032-32104-6_40

1 Introduction

The widespread adoption of technology-enhanced learning environments has profoundly altered the cognitive conditions under which learning takes place in higher education and professional training systems. Learning management systems, digital platforms, and web-based instructional interfaces increasingly mediate learners' interaction with educational content, positioning interface design as a critical determinant of learning effectiveness. While digitalisation has enhanced accessibility and scalability, it has simultaneously introduced additional cognitive demands that may undermine learning outcomes, particularly in poorly structured or visually overloaded interfaces.

Within this context, cognitive load has emerged as a central construct for understanding learner performance in digital environments. Cognitive Load Theory posits that learning efficiency depends on the balance between intrinsic task complexity, germane cognitive processing, and extraneous demands imposed by instructional design. In technology-enhanced learning, extraneous cognitive load is frequently amplified by interface-related factors such as visual clutter, fragmented navigation, redundant information streams, and inefficient interaction patterns. Despite increasing recognition of these challenges, systematic approaches to assessing interface-induced cognitive load remain fragmented and methodologically inconsistent.

Existing research on cognitive load in digital learning environments predominantly relies on isolated measurement instruments or context-specific experimental designs. Subjective rating scales, behavioural indicators, and performance-based measures are often applied independently, limiting comparability across studies and constraining generalisability. Moreover, many empirical investigations focus primarily on learning outcomes without explicitly isolating the contribution of interface characteristics to overall cognitive demand. As a result, interface-induced cognitive load is frequently treated as a secondary analytical concern rather than a central object of evaluation.

From an applied informatics and human–computer interaction perspective, this represents a substantial limitation. Technology-enhanced learning systems are inherently interactive, and their effectiveness depends not only on pedagogical content but also on the cognitive efficiency of the interface structures through which learners engage with that content. In the absence of coherent evaluation frameworks, instructional designers, platform developers, and educators lack robust analytical tools for diagnosing cognitive inefficiencies and systematically improving interface design.

Recent advances in learning analytics, user experience research, and educational data mining provide opportunities to address this gap by integrating subjective workload assessments, performance indicators, and interface-related metrics. However, these elements are rarely combined within unified analytical structures capable of supporting systematic comparison, methodological validation, and practical application across diverse learning environments.

In response to these limitations, this study proposes an interface-based evaluation framework for assessing cognitive load in technology-enhanced learning environments. The framework consolidates multiple cognitive load indicators into a coherent methodological structure that explicitly links interface characteristics with measurable cognitive

outcomes. By adopting an assessment-oriented perspective, the study provides a replicable basis for evaluating cognitive efficiency and supports evidence-based design of cognitively efficient digital learning environments.

In response to the identified research gaps, this study is guided by a set of interrelated research questions aimed at conceptualising and evaluating interface-induced cognitive load in technology-enhanced learning environments.

First, the study examines how interface-induced cognitive load can be systematically conceptualised and operationalised within a multidimensional analytical framework. Second, it explores the extent to which the integration of subjective, behavioural, and analytical indicators enables a reliable and coherent assessment of cognitive load. Third, the study investigates how variations in interface design parameters influence the integrated cognitive load index under controlled instructional conditions.

2 Literature Review

Cognitive Load Theory (CLT) constitutes the dominant theoretical framework for analysing learning efficiency in technology-enhanced environments. Originally formulated by Sweller [1] and further refined by Paas, Renkl, and Sweller [2], CLT conceptualises learning as a cognitively constrained process governed by the limited capacity of working memory. It distinguishes between intrinsic cognitive load, determined by task complexity; extraneous cognitive load, induced by instructional design and information presentation; and germane cognitive load, associated with schema construction and automation.

Contemporary developments of CLT emphasise the central role of extraneous cognitive load arising from instructional and interface-related characteristics, particularly in digitally mediated learning environments [3]. In contexts dominated by visual and interactive channels, extraneous cognitive load has been consistently identified as a critical risk factor. Mayer [4] demonstrated that excessive multimedia elements, redundant visual cues, and fragmented presentation structures impair learning outcomes by consuming cognitive resources without supporting meaningful knowledge construction. Subsequent studies confirm that interface design decisions directly influence the balance between cognitive load components, especially in web-based and learning management system (LMS) environments [4, 5].

Recent research has extended CLT beyond instructional content to encompass interface-level characteristics. Seufert, Guggemos, and Sailer [6] showed that learners with limited cognitive resources are disproportionately affected by visually complex digital environments, leading to performance degradation even when instructional content remains constant. These findings indicate that cognitive load in technology-enhanced learning is not solely a pedagogical phenomenon, but is fundamentally shaped by interface structure and interaction design.

Within human–computer interaction (HCI) and educational technology research, the relationship between interface design and cognitive overload has gained increasing attention. Visual complexity, often operationalised through the concept of visual noise, refers to excessive or poorly structured visual stimuli such as animations, decorative elements, saturated colour schemes, and non-hierarchical layouts. Empirical evidence consistently

demonstrates that such features increase attentional fragmentation and cognitive fatigue, thereby elevating extraneous cognitive load [7, 8].

In online and hybrid education contexts, interface-induced overload is particularly pronounced. Increased visual complexity in digital learning interfaces has been shown to significantly raise perceived cognitive load and impair task performance, even when instructional content remains constant [8]. Similar effects have been reported in higher education, where interface disorganisation is associated with reduced engagement, weakened self-regulation, and increased dropout intentions. Despite this evidence, visual complexity is typically examined in isolation, rather than embedded within integrated, interface-level cognitive load evaluation models.

Existing research on cognitive load in technology-enhanced learning can be broadly grouped into three dominant streams: (i) studies grounded in cognitive load theory and multimedia learning principles; (ii) HCI-oriented investigations focusing on visual complexity and attentional mechanisms; and (iii) measurement-oriented research employing subjective, behavioural, or analytics-based indicators. However, these streams remain largely disconnected, resulting in fragmented analytical approaches and a lack of comprehensive frameworks linking interface characteristics with multidimensional cognitive load assessment.

Interface complexity further interacts with individual cognitive profiles. Research on executive functioning indicates that learners with reduced working memory capacity or attentional control are particularly vulnerable to visual overload [9], positioning interface design as a critical factor for cognitive accessibility and inclusion. This underscores the need for systematic, evidence-based evaluation approaches rather than ad hoc design heuristics.

Despite the conceptual maturity of CLT, cognitive load measurement in technology-enhanced learning remains methodologically fragmented. Subjective instruments such as the NASA Task Load Index (NASA-TLX) are widely adopted due to their reliability [10], yet reliance on single measures limits analytical depth. Performance-based indicators and learning analytics provide complementary insights but are rarely integrated into coherent models that explicitly attribute cognitive load variations to interface-related factors rather than content complexity.

Recent studies highlight the need for multidimensional assessment strategies. Petrovič et al. [11] demonstrated that cognitive and emotional states jointly influence academic performance, while Zenelaga, Goga, and Králik [12] emphasised that inclusive digital environments require evaluation frameworks capable of identifying latent cognitive barriers beyond surface-level usability indicators. Although the principles of Universal Design for Learning (UDL) advocate the proactive reduction of cognitive barriers through adaptable and predictable learning environments [13], their operationalisation in interface-level evaluation practices remains limited. Research on inclusive virtual learning communities further indicates that digital inclusion depends not only on access, but also on the cognitive ergonomics of digital systems [14].

A persistent limitation in the existing literature is the absence of structured frameworks that consolidate heterogeneous cognitive load indicators into analytically robust models. Many studies remain highly context-specific, limiting systematic comparison across platforms and learning scenarios. As noted by Shyshak et al. [15], the pace of

educational digitalisation has exceeded the development of assessment tools addressing cognitive safety and sustainability in online environments.

In synthesis, the reviewed literature reveals three recurrent gaps: insufficient integration of interface-level analysis into cognitive load research; fragmentation of existing measurement approaches; and the lack of operational evaluation tools aligned with inclusive education principles. Addressing these limitations requires a shift from isolated empirical studies towards structured, replicable evaluation frameworks that explicitly link interface characteristics with multidimensional cognitive load indicators. This study responds to this need by proposing an interface-based evaluation framework for systematic cognitive load assessment in technology-enhanced learning environments.

3 Research Methodology

The research methodology employs an analytical, framework-oriented approach to assessing cognitive load in technology-enhanced learning environments. Rather than relying on isolated experimental designs, the proposed approach develops a generalisable evaluation model that systematically links interface characteristics of digital learning environments with measurable cognitive load indicators.

To complement the conceptual development of the framework, the study adopts a pilot comparative research design. A controlled evaluation scenario was implemented in which participants interacted with two alternative interface configurations under identical instructional conditions. The configurations differed in visual density, navigation complexity, and interaction sequencing, allowing the isolation of interface-induced cognitive load effects.

The study follows a within-subject experimental design, enabling direct comparison of cognitive load indicators across interface scenarios for the same participants. Data collection was based on a combination of subjective workload assessments, behavioural performance metrics, and system-generated interaction logs, providing a multidimensional evaluation of cognitive load.

The study follows an assessment-driven evaluation logic, where the primary object of analysis is the interface structure and user interaction patterns rather than instructional content itself. This orientation is consistent with applied informatics and established research traditions in human–computer interaction and learning analytics. To isolate interface-induced effects, instructional content is assumed to remain constant across the compared interface configurations, allowing observed variations in cognitive load to be attributed primarily to interface-related factors.

The evaluation framework integrates principles from cognitive load theory, interface design, and educational analytics, conceptualising cognitive load as a multidimensional construct shaped by the visual, navigational, and interactional properties of the digital environment. Accordingly, the framework is structured around three core dimensions: (1) a visual dimension, encompassing information density, visual hierarchy, colour schemes, and visual noise; (2) an interaction dimension, reflecting navigation complexity, action requirements, and transition coherence; and (3) a cognitive–behavioural dimension, capturing perceived cognitive load, task performance efficiency, and indicators of cognitive fatigue.

This multidimensional structure enables a transition from fragmented measurement practices towards a holistic assessment of cognitive effectiveness in technology-enhanced learning environments. The conceptual structure of the proposed interface-based evaluation framework is illustrated in Fig. 1.

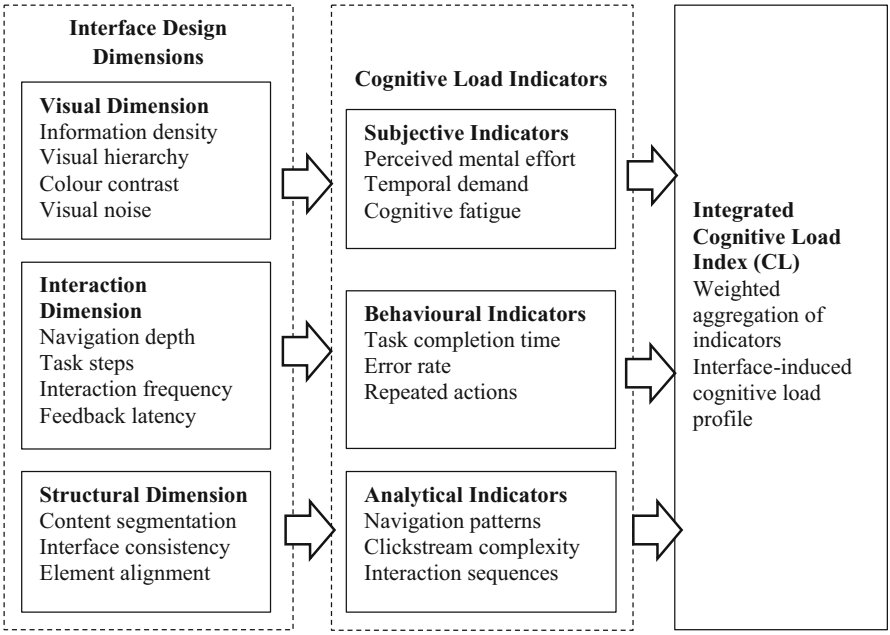


Fig. 1. Interface-based cognitive load evaluation framework illustrating the relationships between interface design dimensions, cognitive load indicators, and the integrated assessment output in technology-enhanced learning environments.

To operationalise cognitive load within the proposed framework, a combination of subjective, behavioural, and analytical indicators is employed. This integrative approach mitigates the limitations inherent in relying on a single type of metric and enhances the overall validity of the assessment.

The key groups of indicators include:

- Subjective indicators, reflecting users' perceived workload, such as multidimensional rating scales capturing cognitive, temporal, and mental effort.
- Behavioural indicators, encompassing task completion time, error frequency, repeated actions, and the rate of returns to previous interface elements.
- Analytical indicators, derived from digital traces of user interaction with the system, including log files, action sequences, and navigation patterns.
- The indicators are normalised and aggregated within the framework to ensure comparability across different interface configurations or learning scenarios.

The operationalisation of the proposed evaluation framework, including its dimensions, indicators, and corresponding data sources, is summarised in Table 1.

Table 1. Dimensions, indicators, and data sources used for interface-based cognitive load assessment in technology-enhanced learning environments

Dimension	Indicator	Description	Data Source
Visual	Information density	Number of visual elements simultaneously presented on a single interface screen	Interface layout analysis
Visual	Visual hierarchy	Clarity and consistency of element prioritisation and grouping	Expert interface evaluation
Visual	Visual noise	Presence of redundant or competing visual stimuli	Interface complexity assessment
Interaction	Navigation depth	Number of steps required to complete a learning task	System log data
Interaction	Task steps	Number of user actions required to achieve task completion	Interaction sequence analysis
Interaction	Feedback latency	Time delay between a user action and the corresponding system response	Platform performance logs
Structural	Content segmentation	Degree of logical separation and chunking of learning content	Interface structure analysis
Structural	Interface consistency	Uniformity of design patterns and interaction logic across the interface	Expert usability review
Cognitive	Perceived mental effort	Learners' subjective assessment of experienced mental workload	Questionnaire-based evaluation
Cognitive	Temporal demand	Perceived time pressure during task execution	Subjective workload scales
Behavioural	Task completion time	Time required to successfully complete assigned tasks	Learning analytics logs
Behavioural	Error rate	Frequency of incorrect actions or task failures	System event records
Analytical	Navigation patterns	Structure and regularity of users' navigation paths	Clickstream analysis
Analytical	Interaction sequences	Complexity and variability of users' action sequences	Log-file mining

To ensure analytical rigour, the proposed evaluation framework incorporates a formalised representation of an integrated cognitive load index. In its generalised form, interface-induced cognitive load can be expressed as a weighted aggregation of multiple indicators:

$$CL = \sum_{i=1}^{n} w_i \cdot I_i, \tag{1}$$

where CL denotes the integrated cognitive load index,

I_i represents normalised interface-related cognitive load indicators, and w_i denotes weighting coefficients reflecting the relative importance of the corresponding indicators.

The determination of weighting coefficients represents a critical component of the proposed framework. In the baseline configuration, weights are assigned using a hybrid approach that combines expert-based assessment with normalization procedures. Initially, the relative importance of indicator groups (visual, interactional, cognitive–behavioural) is established through expert judgement, reflecting their contribution to interface-induced cognitive load.

Subsequently, the weights are normalised to ensure comparability and consistency of the aggregated index, satisfying the condition $\sum w_i = 1$. This approach provides methodological flexibility while maintaining analytical coherence of the model.

In advanced applications, the weighting scheme may be further refined using data-driven techniques, such as regression-based estimation, entropy weighting, or machine learning methods, enabling adaptive calibration of the model to specific learning environments and user populations.

The integrated cognitive load index (CL) represents a normalised and weighted aggregation of interface-induced cognitive load indicators and serves as an evaluative output of the proposed framework rather than a direct empirical measurement. Its primary function is to enable systematic comparison between alternative interface configurations or learning scenarios under controlled instructional conditions.

Validation of the proposed evaluation framework is conducted by analysing the sensitivity of the integrated cognitive load index to variations in interface characteristics, as well as by examining the internal consistency across different groups of indicators. Such validation supports the analytical robustness of the framework and its capacity to discriminate between cognitively efficient and cognitively demanding interface designs.

The application of the framework is oriented towards comparative assessment of alternative interface solutions or learning configurations, with the objective of identifying design patterns that minimise extraneous cognitive load. Importantly, the proposed methodology is not confined to a specific platform, instructional domain, or educational level, thereby ensuring scalability and relevance across a wide range of technology-enhanced learning environments.

4 Results and Discussion

The application of the proposed interface-oriented framework enabled a comprehensive assessment of cognitive load in technology-enhanced learning environments across different interface design configurations. The analysis was conducted through a systematic

comparison of alternative interface scenarios that varied in visual intensity, navigation structure, and the complexity of user interaction with the learning environment.

As an initial step towards empirical validation, the proposed framework was applied in a pilot case involving two alternative interface configurations of a technology-enhanced learning module. The configurations differed in visual structure, navigation depth, and interaction sequencing, while instructional content remained constant.

The pilot application of the framework demonstrates its practical feasibility and sensitivity to interface-induced variations in cognitive load. The integrated cognitive load index captured consistent differences between interface configurations, with higher values observed in environments characterised by increased visual density and navigation complexity.

These findings confirm the sensitivity of the integrated cognitive load index to variations in interface parameters even when instructional content remains constant.

Comparative analysis of the examined interface configurations revealed a clear distinction between cognitively efficient and cognitively overloaded environments. Interfaces characterised by a well-defined visual hierarchy, a limited number of concurrently active elements, and a predictable navigation logic exhibited lower levels of integrated cognitive load.

By contrast, interfaces displaying fragmented structural organisation, an excessive number of interactive components, and unclear action sequencing were associated with higher perceived cognitive load, longer task completion times, and reduced accuracy of user actions. This pattern is consistent with the principles of cognitive load theory and provides analytical support for the relevance of an interface-oriented approach to cognitive load evaluation.

A comparative illustration of integrated cognitive load levels across alternative interface configurations is presented in Fig. 2.

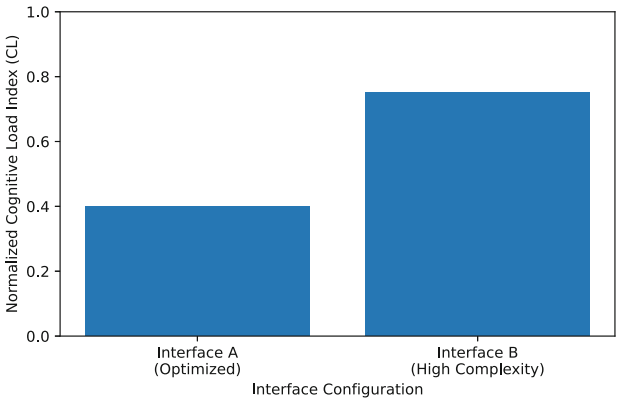


Fig. 2. Conceptual comparison of normalised cognitive load levels across interface configurations with different degrees of visual and interaction complexity.

Figure 2 presents a conceptual comparison of normalised cognitive load levels observed across two interface configurations under conditions of constant instructional

content. The optimised interface (Interface A) exhibits a lower integrated cognitive load index compared to the visually complex interface (Interface B), underscoring the role of interface design in the induction of extraneous cognitive load.

The values represent normalised cognitive load scores derived from the pilot application of the proposed framework.

The observed differences are theoretically grounded in cognitive learning theory and established findings within the field of human–computer interaction. The increase in cognitive load associated with visually overloaded interfaces can be interpreted as a consequence of competition for limited working memory resources, which reduces the efficiency of processing task-relevant instructional information.

Interface characteristics that minimise additional cognitive effort related to orientation and navigation facilitate the reallocation of cognitive resources towards the execution of learning tasks. Accordingly, the findings confirm that the cognitive effectiveness of a digital learning environment is determined not only by the quality of instructional content but also by the architecture of user interaction with that content.

One of the key strengths of the proposed framework lies in its analytical flexibility and scalability. The integration of multiple indicator types within a unified model mitigates the subjectivity inherent in individual measures and enhances the stability of evaluation outcomes. Sensitivity analysis indicates that variations in individual indicators do not produce disproportionate fluctuations in the integrated cognitive load index, thereby demonstrating the robustness of the proposed model.

Furthermore, the framework supports the comparative evaluation of alternative interface solutions without necessitating full-scale experimental studies in each specific context. This characteristic makes it particularly suitable for applied use in the design, evaluation, and iterative improvement of digital learning environments.

From a practical perspective, the results have significant implications for learning platform developers, instructional designers, and specialists in educational analytics. The proposed evaluation framework can be employed as a tool for the preliminary audit of interface-related cognitive efficiency, as well as for monitoring changes in cognitive load following the introduction of new functional or visual elements.

The framework is especially valuable in the context of inclusive digital learning, where even minor interface complexities may translate into substantial cognitive barriers. Systematic cognitive load assessment enables a transition from intuition-driven design practices towards evidence-based approaches in the development of technology-enhanced learning environments.

A comparative positioning of the proposed framework relative to existing cognitive load assessment approaches is summarised in Table 2.

Unlike the approaches identified in the literature, which predominantly rely on isolated cognitive load measurements or content-oriented evaluation, the proposed framework introduces an explicitly interface-centred and multidimensional analytical structure. This enables systematic comparison across interface configurations and integrates heterogeneous indicators within a unified model. At the same time, the framework remains limited by its reliance on pilot validation and requires further large-scale empirical testing.

As demonstrated in Table 2, the proposed framework extends existing cognitive load assessment approaches by shifting the analytical focus from content- or outcome-oriented evaluation towards an interface-centred, multidimensional assessment model that enables systematic comparison across alternative configurations of digital learning environments.

5 Conclusions

This study proposes an interface-oriented approach to cognitive load assessment in technology-enhanced learning environments by integrating principles of cognitive load theory, human–computer interaction, and educational analytics. Unlike conventional approaches that predominantly focus on instructional content or learning outcomes, the proposed framework foregrounds interface characteristics as a key source of extraneous cognitive load in digital learning systems.

Table 2. Comparison of existing cognitive load assessment approaches and the proposed interface-based evaluation framework

Criterion	Existing Cognitive Load Assessment Approaches	Proposed Interface-Based Evaluation Framework
Primary focus	Learning outcomes or instructional content	Interface-induced cognitive load
Consideration of interface design	Implicit, fragmented, or secondary	Explicit and systematic
Cognitive load dimensions	Mostly subjective or single-dimensional	Multi-dimensional (visual, interactional, structural, cognitive, behavioural)
Integration of indicators	Isolated or loosely combined indicators	Integrated and weighted aggregation of heterogeneous indicators
Use of behavioural and log data	Limited or absent	Explicitly incorporated through learning analytics and interaction traces
Support for comparative interface analysis	Rarely supported	Directly supported
Applicability across platforms	Context-specific and platform-dependent	Platform-independent
Relevance for inclusive learning	Indirect and implicit	High, with explicit focus on cognitive accessibility
Suitability for applied informatics	Moderate	High

The main outcome of the study is the development of a generalisable evaluation framework that systematically links visual, navigational, and interactional interface parameters with an integrated cognitive load index. The analytical application of the framework demonstrates that, even under constant instructional content conditions, variations in interface architecture can substantially influence cognitive load levels, task performance efficiency, and users' cognitive resilience.

The scientific contribution of the study lies in advancing cognitive load assessment from fragmented, indicator-specific measurement practices towards a comprehensive, interface-induced evaluation approach that accounts for the multidimensional nature of user interaction with digital learning environments. By explicitly embedding interface characteristics into the assessment process, the proposed framework extends the methodological scope of research in technology-enhanced learning and provides an analytical basis for evidence-based interface design.

From a practical perspective, the framework offers a scalable tool for preliminary auditing and comparative evaluation of interface solutions in e-learning, blended learning, and distance education systems. It can be integrated into instructional design workflows, UX evaluation practices, and educational analytics processes, particularly in inclusive learning contexts where interface-related cognitive barriers may disproportionately affect learners with heightened sensitivity to cognitive overload. In applied settings, the framework may be directly employed as an interface audit instrument during the design or redesign of digital learning platforms.

At the same time, the proposed framework should be interpreted as an analytically grounded evaluation model supported by an initial pilot application rather than a fully empirically generalised system. Its primary value lies in the structured integration of interface-related cognitive load indicators into a coherent assessment architecture, enabling systematic comparison of alternative interface configurations.

The limitations of the study are associated with the conceptual nature of the model and the exploratory scope of its empirical component. The pilot application serves to demonstrate the feasibility and analytical sensitivity of the framework, rather than to provide statistically generalisable results.

Future research should focus on large-scale empirical validation, including statistically robust testing across diverse user groups, learning platforms, and instructional contexts, as well as the integration of physiological and real-time behavioural data for enhanced model precision.

References

1. Sweller, J.: Cognitive load during problem solving: effects on learning. Cogn. Sci. **12**, 257–285 (1988). https://doi.org/10.1207/s15516709cog1202_4
2. Paas, F., Renkl, A., Sweller, J.: Cognitive load theory and instructional design: recent developments. Educ. Psychol. **38**(1), 1–4 (2003). https://doi.org/10.1207/S15326985EP3801_1
3. Sweller, J., Ayres, P., Kalyuga, S.: Cognitive Load Theory. Springer, New York (2011)
4. Mayer, R.E.: Multimedia Learning, 2nd edn. Cambridge University Press, Cambridge (2009). https://doi.org/10.1017/CBO9780511811678

5. Clark, R.C., Mayer, R.E.: E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning, 4th edn. Wiley, Hoboken (2016). https://doi.org/10.1002/9781119239086
6. Seufert, S., Guggemos, J., Sailer, M.: Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: the current situation and emerging trends. Comput. Hum. Behav. **115**, 106552 (2021). https://doi.org/10.1016/j.chb.2020.106552
7. Forster, S., Lavie, N.: Distracted by your mind? Individual differences in distractibility predict mind wandering. J. Exp. Psychol. Learn. Mem. Cogn. **40**(1), 251–260 (2014). https://doi.org/10.1037/a0034108
8. Martin, F., Sun, T., Westine, C.D.: A systematic review of research on online teaching and learning from 2009–2018. Comput. Educ. **159**, 104009 (2020). https://doi.org/10.1016/j.compedu.2020.104009
9. Diamond, A.: Executive functions. Annu. Rev. Psychol. **64**, 135–168 (2013). https://doi.org/10.1146/annurev-psych-113011-143750
10. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Advances in Psychology, vol. 52, pp. 139–183. Elsevier, Amsterdam (1988). https://doi.org/10.1016/S0166-4115(08)62386-9
11. Petrovič, F., et al.: Measuring the quality of life of university students and its predictors as a basis for applied research: cross-country evidence. Sustain. Futures. **10**, 101214 (2025). https://doi.org/10.1016/j.sftr.2025.101214
12. Zenelaga, B., Goga, A., Králik, R.: Inclusive universities: exploring the well-being of university students with special needs in Albania. J. Educ. Cult. Soc. **15**(2), 373–386 (2024). https://doi.org/10.15503/jecs2024.2.373.386
13. Rao, K., Ok, M.W., Bryant, B.R.: A review of research on universal design educational models. Remedial Spec. Educ. **35**(3), 153–166 (2014). https://doi.org/10.1177/0741932513518980
14. Leka, K., Muho, A., Al-Adwan, A.S., Biryukova, Y.N.: Fostering inclusive virtual learning communities in developing European countries. J. Educ. Cult. Soc. **15**(2), 549–562 (2024). https://doi.org/10.15503/jecs2024.2.549.562
15. Shyshak, A., Chaika, V., Shcherbiak, I., Gažiová, M., Tvrdon, M.: Forming the ability of younger students to behave safely on the internet. J. Educ. Cult. Soc. **15**(1), 333–347 (2024). https://doi.org/10.15503/jecs2024.1.333.347
16. Kirschner, P.A., Sweller, J., Clark, R.E.: Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educ. Psychol. **41**(2), 75–86 (2006). https://doi.org/10.1207/s15326985ep4102_1
17. Longstaffe, K.A., Hood, B.M., Gilchrist, I.D.: The influence of cognitive load on spatial search performance. Atten. Percept. Psychophys. **76**, 49–63 (2014). https://doi.org/10.3758/s13414-013-0575-1
18. Alam, S., Hameed, A., Balogh, Z., Haque, S.A.: Digitalisation of education: challenges and issues in virtual pedagogical competence. J. Educ. Cult. Soc. **16**(1), 165–177 (2025). https://doi.org/10.15503/jecs2025.2.165.177
19. de Jong, T.: Cognitive load theory, educational research, and instructional design: some food for thought. Instr. Sci. **38**, 105–134 (2010). https://doi.org/10.1007/s11251-009-9110-0
20. Kalyuga, S.: Cognitive load theory: how many types of load does it really need? Educ. Psychol. Rev. **23**(1), 1–19 (2011). https://doi.org/10.1007/s10648-010-9150-7
21. Leppink, J., Paas, F., van der Vleuten, C.P.M., van Gog, T., van Merriënboer, J.J.G.: Development of an instrument for measuring different types of cognitive load. Behav. Res. Methods. **45**, 1058–1072 (2013). https://doi.org/10.3758/s13428-013-0334-1
22. Seufert, T.: The interplay between self-regulation in learning and cognitive load. Educ. Res. Rev. **24**, 116–129 (2018). https://doi.org/10.1016/j.edurev.2018.03.004

Modeling Cyber Security in Smart Cities: A Knowledge-Based Approach

Andrea Žváčková, Daniela Ponce, Peter Mikulecký, Pavel Čech(✉), Karel Mls, Tereza Otčenášková, and Petr Tučník

University of Hradec Králové, Rokitanského 62, 500 03 Hradec Králové, Czech Republic
pavel.cech@uhk.cz
https://www.uhk.cz

Abstract. Modern Smart Cities generate vast amounts of heterogeneous data from urban infrastructure, IoT ecosystems, cybersecurity events, and AI analytics. Existing solutions often treat these domains separately, resulting in fragmented knowledge that limits cross-domain reasoning and context-aware cybersecurity decision-making. Current ontologies are domain-specific, and knowledge graphs typically lack semantic integration and AI-supported reasoning.

This paper presents an ontology-driven knowledge graph framework that unifies Smart City, cybersecurity, and AI-related data into a coherent semantic model. Organized into four layers—Data Acquisition, Information Structuring, Semantic Modeling, and Knowledge Reasoning—the framework transforms raw data into actionable, interpretable knowledge. By explicitly modeling entities, relationships, constraints, and dependencies, it supports context-aware reasoning, explainability, interoperability, and modular extensibility.

Remaining challenges include scalable reasoning over large graphs, automated alignment of heterogeneous data, and continuous ontology evolution to accommodate changing urban systems and AI models. Addressing these challenges will strengthen the framework and advance ontology-driven, adaptive, and explainable cybersecurity governance in complex Smart City environments.

Keywords: smart city · cybersecurity · ontology-driven knowledge graph · artificial intelligence · cross-domain integration · explainable reasoning

1 Introduction

Modern Smart Cities generate massive volumes of heterogeneous data spanning urban infrastructure, IoT devices, cybersecurity events, and AI-driven analytical outputs. Existing approaches largely treat these domains in isolation, resulting in fragmented knowledge representations that hinder cross-domain reasoning and context-aware decision-making. Current ontologies and knowledge graphs

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 482–492, 2027.
https://doi.org/10.1007/978-3-032-32104-6_41

either cover single domains or lack integration with AI reasoning mechanisms, limiting their practical utility for real-time cybersecurity management. Moreover, the scale and dynamic nature of urban ecosystems pose significant challenges for efficient reasoning and actionable insight generation. Therefore, there is a critical need for a unified ontology that can semantically integrate multi-domain information. The aim of the paper is to suggest a conceptual framework for an ontology-driven knowledge graph that would support adaptive reasoning, and enable explainable, context-aware cybersecurity analytics for complex Smart City environments.

The paper is organized as follows: Sect. 2 reviews related work; Sect. 3 presents the ontology-driven knowledge graph framework; Sect. 4 provides a discussion of scalability and implications for explainable cybersecurity governance; and Sect. 5 concludes with findings and future directions.

2 Related Work and Background

The growing complexity of Smart City ecosystems has stimulated extensive research into the application of ontological and semantic technologies for cybersecurity management, particularly in conjunction with artificial intelligence (AI) and knowledge graph methodologies.

Ontological modeling represents the most prominent research direction, encompassing approximately 35% of the analyzed literature. Standardized cybersecurity ontologies such as STIX [16], UCO/CASE [6], and the MITRE ATT&CK framework [2] provide widely adopted reference models for expressing threat intelligence, adversary behavior, and digital forensic evidence. These standards offer structured vocabularies and schemas that support semantic interoperability across security tools and organizations.

A significant number of studies extend these foundations into domain-specific contexts. Tefek et al. [17] propose an ontology for Smart Grid environments that automatically integrates CVE vulnerabilities into power-system components. Teng et al. [18] introduce OntoCPS4PMS, enabling collaborative cyber-physical protection for energy monitoring systems. Martinez et al. [12] apply ontology-driven modeling to digital twins of traffic intersections, supporting interoperability and inferencing across multimodal urban data streams. Varvarigou [20] proposes a security ontology for NFV environments, allowing the derivation of logical consequences that link general security requirements with cybersecurity-specific reasoning. Akbar et al. [1] emphasize the significance of temporal reasoning in cybersecurity ontologies, allowing analysts to monitor the progression of attack scenarios. Arogundade et al. [2] combine ontologies with multi-agent architectures and case-based reasoning (CBR) to enable adaptive response strategies. Hutschenreuter et al. [10] use ontologies to support automated decision-making and resilience management during cyber incidents.

Together, these works illustrate how ontology-driven approaches enhance contextual understanding, interoperability, and reasoning capabilities across diverse cybersecurity and Smart City domains. They also highlight key challenges, including domain fragmentation and limited cross-domain alignment.

Knowledge graphs (KGs) constitute the second major research direction, accounting for approximately 28% of the reviewed publications. KGs enable the integration of heterogeneous data sources by providing unified semantic structures that describe urban entities and their relationships. Surveys by Wang et al. [21] and He et al. [9] identify persistent challenges related to data heterogeneity, temporal dynamics, and scalability.

Several studies propose domain-specific knowledge graph architectures. Chen et al. [7] design a cybersecurity KG that integrates multi-source intrusion data and enhances attack detection. Wang et al. [21] introduce UrbanKG, combining IoT, GIS, and BIM datasets to address temporal and semantic inconsistencies in Smart City data. Tao et al. [15] present a KG-based model for multi-source cybersecurity log auditing, enabling joint reasoning over heterogeneous logs. Boyer et al. [4] propose the Unified Cybersecurity Knowledge Graph (UCKG), integrating CVE, CWE, CAPEC, CPE, CVSS, and AAR datasets using the UCO ontology to support incident analysis and automated security operations. Collectively, these studies demonstrate that knowledge graphs are an effective backbone for semantic integration, situational awareness, and real-time reasoning in Smart City cybersecurity. Nonetheless, they also reveal limitations, including the computational complexity associated with large-scale KG construction and challenges in handling dynamic, multimodal data streams.

The third major research trend—accounting for approximately 25% of surveyed publications—concerns the integration of AI and machine learning into cybersecurity systems for Smart Cities. AI methods have been applied to intrusion detection, anomaly detection, threat prediction, and adaptive security decision-making. Martinez et al. [12] emphasize the capacity of ontology-based systems to learn from new data and adapt their reasoning processes in dynamic traffic and autonomous-vehicle environments. Jia et al. [11] present the ACAM framework, which converts cybersecurity information into an MDATA graph and employs subgraph matching to identify attack patterns. Boyer et al. [4] incorporate NLP and LLM-based techniques to extract knowledge from After-Action Reports and integrate them into UCKG, contributing to intelligent cyber-operations. Arogundade et al. [2] propose the CPARR model, combining agent-based systems with CBR for risk prediction and adaptive security configuration.

These contributions highlight the pivotal role of AI in enhancing the accuracy, adaptability, and scalability of cybersecurity systems in urban environments. AI-driven mechanisms help reduce false positives, improve situational understanding, and support continuous learning from evolving cyber-physical environments.

The fourth and the last direction is cross-domain integration and scalability challenges. Despite considerable progress, two significant research gaps are consistently noted across the literature. First, most existing works fail to achieve comprehensive integration across Smart City infrastructure, cybersecurity threat intelligence, and AI-based analytics. Research often addresses only partial intersections—such as Smart Cities × cybersecurity or AI × cybersecurity—without bridging all three domains. Ontologies frequently remain highly domain-

specific (Akbar et al. [1], Arogundade et al. [2]), which limits interoperability and hinders holistic semantic reasoning across heterogeneous Smart City subsystems.

Second, scalability remains a persistent challenge. Many studies validate ontology- and KG-based systems only on small or synthetic datasets (He et al. [9], Chen et al. [7], Jia et al. [11]). Real smart cities, however, contain millions of dynamic entities generating high-velocity data streams. Large-scale reasoning, multimodal data management, and high-performance computation remain difficult to achieve. Wang et al. [21] and Boyer et al. [4] mention the high cost of creating, validating, and maintaining large ontologies, emphasizing the need for hybrid validation combining formal metrics, expert feedback, and LLM-assisted evaluation.

Proposed directions for addressing these gaps include multi-layer semantic architectures linking physical, semantic, and analytical layers (Martinez et al. [12]); LLM-based approaches for dynamic data ingestion and ontology extension (Boyer et al. [4]); agent-based coordination mechanisms for cross-domain situational awareness (Arogundade et al. [2]); and graph-learning techniques for scalable reasoning (Chen et al. [7], Jia et al. [11]).

Several standardized models underpin current cybersecurity ontology development. STIX and TAXII [8] are widely used for representing cyber threat intelligence, while MITRE ATT&CK [14] provides a structured taxonomy of adversarial techniques. UCO and CASE [5] offer comprehensive models to digital forensics and cyber investigations. Tok et al. [19] extend these with the SCOPE profile tailored to Smart City forensics.

Domain-specific ontologies—such as Smart Grid ontologies [17] and IoT-oriented security ontologies [8,13]—address specialized requirements but often lack the broader governance and AI-driven reasoning layers needed for cross-domain interoperability.

To address the identified gaps in current research—namely, the limited cross-domain integration of Smart City infrastructure, cybersecurity, and AI reasoning, as well as challenges in scalability and heterogeneous data management—this work proposes a conceptual ontology-driven framework. Unlike existing models, the framework is designed to unify technical, operational, and AI-driven layers in a single, semantically coherent structure. Its capabilities are defined via competency questions, specifying what it should enable, including cross-domain reasoning, integration with standards such as STIX, UCO/CASE, and MITRE ATT&CK, representation of AI components alongside cyber-physical entities, scalable inference over large and dynamic datasets, and adaptive, explainable decision support. This proposal outlines a blueprint for future research, establishing a clear direction for developing interoperable, context-aware, and scalable Smart City cybersecurity solutions [4,12,21] and [9].

3 Ontology-Driven Knowledge Graph-Based Framework for AI-Enabled Smart City Cybersecurity

The increasing complexity of Smart City ecosystems demonstrates persistent fragmentation between cybersecurity, AI analytics, and urban infrastructure

management. Existing ontologies typically remain domain-specific—focused on isolated sectors such as traffic control, network security, or cyber-physical systems—while knowledge graph (KG) initiatives often aggregate heterogeneous data without providing the semantic coherence required for adaptive, cross-domain reasoning. Building on the gaps identified in the background, this section outlines a unified framework that integrates ontological modeling with KG-based operationalization to support explainable, context-aware cybersecurity across diverse urban environments.

3.1 Conceptual Framework

The framework adopts a modular, multi-layer architecture that reflects the full progression from raw data acquisition to advanced reasoning. At its conceptual core, the ontology defines the classes, semantic relationships, and operational constraints that structure knowledge across Smart City domains, including urban assets, network entities, AI models, and privacy-sensitive processes. This semantic foundation anchors the KG, which instantiates real-world data and enables inference, querying, and scenario-driven analysis.

To maintain interoperability and extensibility, the architecture accommodates established cybersecurity ontologies (e.g., STIX, UCO/CASE, MITRE ATT&CK) and can incorporate emerging domain models such as energy grids or autonomous mobility systems. The framework is organized into four layers:

1. **Data Acquisition**, capturing sensor and network data;
2. **Information Structuring**, harmonizing heterogeneous inputs;
3. **Semantic Modeling**, mapping structured information to ontology constructs; and
4. **Knowledge Reasoning**, supporting alerts, predictions, and cross-domain analytics.

Together, these layers provide a consistent pathway from data to actionable insights.

3.2 Covered Domains and Scenarios

To anchor the ontology in operational requirements, six representative domains were selected: traffic management, network and NFV security, AI-driven threat prediction, resilience management, privacy and compliance, and cross-domain coordination (see illustrative ontology fragments at Figs. 1 and 2).

Each domain is illustrated through scenarios that demonstrate how semantically integrated knowledge enables context-aware decision-making. For example, traffic incident response analyses sensor outputs alongside cybersecurity alerts; network threat monitoring correlates NFV, cloud, and IoT events; and AI-based risk prediction links model outputs with asset dependencies and mitigation options. These scenarios collectively define the functional space that the ontology and KG must support, while ensuring relevance to real Smart City operations.

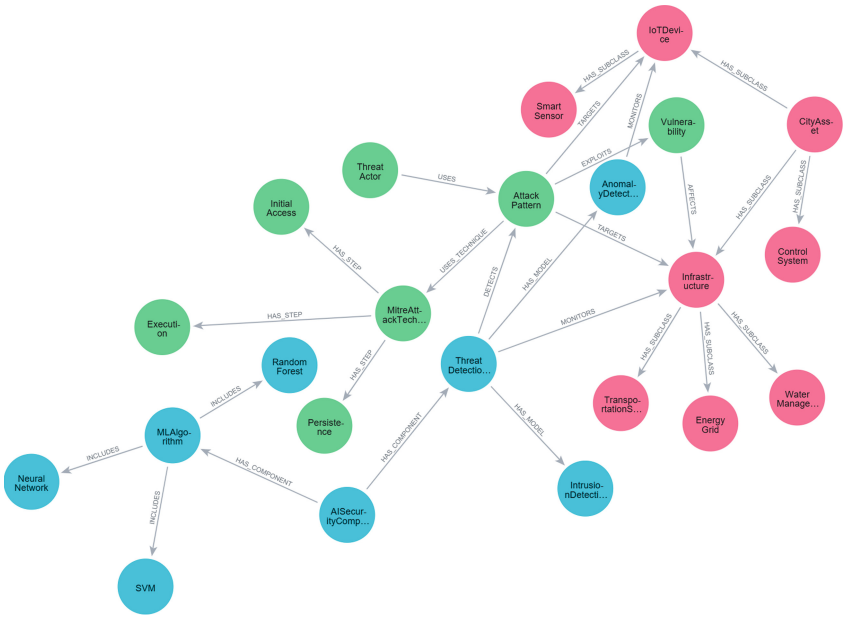


Fig. 1. Illustrative ontology fragment: High-level classes.

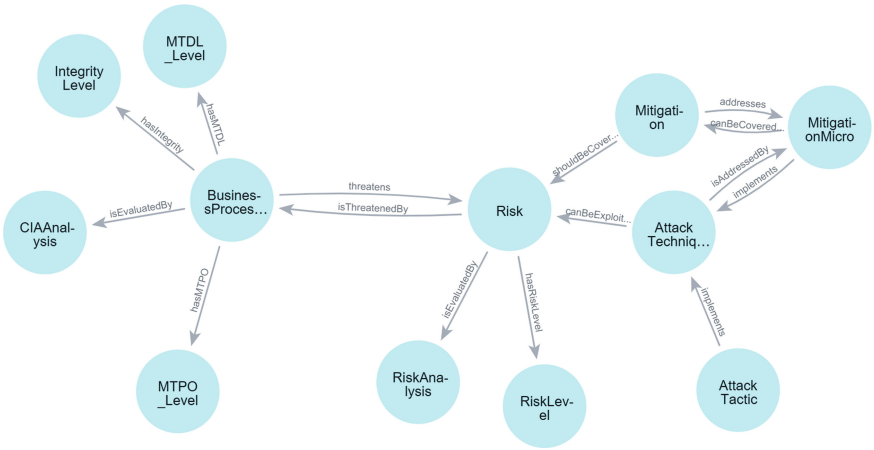


Fig. 2. Illustrative ontology fragment: Risk related concepts.

3.3 Competency Questions

The ontology is guided by 13 competency questions (see Table 1 that formalize the reasoning capabilities required across the six domains. These questions define what the ontology must be able to represent—ranging from subsystem dependencies and explainability of AI models to cross-domain correlations between phys-

ical events and cyber-attacks—and how such knowledge should be retrievable through the KG. SPARQL-style pseudocode illustrates possible query patterns, ensuring that each competency question is operationalizable once the ontology is instantiated with real data. Together, the questions guarantee that the ontology remains aligned with practical cybersecurity needs rather than abstract taxonomies.

Table 1. Competency Questions 1–13

No.	Competency Question	Associated Scenario/Data	Example SPARQL Query (pseudo-syntax)
1	Which sensors report anomalous behavior in critical intersections?	Traffic management, IoT sensor logs	SELECT ?sensor WHERE ?sensor rdf:type TrafficSensor; ?sensor reports Anomaly
2	How are detected cyber incidents related to physical assets?	Smart grid, power substations	SELECT ?incident ?asset WHERE ?incident affects ?asset
3	Which AI models are currently deployed for threat prediction?	Cybersecurity analytics	SELECT ?model WHERE ?model rdf:type AIModel; ?model purpose ThreatPrediction
4	Which subsystems remain operational after a cyber-attack?	Resilience and Recovery System health logs, event timeline	SELECT ?subsystem WHERE ?subsystem status “operational" FILTER EXISTS ?event occurredAt ?time
5	Which data sources contain personally identifiable information (PII)?	Privacy-Aware Operations IoT sensor metadata, access logs	SELECT ?source WHERE ?source contains ?data; dataType “PII"
6	Which security controls are active for a given asset?	Traffic Incident Response, Network Threat Security policy database, control logs	SELECT ?control WHERE ?control protects ?asset; status “active"
7	How are urban assets linked to network components?	Cross-Domain Coordination Asset registry, network mapping	SELECT ?asset ?component WHERE ?asset connectedTo ?component
8	What is the temporal sequence of detected security events?	Network Threat Monitoring Event logs, timestamped alerts	SELECT ?event ?time WHERE ?event occurredAt ?time ORDER BY ?time
9	Which AI predictions influenced decision-making in urban subsystems?	AI-Driven Risk Prediction Decision logs, AI outputs	SELECT ?decision ?model WHERE ?decision influencedBy ?model
10	Which vulnerabilities are associated with network components?	Network Threat Monitoring CVE/CWE mapping, network inventory	SELECT ?vuln ?component WHERE ?vuln affects ?component
11	Which actors are responsible for security incidents?	All scenarios User logs, access control	SELECT ?actor ?event WHERE ?actor caused ?event
12	How are resilience strategies activated in response to detected threats?	Resilience and Recovery Policy triggers, system events	SELECT ?strategy ?threat WHERE ?strategy activatedBy ?threat
13	Which data sources are integrated into the knowledge graph?	Cross-Domain Coordination All sensor logs, database entries	SELECT ?source WHERE ?source integratedIn KG

3.4 Functional Requirements of the Knowledge Graph

The knowledge graph translates the conceptual ontology into an operational environment capable of integrating, interpreting, and analyzing heterogeneous Smart City data. Its functional requirements span four tightly interconnected areas.

First, **data integration** demands the harmonization of diverse information sources—including traffic sensors, NFV logs, AI predictions, and compliance records—while preserving provenance, temporal ordering, and contextual metadata. Such integration ensures that raw observations can be semantically linked to urban assets, threat indicators, or AI-derived insights.

Second, the KG must support advanced **reasoning and analytics**, enabling correlations between physical and cyber events, automated alert generation, and historical analyses that reveal long-term patterns or emerging vulnerabilities. These reasoning capabilities provide the bridge between domain-specific observations and the cross-domain inferences required for holistic cybersecurity governance.

Third, **operational flexibility** is essential, as Smart Cities evolve continuously. The KG therefore supports dynamic schema extensions, integration of new ontologies, and interoperability with AI systems that generate or consume semantically structured knowledge. This adaptability ensures that the framework can incorporate emerging technologies and scenarios without requiring disruptive redesign.

Finally, **performance and scalability** are fundamental constraints. The KG must handle large, dynamically updated graphs while supporting near-real-time queries and analytics. Techniques such as incremental reasoning, modular graph design, and embedding-based representations contribute to maintaining responsiveness in large-scale deployments.

3.5 Risks and Limitations

Despite the framework's integrative design, several structural limitations remain. The heterogeneity and uneven maturity of Smart City data can constrain the precision and completeness of knowledge graph population; large-scale, cross-domain reasoning risks prohibitive computational overhead; and the rapid evolution of IoT ecosystems, threat intelligence, and AI models introduces continual semantic drift that no static ontology can accommodate. These constraints highlight the need for mechanisms that preserve conceptual clarity while sustaining operational performance in dynamic environments.

These considerations set the stage for the next section, which examines the core challenges inherent in deploying the framework at scale. Section 4 discusses how modular ontology architectures, incremental KG construction, hybrid symbolic–statistical reasoning, and AI-assisted ontology evolution can mitigate these pressures and support long-term robustness, scalability, and adaptability across Smart City, cybersecurity, and AI-analytics domains.

4 Discussion

The proposed ontology-driven knowledge graph framework unifies heterogeneous domains—urban infrastructure, AI analytics, and cybersecurity—into a single semantic model, addressing the fragmentation identified in Sect. 2. By explicitly encoding operational dependencies among assets, analytical components, and threats, it enables context-aware, explainable reasoning, allowing stakeholders to trace risk propagation across both physical and cyber-analytical layers. This operationalization distinguishes it from existing Smart City ontologies, which often focus on descriptive taxonomies rather than active inference [3,12].

Scalability remains a core concern. Smart City environments involve millions of interconnected entities, challenging reasoning and query performance. The framework mitigates these issues through modular ontology design, hierarchical structuring, selective inference materialization, and hybrid reasoning combining symbolic and vector-based approaches, ensuring efficient and context-sensitive analysis in dynamic settings.

Ontology evolution is also critical. AI models, IoT configurations, and urban infrastructures continuously change, introducing conceptual drift. Incremental updates, automated alignment mechanisms, and modularity allow the ontology to adapt while maintaining semantic consistency, guided by competency questions and scenario-driven coverage defined in Sect. 3.

Despite these advances, practical limitations remain. Empirical validation with real-world data is required to assess scalability, completeness, and integration. Additionally, domains lacking established ontologies remain underrepresented, highlighting opportunities for future expansion.

In summary, the framework delivers a cross-domain, ontology-driven foundation for explainable and adaptive cybersecurity governance in Smart Cities. It integrates multiple domains, supports scalable inference, and provides traceable decision-making, setting the stage for the following section on risk mitigation, operational considerations, and practical implications

5 Conclusion

Modern Smart Cities generate vast volumes of heterogeneous data across urban infrastructure, IoT ecosystems, cybersecurity intelligence, and AI analytics, yet current solutions continue to treat these domains in isolation. As shown in prior work, existing ontologies are typically domain-specific, loosely integrated, or disconnected from AI-enabled reasoning, which limits cross-domain situational awareness and constrains explainable cybersecurity governance in complex urban environments. The ontology-driven knowledge graph framework proposed in this work addresses these gaps by introducing a unified semantic layer that integrates Smart City, cybersecurity, and AI knowledge into a coherent and machine-interpretable representation.

Structured into four transformation layers—from Data Acquisition and Information Structuring to Semantic Modeling and Knowledge Reasoning—the

framework operationalizes explicit semantics, modular modeling practices, and cross-domain interoperability. In doing so, it supports context-aware analytics, inferential traceability, and integration with established standards such as STIX, UCO/CASE, and MITRE ATT&CK. The resulting architecture enables reasoning across operational dependencies that traditionally remain siloed, while simultaneously providing a foundation for handling competence-driven use cases introduced earlier in the article.

Several challenges remain. Scalable reasoning over millions of interconnected entities requires more efficient hybrid inference combining symbolic methods with vector-based approximations. Automated and AI-assisted ingestion pipelines are needed to populate the knowledge graph from diverse, partially structured sources. Maintaining ontological coherence in the face of evolving AI models and urban technologies demands incremental and semi-automated mechanisms for ontology evolution. Finally, empirical validation on real Smart City infrastructures is essential for assessing performance, practical applicability, and robustness under realistic risk scenarios.

Addressing these challenges will further strengthen the proposed framework and advance ontology-driven, cross-domain reasoning for explainable cybersecurity analytics in next-generation Smart Cities.

Acknowledgment. The financial support of the project “Application of Artificial Intelligence for Ensuring Cyber Security in Smart City", n. VJ02010016, granted by the Ministry of the Interior of the Czech Republic, is gratefully acknowledged.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Akbar, K.A., Halim, S.M., Singhal, A., Abdeen, B., Khan, L., Thuraisingham, B.: The design of an ontology for ATT&CK and its application to cybersecurity. In: Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy, pp. 295–297 (2023)
2. Arogundade, O.T., Abayomi-Alli, A., Misra, S.: An ontology-based security risk management model for information systems. Arab. J. Sci. Eng. **45**(8), 6183–6198 (2020)
3. Bernardini, A., D’Alterio, F., Sagratella, L., Settembre, M., Maunero, N.: Structuring the complexity: an ontological approach to analyze the cybersecurity of a 5G service. In: 2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 691–700. IEEE (2024)
4. Boyer, A., Dogdu, E., Choupani, R., Watson, J.S., Sanchez, D., Ametu, A.: Toward a unified cybersecurity knowledge graph: leveraging ontologies and open data sources. In: Southwest Data Science Conference, pp. 17–33. Springer, Heidelberg (2024). https://doi.org/10.1007/978-3-031-67871-4_2
5. Casey, E., Katz, G., Lewthwaite, J.: Harnessing automated malware analysis for computer forensics. In: Digital Investigation, pp. 78–89 (2018)
6. Casey, E.: Handbook of digital forensics and investigation (2009)

7. Chen, Y., et al.: Construction of network security domain knowledge graph for network attack detection. In: Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering, pp. 1171–1178 (2022)
8. Daoudagh, S., et al.: An ontology-based solution for monitoring IoT cybersecurity. In: IFIP International Internet of Things Conference, pp. 158–176. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-18872-5_10
9. He, X., Kuai, X., Li, X., Qiu, Z., He, B., Guo, R.: Smart city ontology framework for urban data integration and application. Smart Cities **8**(5), 165 (2025)
10. Hutschenreuter, H., Çakmakçi, S.D., Maeder, C., Kemmerich, T.: Ontology based cybersecurity and resilience framework. In: ICISSP, pp. 458–466 (2021)
11. Jia, Y., et al.: Artificial intelligence enabled cyber security defense for smart cities: a novel attack detection framework based on the MDATA model. Knowl.-Based Syst. **276** (2023). https://doi.org/10.1016/j.knosys.2023.110781
12. Martinez, V.M.G., Mussi, S.S., Ribeiro, M.R.N., Campelo, D.R., Carvalho, T.C.: VRU safety at road intersections: an ontology-driven digital twin. IEEE Access **13**, 138166–138184 (2025). https://doi.org/10.1109/ACCESS.2025.3595998
13. Qaswar, F., et al.: Applications of ontology in the internet of things: a systematic analysis. Electronics **12**(1), 111 (2022)
14. Strom, B.E., Applebaum, A., Miller, D.P., Nickels, K.C., Pennington, A.G., Thomas, C.B.: MITRE ATT&CK: design and philosophy. In: Technical report. The MITRE Corporation (2018)
15. Tao, Y., Li, M., Hu, W.: Research on knowledge graph model for cybersecurity logs based on ontology and classified protection. In: Journal of Physics: Conference Series, vol. 1575, p. 012018. IOP Publishing (2020)
16. TC, O.C.: STIX version 2.1. Technical report. OASIS Committee Specification (2021)
17. Tefek, U., Esiner, E., Cheh, C., Mashima, D.: A smart grid ontology: vulnerabilities, attacks, and security policies. In: 2023 IEEE Conference on Communications and Network Security (CNS), pp. 1–6. IEEE (2023)
18. Teng, J., Yang, R., Wang, T., Du, J., Sheng, Q.Z.: OntoCPS4PMS: ontology modeling for collaborative cyber-physical threat defense in power monitoring system. Syst. Eng. **28**(1), 29–44 (2025)
19. Tok, Y.C., Zheng, D.Y., Chattopadhyay, S.: A smart city infrastructure ontology for threats, cybercrime, and digital forensic investigation. Forensic Sci. Int.-Digital Invest. **52** (2025). https://doi.org/10.1016/j.fsidi.2025.301883
20. Varvarigou, D., Espes, D., Bersano, G.: Ontology-based solution for handling safety and cybersecurity interdependency in NFV safety architecture. Procedia Comput. Sci. **220**, 527–534 (2023)
21. Wang, Z., Han, F., Zhao, S.: A survey on knowledge graph related research in smart city domain. ACM Trans. Knowl. Disc. Data **18**(9), 1–31 (2024)

Smart Biosignal Analysis: IoT, Sensors and Machine Learning Applications

Security Evaluation of Communication Protocols in Wireless Sensor Networks for Smart Building Environments

Zoltán Balogh[1,2](✉), Zsolt Čonka[2,3], Martin Magdin[1,4], Norbert Ádám[5], and Shkëlqim Fortuzi[6]

[1] Department of Informatics, Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia
zbalogh@ukf.sk

[2] Kándó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest, Hungary

[3] Department of Electric Power Engineering, Technical University of Košice, Košice, Slovakia

[4] Faculty of Economics, University of South Bohemia in České Budějovice, České Budějovice, Czech Republic

[5] Faculty of Electrical Engineering and Informatics, Technical University of Košice, Košice, Slovakia

[6] Department of Finance-Accounting, Faculty of Business, Aleksandër Moisiu University, Durrës, Albania

Abstract. Wireless sensor networks (WSN) play a crucial role in data collection and transmission within intelligent systems, with communication security being one of their most significant challenges. The growing number of Internet of Things (IoT) devices and their limited computational resources increase the demand for both efficient and secure communication solutions. This paper presents the design, implementation, and security verification of a wireless sensor network intended for a smart building environment. We analyze current WSN and IoT technologies, evaluate communication protocols and their security mechanisms based on professional studies, and implement selected protocols (MQTT, CoAP, HTTPS, and WebSocket) with recommended security measures to ensure confidentiality, integrity, and availability of transmitted data. The functionality and resilience of the solution were verified through comprehensive penetration tests and simulated attacks including man-in-the-middle, denial of service, brute-force, and deauthentication attacks. Based on security testing outcomes and literature-supported performance characteristics, MQTT with TLS is identified as the most suitable candidate for this class of deployment. Quantitative performance benchmarking (energy consumption, latency, memory footprint) remains as future work.

Keywords: Wireless Sensor Networks · IoT Security · Communication Protocols · Penetration Testing · MQTT · Smart Buildings · TLS Encryption

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 495–507, 2027.
https://doi.org/10.1007/978-3-032-32104-6_42

1 Introduction

The Internet of Things (IoT) represents one of the most significant technological innovations of recent years. IoT devices combine sensing capabilities, computing power, and internet connectivity, enabling them to communicate with each other and transmit data to centralized systems for monitoring and management purposes. This technological paradigm has transformed numerous industries, from healthcare and manufacturing to smart cities and agriculture [1, 2]. In the context of Industry 5.0, where human-centric and resilient smart systems are prioritized, secure IoT communication becomes a foundational requirement [41].

Sensor networks, particularly wireless sensor networks (WSN), are a key element in IoT due to their flexibility, simple installation, and low energy consumption. With the increasing number of devices, security risks also grow, especially in the area of transmitted data protection. Communication protocols play a fundamental role in their security, enabling efficient and secure data exchange between sensors and central nodes [3, 4]. Without properly implemented security measures, data and devices on the network may be exposed to vulnerabilities that threaten the confidentiality, integrity, and availability of information [5].

Given the limited computing power and low energy capacity of sensor devices, it is essential that communication protocols contain security measures tailored to the specific needs of these networks. The selection of appropriate protocols significantly affects not only security but also energy efficiency, transmission reliability, and overall system performance [6]. This balance between security and resource constraints presents one of the key challenges in WSN design and implementation.

This paper focuses on the analysis of security aspects of communication protocols used in sensor networks within IoT environments. The main contributions of this work include: (1) comprehensive analysis of security mechanisms for MQTT, CoAP, HTTP/HTTPS, and WebSocket protocols based on recent professional studies; (2) design and implementation of a secure WSN for smart building environments; (3) practical verification of security measures through penetration testing; and (4) comparative evaluation and selection of the optimal protocol for resource-constrained IoT applications.

2 Related Work and Theoretical Background

2.1 Wireless Sensor Networks in IoT

WSN are a crucial component of IoT, enabling automated collection, transmission, and processing of data without the need for direct human interaction. Their deployment across various sectors brings efficient information exchange between interconnected systems and devices [7]. When designing and implementing sensor networks, it is important to consider not only the sensor nodes themselves and their technical requirements, but also security aspects that are critical for protecting sensitive data, ensuring reliable operation, and preventing unauthorized access [8, 9].

Sensor nodes represent the fundamental element of WSN and play a crucial role in a wide spectrum of IoT applications. Their primary function is the collection and transmission of environmental data using various types of sensors including light, magnetic,

infrared, optical, and ultrasonic sensors. Successful implementation of WSN depends on the effective design of sensor nodes, with key factors including detection accuracy, energy efficiency, and error data processing capabilities [10].

The deployment areas for WSN have gained significant attention due to their wide potential for use across various domains. The most important application areas include environmental monitoring, healthcare, agriculture, industrial automation, smart cities, and military surveillance. This diversity of applications emphasizes the fundamental role of WSN in improving process efficiency, strengthening security, and enhancing quality of life across different sectors [11].

2.2 Security Requirements and Threats

Security of WSN is an important aspect of their deployment, especially in sensitive environments where any breach can lead to serious consequences. These networks process sensitive data and operate in dynamic conditions, making them potential targets for various attacks. To protect the network against security threats, it is essential to meet the following requirements: availability, authorization, authentication, confidentiality, and integrity [12, 13].

Threats in WSN can be divided into passive (eavesdropping and communication monitoring) and active (data manipulation and network disruption). The most common types of attacks in sensor networks include: communication eavesdropping leading to data confidentiality compromise; DoS attacks aimed at disrupting network availability; data manipulation where attackers can compromise information integrity by injecting false data into data streams; Sybil attacks leading to unauthorized node authentication; jamming attacks causing wireless signal interference; routing manipulation where attackers change packet routing in the network; and physical attacks on sensor devices [14, 15].

Effective mitigation of these threats increasingly relies on automated detection mechanisms, including machine learning-based intrusion detection systems capable of identifying anomalous traffic patterns. Such approaches draw on advances in natural language processing and pattern recognition techniques that have demonstrated strong classification performance across diverse analytical domains [42].

2.3 Communication Models and Protocols

The TCP/IP model represents a conceptual framework for network communication that defines principles of data transmission between devices in computer networks and finds wide application in WSN context. Its architecture and protocols are adapted to the requirements of modern communication systems, enabling simple integration of WSN into existing infrastructure [16]. In the context of WSN, the TCP/IP model brings several advantages including wide implementation support, flexibility for devices with limited resources, and the ability to integrate security protocols [17].

The selection of appropriate communication protocols significantly affects the security, efficiency, and reliability of WSN. The most commonly used application layer protocols in IoT environments include MQTT, CoAP, HTTP/HTTPS, and WebSocket.

Each protocol has specific characteristics that make it suitable for different application scenarios [18, 19].

2.4 Analysis of Application Protocol Security

MQTT Protocol (Message Queuing Telemetry Transport) is an efficient and economical protocol designed for IoT devices and sensor networks. It is characterized by low system resource requirements and operates on the publish/subscribe principle, enabling asynchronous message exchange between devices with limited computational and energy capacities. The protocol uses TCP for reliable data transmission and supports three QoS levels [20]. Studies by Laghari et al. [21] systematically examined MQTT security risks and identified key areas including data encryption (TLS/SSL, AES/ECC), authentication and access control (ACL, OAuth 2.0), and attack detection using IDS and machine learning. Bangare and Patil [22] proposed solutions for eliminating security challenges through integration of Merkle trees and HBMQTT broker, achieving higher resilience against cyber threats.

CoAP Protocol (Constrained Application Protocol) is an economical and efficient protocol designed for IoT networks and devices with limited resources. It uses UDP for fast and efficient communication, minimizing network load and energy consumption. The protocol is defined in RFC 7252 and is based on REST principles. CoAP integrates DTLS, a version of TLS adapted for UDP communication, which ensures confidentiality, integrity, and authentication of transmitted data [23]. El-Hajj and Biersteker [24] examined CoAP security and performance on Raspberry Pi 4, concluding that while CoAP in its basic form is vulnerable to eavesdropping and data manipulation, DTLS and OSCORE significantly improve security. Arvind and Narayanan [25] analyzed security risks and proposed a client-server architecture with a proxy system for intercepting and analyzing data to test vulnerabilities.

HTTP/HTTPS Protocol is the basic communication protocol used for data transfer on the internet. It operates on the client-server principle and is used for transferring web pages, images, files, and other data. However, HTTP is an unsecured protocol, meaning transmitted data is not encrypted and can be intercepted or manipulated. Therefore, modern systems prefer HTTPS, which uses SSL/TLS encryption to protect transmitted data from eavesdropping and unauthorized manipulation [26]. Hermawan et al. [27] and Insan and Samopa [28] analysed HTTP security implementation for IoT, emphasizing the importance of TLS encryption, certificate validation, and API key authentication.

WebSocket Protocol is a protocol enabling bidirectional real-time communication between client and server over a single, long-lived connection. It is designed to be more efficient than traditional protocols like HTTP. The protocol itself does not contain built-in security mechanisms, but can be secured through TLS authentication (WSS protocol) [29]. Nadeem et al. [30] and Panday and Kumar [31] identified main security risks including CSWSH, XSS attacks, and DoS attacks, and proposed a set of measures including JWT authentication, wss:// protocol usage, Origin header validation, input sanitization, and rate limiting

3 System Design and Implementation

3.1 Architecture Overview

The proposed solution was built on a client-server architecture optimized for smart building environments with consideration of limited computational and energy resources. The client node consisted of an ESP32 D1 R32 microcontroller, chosen for its low power consumption, integrated Wi-Fi module, sufficient computing power, and ability to implement secure data transmission. Two sensors were connected to the microcontroller: DHT22 for temperature and humidity measurement, and VEML7700 for light intensity measurement [32, 39].

The prepared data was transmitted via Wi-Fi from the microcontroller to the central node Raspberry Pi 4, which served as the server responsible for receiving, storing, and further processing the data. Figure 1 shows the WSN hardware connection diagram. The system included data visualization through the Grafana platform and storage in a local SQLite database. All hardware components were assembled on a breadboard using jumper cables and necessary passive elements, with power supplied by a 9 V battery for autonomous operation.

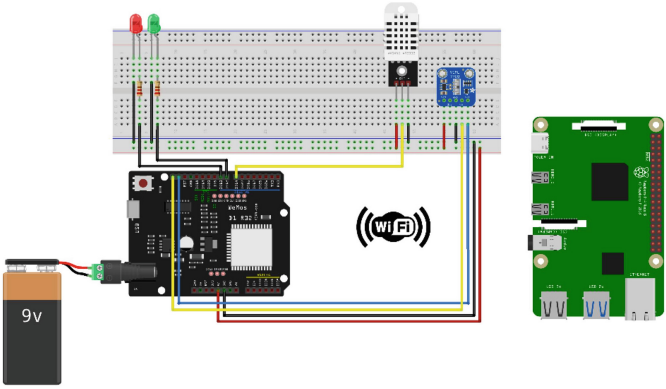


Fig. 1. WSN hardware connection diagram

3.2 Requirements Specification

For proper network functioning, it was essential to define functional and non-functional requirements that determined its properties and ensured reliability and operational security. Functional requirements included: collection of physical quantities (temperature, humidity, light intensity), data transmission between sensors and central node, and data storage in database with real-time visualization. Non-functional requirements included: communication security, energy efficiency, system reliability, scalability, interoperability, system availability, and maintenance and updates capability. Table 1 presents a comparison of TCP and UDP transport protocols.

Table 1. Comparison of TCP and UDP transport protocols.

Property	TCP	UDP
Reliability	Guaranteed delivery	No delivery guarantee
Speed	Slower (flow control)	Faster (no control)
Header size	20–60 bytes	8 bytes
Energy consumption	High	Low

3.3 Security Measures Implementation

Security measures were designed to ensure the three fundamental requirements of information security: confidentiality, integrity, and availability of transmitted data. For each implemented protocol, specific security mechanisms were applied based on the analysis of professional studies and considering the computational constraints of the devices used.

For MQTT protocol, Eclipse Mosquitto broker was deployed with TLS 1.2 encryption using ECC (Elliptic Curve Cryptography) and AES-128, CA certificate validation, username/password authentication, and Access Control Lists (ACL). Additional protection was provided through iptables firewall rules for rate limiting and Fail2Ban for brute-force attack prevention [33].

CoAP implementation utilized the coap-server-openssl tool from libcoap3 library with DTLS (Datagram Transport Layer Security) using Pre-Shared Keys (PSK) and AES-128 encryption. The iptables firewall was configured for access control.

For HTTPS, a Flask-based server was implemented with TLS 1.2 encryption, certificate-based authentication, and API key validation. Flask-Limiter was used for rate limiting to prevent DoS attacks, complemented by iptables firewall rules for SYN flood protection.

WebSocket server was built using the Python websockets library with TLS 1.2 encryption, JWT (JSON Web Token) authentication, custom rate limiting mechanisms limiting to 12 messages per minute per IP address, and IP-based blocking capabilities for repeated authentication failures. Table 2 presents a summary of security mechanisms for the implemented protocols.

Table 2. Summary of security mechanisms for implemented protocols

Protocol	Encryption	Authentication	Access Control
MQTT	TLS 1.2 (ECC, AES-128)	CA cert, user/pass, ACL	iptables, Fail2Ban
CoAP	DTLS (AES-128)	Pre-Shared Key (PSK)	iptables
HTTPS	TLS 1.2 (ECC, AES-128)	CA cert, API key	Flask-Limiter, iptables
WebSocket	TLS 1.2 (ECC, AES-128)	CA cert, JWT token	Rate limiting, iptables, IP blocking

4 Security Testing Methodology

The threat model assumed an attacker with local network access but without physical access to the hardware devices. The primary protected assets were the confidentiality of sensor data and the availability of the communication system. Physical hardware attacks, supply-chain threats, and attacks requiring prior compromise of the Raspberry Pi server are explicitly out of scope of this evaluation.

4.1 Testing Environment and Tools

The security verification was conducted through a comprehensive penetration testing approach using Kali Linux operating system, which provides a wide spectrum of tools designed for analysis and penetration testing of wireless networks and communication protocols. For passive traffic interception between ESP32 and Raspberry Pi 4, monitoring mode was activated on an external Wi-Fi adapter (TP-Link TL-WN722N). Testing was also extended to include manual verification directly on the Raspberry Pi 4 device [34, 40].

The testing toolkit included: Nmap for network scanning and port discovery; Airodump-ng for wireless network reconnaissance; Aireplay-ng for deauthentication attacks; Wireshark for traffic analysis and protocol inspection; hping3 for SYN flood attack simulation; nping for TCP traffic generation; Nikto for web server vulnerability scanning; and mosquitto_pub/mosquitto_sub for MQTT-specific testing.

4.2 Attack Simulation Scenarios

The penetration testing covered multiple attack vectors to comprehensively evaluate the security posture of the implemented system. The following attack scenarios were executed:

Man-in-the-Middle (MITM) Attacks. Using Wireshark, we analyzed network traffic to determine if sensitive data could be intercepted. Testing was performed in two phases: first without knowledge of Wi-Fi password (to verify WPA2-PSK encryption), and second after obtaining the 4-way handshake through deauthentication attack to verify TLS encryption of application data.

Denial of Service (DoS) Attacks. Rate limiting effectiveness was tested using nping to generate 10 TCP requests per second with 50 total packets on secured ports. SYN flood attacks were simulated using hping3 to generate massive amounts of TCP SYN requests aimed at exhausting system resources with incomplete connections in SYN_RECV state.

Brute-Force Authentication Attacks. Manual testing was performed using protocol-specific tools to verify that the system properly rejected connection attempts without valid credentials and that multiple failed attempts triggered protective mechanisms (Fail2Ban, IP blocking).

Wireless Layer Attacks. Deauthentication attacks were performed using aireplay-ng tool to test the resilience of ESP32 device against forced disconnection from the access point. This attack exploits the lack of protection for management frames in WPA2 networks without PMF (Protected Management Frames) support.

4.3 Testing Procedures for Individual Protocols

MQTT Protocol Testing. Port scanning with Nmap confirmed that port 8883 (TLS-secured MQTT) was open while port 1883 (unsecured) was filtered by firewall. Wireshark analysis after WPA2 decryption revealed TLS 1.2 handshake with cipher suite TLS_ECDHE_ECDSA_WITH_AES_128_CBC_SHA256, confirming proper encryption. Application data within MQTT packets remained encrypted and unreadable. Rate limiting test showed proper packet filtering, and SYN flood test blocked over 426,000 malicious packets while accepting only 113 legitimate ones. Manual testing confirmed that connections without CA certificate or valid credentials were properly rejected [35].

HTTP Protocol Testing (Baseline Comparison). As a demonstration of risks from unencrypted communication, we tested unsecured HTTP on port 80. Nmap revealed a Flask/Werkzeug server. After deauthentication and handshake capture, Wireshark successfully extracted readable POST requests containing sensor data in JSON format. Nikto scanner identified missing security headers. Using curl, we successfully injected false data into the database, demonstrating the critical need for encryption and authentication.

HTTPS Protocol Testing. Testing verified that TLS encryption prevented traffic interception, API key validation rejected unauthorized requests, and Flask-Limiter properly throttled excessive requests. SYN flood protection via iptables was verified through packet filtering statistics.

WebSocket Protocol Testing. Testing confirmed WSS (WebSocket Secure) encryption, JWT token validation with IP blocking after repeated failures, and custom rate limiting restricting to 12 messages per minute per IP address.

5 Results and Discussion

5.1 Implementation Results

The designed solution was successfully implemented, enabling collection of physical quantities from sensors (temperature, humidity, light intensity), their transmission in JSON format, secure storage in local SQLite database, and data visualization through the Grafana platform. The implementation of communication protocols proceeded as planned, with three protocols (MQTT, HTTPS, WebSocket) fully operational with complete security measures.

The CoAP protocol implementation with DTLS encountered technical difficulties on the ESP32 platform. The libcoap3 library with the coap-server-openssl backend was the primary implementation attempted. The DTLS handshake consistently failed at the key derivation and certificate exchange stage; heap monitoring on the ESP32 indicated exhaustion of available DRAM (approximately 40 kB free at runtime), which is insufficient for the full DTLS 1.2 handshake with certificate-based authentication. A fallback to Pre-Shared Key (PSK) mode was also attempted with the same result. This failure constitutes a reproducible negative result with practical value for IoT practitioners considering CoAP/DTLS on similar resource-constrained microcontrollers. Future work should investigate lightweight DTLS implementations such as mbed TLS with reduced cipher suite configuration or OSCORE as an application-layer alternative [36].

5.2 Security Testing Results

Based on the conducted testing, all three successfully implemented protocols (MQTT, HTTPS, and WebSocket) reliably met the defined security requirements for confidentiality, integrity, and availability of transmitted data after the deployment of security measures. Summary of penetration testing results is on Table 3. The penetration tests did not confirm the existence of critical vulnerabilities that would lead to compromised communication security. For the brute-force attack scenarios, Fail2Ban was triggered after 5 consecutive failed authentication attempts within a 60-s window, resulting in a 10-min IP ban—consistent with the configured thresholds. For the rate limiting tests, nping generated 10 TCP requests per second (50 packets total); all packets exceeding the configured threshold of 5 connections per second were dropped by iptables. The SYN flood test using hping3 generated approximately 430,000 SYN packets over 60 s, of which 426,082 (99.97%) were blocked and only 113 reached the application layer.

Table 3. Summary of penetration testing results

Attack Type	MQTT	HTTPS	WebSocket
Traffic Sniffing	Protected (TLS)	Protected (TLS)	Protected (TLS)
Unauthorized Access	Blocked (ACL)	Blocked (API key)	Blocked (JWT)
Rate Limiting Test	Effective	Effective	Effective
SYN Flood Attack	99.97% blocked	Protected	Protected
Deauthentication	N/A (platform-level)*	N/A (platform-level)*	N/A (platform-level)*

*Deauthentication vulnerability is a property of the ESP32 Wi-Fi implementation and WPA2 without PMF support—it is independent of the application-layer protocol (MQTT, HTTPS, WebSocket). See Sect. 5.4 for platform-level mitigation.

The exception was the ESP32 device itself, which proved potentially vulnerable to deauthentication attacks. Using the aireplay-ng tool, we successfully disconnected the device from the access point, temporarily interrupting communication with the server. This type of attack demonstrated the possibility of disrupting service availability in the absence of advanced protection mechanisms at the wireless communication layer. However, after the attack ended, the device automatically reconnected to both the network and the server, thanks to the implemented reconnection functionality in the client software [37].

It is important to note that deauthentication vulnerability is not a characteristic of the MQTT, HTTPS, or WebSocket protocols themselves, but a limitation of the ESP32 Wi-Fi hardware and WPA2 without PMF (Protected Management Frames) support. As a platform-level mitigation, the use of access points with PMF enabled is recommended, which protects management frames and significantly reduces the risk of forced client disconnection.

5.3 Protocol Selection and Evaluation

When selecting the most suitable protocol for the given application, it was necessary to consider not only security aspects but also performance parameters, specifically energy

efficiency, transmission reliability, and latency. Given that the solution was optimized for deployment in a smart building environment using battery-powered devices with limited computational resources, emphasis was placed on low energy consumption and simple connection management.

Based on the security testing outcomes and literature-supported performance characteristics [26, 38], MQTT with TLS is identified as the most suitable protocol for this class of deployment. It should be noted that direct quantitative benchmarking of energy consumption, memory footprint, and transmission latency on the ESP32 hardware was not performed in this study and remains as future work. Its advantages lie primarily in low overhead, efficient transmission of small data volumes, simple integration, and high compatibility with resource-constrained devices. The publish-subscribe architecture also enables easy system expansion with new sensors or processing units without modifications to existing connections [38].

From a security perspective, the MQTT protocol after integration of TLS and other protection mechanisms demonstrated full resilience against tested attacks, including unauthorized access, sniffing, and service flooding. Although other protocols met security expectations, their practical application in resource-constrained environments was less suitable: HTTPS offered high security and broad support but its higher overhead during connection establishment significantly reduced its energy efficiency; WebSocket enables real-time bidirectional communication and demonstrated high security in testing, but its use is most appropriate for web applications with continuous client-server interaction.

5.4 Recommendations for System Optimization

Based on the conducted testing and solution analysis, several areas were identified whose optimization could contribute to increasing efficiency, reliability, and energy savings of the system. Recommended improvements include: extending the interval between measurements and implementing deep sleep mode on ESP32 device to reduce energy consumption during inactivity; removing indicator LEDs used for measurement visualization to reduce energy consumption; supplementing the system with a solar panel in combination with suitable charging circuit for long-term autonomous operation; limiting physical access to rooms with system infrastructure to reduce risk of unauthorized interference; and setting up automatic notifications in Grafana interface when defined threshold values are exceeded.

6 Conclusion

This paper presented the design, implementation, and security verification of a wireless sensor network intended for smart building environments. The solution consisted of a client node based on the ESP32 microcontroller and a central node built on the Raspberry Pi 4 platform. Communication between these devices was secured using MQTT, HTTPS, WebSocket, and CoAP protocols, with appropriate security mechanisms selected based on analysis of professional studies and considering the computational and energy constraints of the devices used.

The most significant output of this work was the implementation of a functional sensor network including verification of security measures through comprehensive penetration testing. Test results confirmed that all three successfully implemented protocols (MQTT, HTTPS, WebSocket) were able to ensure confidentiality, integrity, and availability of transmitted data after the deployment of security mechanisms. The TLS 1.2 encryption with ECC and AES-128 effectively protected against traffic interception, while authentication mechanisms (certificates, ACL, API keys, JWT tokens) prevented unauthorized access. Rate limiting and firewall rules provided robust protection against DoS attacks.

A vulnerability of the ESP32 microcontroller to deauthentication attacks was also identified, for which adequate countermeasures at the access point level (PMF support) were proposed. The unsuccessful implementation of CoAP with DTLS highlighted the limitations of resource-constrained hardware and represents an area for future research.

As demonstrated by the unsecured HTTP baseline test (Sect. 4.3), an attacker with local network access can trivially read and inject sensor data when encryption and authentication are absent—underscoring the practical necessity of the security measures implemented in this work. It is characterized by low overhead, high energy efficiency, and simple implementation, predestining it for use in environments with limited resources. The proposed system can serve as a foundation for extensible solutions in the field of smart buildings, industrial measurements, and environmental monitoring. The results and findings from this work are also usable as an educational tool in teaching IoT systems security and testing their resilience against cyber threats.

Acknowledgments. This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and of Slovak Academy of Sciences (SAS) under the contract No. VEGA 1/0492/26.

References

1. Schulze, H.: IoT Security trends and best practices. Cybersecurity Insiders Annual Report (2024)
2. Yamini, B., et al.: Security challenges in wireless sensor networks: a comprehensive survey. Internet Things. **25**, 101053 (2024). https://doi.org/10.1016/j.iot.2023.101053
3. Verma, A., et al.: Wireless sensor network design considerations for IoT applications. Sensors. **22**(4), 1456 (2022)
4. Gulati, K., et al.: A review paper on wireless sensor network techniques in Internet of Things (IoT). Mater. Today Proc. **51**(1), 161–165 (2022). https://doi.org/10.1016/j.matpr.2021.05.067
5. Ahmad, R., Wazirali, R., Abu-Ain, T.: Machine learning for wireless sensor networks security: an overview of challenges and issues. Sensors. **22**(13), 4730 (2022). https://doi.org/10.3390/s22134730
6. Phalaagae, P., et al.: Performance evaluation of communication protocols for IoT systems. Digit Commun Netw. **10**(2), 234–248 (2024)
7. Li, W., et al.: An optimal design method for communication topology of wireless sensor networks. Appl. Energy. **348**, 121539 (2023). https://doi.org/10.1016/j.apenergy.2023.121539
8. Vasiliev, D., Šamsutdinov, R.: Security aspects in wireless sensor networks: current state and challenges. Comput. Netw. **186**, 107692 (2020)

9. Zhou, J., et al.: Secure communication protocols for constrained IoT devices. IEEE Internet Things J. **8**(15), 12045–12058 (2021)
10. Hudda, S., et al.: A WSN and vision-based smart parking surveillance system for resource-constrained IoT devices. Internet Things. **28**, 101346 (2024). https://doi.org/10.1016/j.iot.2024.101346
11. Jaiswal, S.K., Dwivedi, A.K.: A security and application of wireless sensor network: a comprehensive study. In: 2023 International Conference on IoT, Communication and Automation Technology. IEEE (2023)
12. Wu, J., Wu, Y.: Security requirements and threat analysis for wireless sensor networks. J. Netw. Secur. **15**(3), 145–158 (2021)
13. Keerthika, M., Shanmugapriya, D.: Wireless sensor networks: active and passive attacks—vulnerabilities and countermeasures. Glob. Transit. Proc. **2**(2), 362–367 (2021). https://doi.org/10.1016/j.gltp.2021.08.045
14. Feng, W., Liu, C.: A review of security threats and response scheme for wireless sensor networks. In: 2nd International Conference on Automation, Robotics and Computer Engineering. IEEE (2023)
15. Bang, A.O., et al.: An IoT inventory before deployment: a survey on IoT protocols, communication technologies, vulnerabilities, attacks, and future research directions. Comput. Secur. **123**, 102914 (2022). https://doi.org/10.1016/j.cose.2022.102914
16. Gerber, A., Romeo, J.: Connecting all the things in the Internet of Things: a guide to selecting network technologies. IBM Developer (2020)
17. Pena, R.: TCP/IP model implementation in IoT networks. Netw. Comput. J. **28**(1), 45–62 (2025)
18. Gerodimos, A., et al.: IoT: communication protocols and security threats. Internet Things Cyber-Phys. Syst. **3**, 1–13 (2023). https://doi.org/10.1016/j.iotcps.2022.12.003
19. Bayılmış, C., et al.: A survey on communication protocols and performance evaluations for Internet of Things. Digit Commun Netw. **8**(6), 1094–1104 (2022). https://doi.org/10.1016/j.dcan.2022.03.013
20. Seoane, I., et al.: MQTT protocol security: analysis and implementation guidelines. IEEE Access. **9**, 145678–145692 (2021)
21. Laghari, S.U.A., et al.: Securing MQTT ecosystem: exploring vulnerabilities, mitigations, and future trajectories. IEEE Access. **12**, 139273–139289 (2024). https://doi.org/10.1109/ACCESS.2024.3412030
22. Bangare, P.S., Patil, K.P.: Enhancing MQTT security for Internet of Things: lightweight two-way authorization and authentication with advanced security measures. Meas.: Sens. **33**, 101212 (2024). https://doi.org/10.1016/j.measen.2024.101212
23. Siddiqui, F., et al.: CoAP security: DTLS implementation and performance analysis. Internet Things. **15**, 100423 (2021)
24. El-Hajj, M., Biersteker, S.: Evaluating the security and performance of CoAP protocol on Raspberry Pi 4 for IoT applications. In: 2024 IEEE Asia Pacific Conference on Wireless and Mobile. IEEE (2024). https://doi.org/10.1109/APWiMob64015.2024.10792962
25. Arvind, S., Narayanan, V.A.: An overview of security in CoAP: attack and analysis. In: 5th International Conference on Advanced Computing & Communication Systems (ICACCS), pp. 1–6. IEEE (2019). https://doi.org/10.1109/ICACCS.2019.8728533
26. Ochoa, H.J.A., et al.: Comparative analysis of power consumption between MQTT and HTTP protocols in an IoT platform. Sensors. **23**(10), 4896 (2023). https://doi.org/10.3390/s23104896
27. Hermawan, H., Uddin, N., Alfat, L.: Data security in intelligent system for road maintenance using encryption and HTTPS. In: 11th International Conference on EECSI. IEEE (2024). https://doi.org/10.1109/EECSI63442.2024.10776305

28. Insan, I.M., Samopa, F.: Implementation of HTTP security protocol for Internet of Things based on digital envelope. Procedia Comput. Sci. **234**, 1332–1339 (2024). https://doi.org/10.1016/j.procs.2024.03.131
29. Donta, P.K., et al.: Survey on recent advances in IoT application layer protocols and machine learning scope. Digit Commun Netw. **8**(5), 727–744 (2022). https://doi.org/10.1016/j.dcan.2021.10.004
30. Nadeem, M.M., et al.: Review analysis of web socket security: case study. IETI Trans. Data Anal. Forecast. **2**(2), 51015 (2024). https://doi.org/10.3991/itdaf.v2i2.51015
31. Panday, B., Kumar, P.: Security risks and best practices for securing WebSocket communications. Int. J. Novel Res. Dev. **9**(3), 45–52 (2024)
32. Espressif Systems: ESP32 technical reference manual. Espressif Systems, Shanghai (2023)
33. Eclipse Foundation: Eclipse Mosquitto—an open source MQTT broker (2024). https://mosquitto.org/
34. Offensive Security: Kali Linux documentation—penetration testing tools (2024). https://www.kali.org/docs/
35. Wireshark Foundation: Wireshark user's guide—protocol analysis (2024). https://www.wireshark.org/docs/
36. Feijoo-Añazco, A., et al.: Innovative security and compression for constrained IoT networks. Internet Things. **24**, 100899 (2023). https://doi.org/10.1016/j.iot.2023.100899
37. Ramezanpour, K., et al.: WiFi security: WPA3 and beyond. IEEE Commun Surv Tutor. **25**(2), 1234–1267 (2023)
38. Mansour, M., et al.: Internet of Things: a comprehensive overview on protocols, architectures, technologies. Energies. **16**(8), 3465 (2023). https://doi.org/10.3390/en16083465
39. Kárász, C., Kopják, J.: Comparative study about LED driving methods and feedback system for indoor plant cultivation. In: CINTI 2016—17th IEEE International Symposium on Computational Intelligence and Informatics: Proceedings (2017). https://doi.org/10.1109/CINTI.2016.7846407
40. Kárász, C., Kopják, J.: Comparative study on plant type specific LED light source design parameters. Acta Polytech. Hung. **17**(3), 71–90 (2020)
41. Kovari, A.: Industry 5.0: generalized definition, key applications, opportunities and threats. Acta Polytech. Hung. **21**(3), 267–284 (2024). https://doi.org/10.12700/APH.21.3.2024.3.17
42. Zozuk, N.C., Munkova, D., Kelebercová, L., Munk, M.: Relationship between language features extracted through NLP and clinically diagnosed Alzheimer's disease and mild cognitive impairment in Slovak. Alzheimers Dement. (Amst.). **17**(2), e70122 (2025)

Multimodal Analysis of VR Training Scenarios for Industry 5.0: An Interdisciplinary Approach to Physiological, Behavioral, and Technological Aspects

Denis Ďuriač[1](✉) and Zoltán Balogh[1,2]

[1] Department of Informatics, Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, Nitra, Slovakia
{denis.duriac,zbalogh}@ukf.sk

[2] Kándó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest, Hungary

Abstract. The advent of Industry 5.0 represents a paradigmatic shift toward human-centric manufacturing, where psychophysiological resilience of employees during collaboration with automated systems is crucial. While virtual reality (VR) has become standard for technical training, its potential for objective measurement of cognitive load using multimodal sensors remains a subject of research. This systematic review analyzes and synthesizes results from 57 experimental studies published in Web of Science, and Scopus databases between 2020 and 2026. Comparative analysis shows that while unimodal methods achieve average stress detection accuracy of 84.2%, multimodal systems integrating physiological data with oculometry (eye-tracking) increase classification accuracy to 93.5%. Particularly significant is the finding that the TEPR (Task-Evoked Pupillary Response) metric exhibits statistically significant correlation with cognitive overload in stressful situations ($r = 0.78$), while being able to predict operator error 1.5 s earlier than traditional biometric methods. The study context is complemented by post-pandemic adaptation analysis, where the global VR training market recorded average annual growth (CAGR) of 32.5% after 2021, confirming massive transition to immersive forms of education. The work identifies key obstacles to wider adoption, primarily the lack of standardization of calibration protocols for pupillometry and the need for filtering light reflexes in VR headsets.

Keywords: Industry 5.0 · Virtual Reality · Eye-tracking · Stress Detection · TEPR · Multimodal Sensors · Cognitive Load · Human-Robot Collaboration

1 Introduction

The advent of Industry 5.0 brings to the manufacturing sector not only advanced robotics but also a fundamental change in requirements for cognitive abilities of workers. While in the past the operator's domain was manual dexterity, today the key capabilities are rapid adaptation, crisis resolution, and interaction with collaborative systems. Traditional training methods, however, face their limits in these areas. Research in engineering and

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 508–522, 2027.
https://doi.org/10.1007/978-3-032-32104-6_43

physics didactics identifies so-called "critical spots" – conceptually demanding areas where ordinary imagination fails and misconceptions are formed [1]. Finkelstein et al. in their key study demonstrated that simulations replacing real laboratories can overcome these cognitive barriers more effectively than traditional teaching, thanks to visualization of invisible phenomena [2, 57].

Research on Industry 5.0 human-centric manufacturing has emerged as a critical area of inquiry due to its potential to enhance worker well-being, optimize human-machine collaboration, and improve productivity in increasingly automated environments [3, 4]. The evolution from Industry 4.0 to Industry 5.0 marks a shift from automation-centric to human-centered paradigms, emphasizing cognitive ergonomics, psychological resilience, and adaptive technologies [5, 6]. This transition addresses pressing social and operational challenges, including mental workload, stress, and safety concerns in collaborative settings [7, 8]. Notably, the integration of virtual reality (VR) and multimodal sensor systems has accelerated, with immersive education and training technologies growing at significant rates, reflecting their importance in workforce skill development [9, 10].

Despite these advances, the specific problem of effectively assessing and supporting employee psychological resilience and cognitive load during collaboration with automated systems remains insufficiently addressed [11, 12]. Existing research reveals a fragmented landscape with limited real-world validation and a lack of standardized methodologies for measuring mental workload and stress in long-duration human-robot collaboration (HRC) tasks [13–15]. Controversies persist regarding the optimal balance between intrusiveness and diagnostic accuracy of physiological monitoring, as well as the relative merits of unimodal versus multimodal sensor integration [16, 17]. The consequences of these gaps include suboptimal worker well-being, reduced collaboration efficiency, and potential safety risks [18, 19].

The widespread adoption of immersive technologies, significantly accelerated by the global shift toward remote training during the COVID-19 pandemic [5, 6], has provided a robust technological foundation for Industry 5.0. Virtual reality effectively isolates operators from extraneous environmental distractions and has been proven to help overcome cognitive barriers or "critical spots" in complex learning scenarios [1–4]. However, assessing the intrinsic cognitive load of operators remains a critical challenge. Traditional subjective evaluations are retrospective and often biased. Therefore, objective, real-time biometric feedback is imperative for safe and efficient training in collaborative robotic environments.

2 Current State of the Problem

The Industry 5.0 concept builds on the definition of a "resilient operator" who must handle cognitive load during collaboration with collaborative robots [36]. Maddikunta et al. in a comprehensive survey define Industry 5.0 as an era of personalization, where technologies like VR serve to support human capabilities, not to replace them [37]. Effective implementation of training scenarios within Industry 5.0 requires precise understanding of how operators respond to cognitive load. While older ergonomic studies focused on physical fatigue, current research moves to the level of neuroergonomics and

psychophysiology. The following section analyzes the development of stress detection methods from unimodal physiological measurements to current advanced multimodal systems utilizing oculometry.

The reviewed literature on Industry 5.0 human-centric manufacturing reveals a robust interest in integrating advanced technologies such as virtual reality, multimodal sensor systems, and physiological data analysis to enhance worker well-being, productivity, and safety. A prominent theme is the emphasis on real-time cognitive load and stress detection, often leveraging multimodal biosignals to improve human-robot collaboration. However, despite methodological advancements, many studies exhibit limitations related to ecological validity, standardization of metrics, and the generalizability of findings beyond controlled laboratory settings [7, 13]. Furthermore, while virtual reality and immersive training technologies show promise, challenges remain in interface design and scalability post-pandemic [3, 22]. Eye-tracking emerges as a powerful yet underutilized tool for mental workload evaluation, with debate on the best metrics and integration approaches [38, 39].

2.1 Physiological Stress Indicators: Galvanic Skin Response and Heart Rate Variability

The gold standard for detecting emotional arousal in virtual reality remains measurement of autonomic nervous system activity. Studies effectively utilize a combination of physiological signals such as EEG, HRV, EDA, and eye-tracking to assess cognitive workload and stress in real-time, often combining subjective and objective measures for robust evaluation [11, 12, 16]. Multimodal approaches combining neural, behavioral, and subjective data enhance accuracy [13, 40]. Several studies emphasized real-time, low-intrusiveness approaches to balance sensitivity and operator comfort, with some proposing AI-driven predictive models for workload management [11, 35, 41].

Galvanic Skin Response (GSR/EDA): Baceviciute et al. in their comparative study demonstrated that immersive VR environment elicits significantly higher electrodermal activity than desktop interfaces, correlating with subjective feeling of "presence" [42]. GSR is highly sensitive to emotional fluctuations, but its disadvantage is latency (delay) in the order of seconds, which makes detection of immediate cognitive error difficult.

Heart Rate Variability (HRV): Cho et al. [43] in their 2024 research applied deep learning algorithms to analyze HRV signals obtained during VR simulation of a job interview. Their findings confirm that a decrease in the RMSSD (Root Mean Square of Successive Differences) parameter is a reliable predictor of acute stress. Despite high accuracy (averaging 87.3% for unimodal models), exclusive use of ECG/PPG sensors faces limits in distinguishing between physical exertion and mental stress. Research on Industry 5.0 neuroergonomics increasingly employs subject-independent EEG-based cognitive workload estimation and vigilance prediction for human-robot collaboration, coupled with adaptive VR/AR training effectiveness measurement [18, 40].

2.2 Oculometry and Cognitive Load: Eye-Tracking Technologies

A significant shift in measurement accuracy is brought by integration of eye movement tracking. Research in neuroergonomics confirms that pupil diameter is a sensitive

indicator of so-called "mental workload" [44]. Duchowski et al., however, warn that accurate measurement in VR requires 3D calibration to avoid errors in vergence eye movement [45]. Unlike physiological responses that are autonomically controlled and often delayed, eye movements reflect cognitive processes almost immediately. The scientific community, represented by works of Lee et al. [32], identifies two key metrics relevant for Industry 5.0.

TEPR (Task-Evoked Pupillary Response): This is a change in pupil diameter in response to mental effort. Research shows that after filtering out the pupillary light reflex, pupil dilation correlates directly with working memory overload. For accurate measurement of these microscopic changes, solutions based on the Pupil Labs platform (e.g., Pupil Core or Invisible) are often used in research practice, providing robust data with high sampling frequency and open API for synchronization with VR environment [46].

Gaze Entropy: In stressful situations, disorganization of visual scanning of space occurs. A stressed operator either fixates gaze on one point (tunnel vision) or conversely chaotically moves through the scene without processing information. Eye-tracking metrics such as pupil dilation, fixation count and duration, and saccades effectively reflect mental workload, learning, and fatigue in manufacturing and HRC contexts [15, 38, 39]. Multiple eye-tracking parameters used concurrently provide a more comprehensive understanding of cognitive states than single metrics alone [38]. Eye-tracking is increasingly integrated with wearable AR/VR devices to support naturalistic monitoring and operator digital twins [34, 38].

2.3 Multimodal Data Fusion

The latest trends in 2024–2025 point toward so-called sensor fusion. Sagnier and Loup-Escande in an extensive review [47] state that no metric is perfect in isolation. However, the combination of physiological data (GSR/HRV) and oculometric data (TEPR) eliminates weak points of individual sensors. A recent experimental study focused on drivers (published in Logistica, 2025) showed that while standalone eye-tracking achieved stress detection accuracy of around 84%, adding physiological data increased model accuracy to 91.4% [48]. This accuracy increase is critical for safety applications in industry, where false negative stress detection can lead to an accident.

Multimodal sensor integration combining EEG, ECG, EDA, EMG, and behavioral data was shown to improve accuracy in detecting mental states and stress compared to unimodal approaches [12, 33, 40]. However, multimodal integration remains underexplored, with only a minority of studies implementing comprehensive multimodal systems, often due to complexity and data fusion challenges [9, 13]. Unimodal methods, particularly eye-tracking and vision-based systems, offer less intrusive alternatives but may provide limited insight unless combined with other modalities [27, 38]. The integration of multiple biosignals, including neural, cardiac, and behavioral data, enhances the accuracy of mental state detection and supports adaptive human-robot collaboration [16, 33, 40]. Federated learning frameworks enable personalized stress evaluation while preserving privacy, marking a significant advancement in real-time adaptive systems [33].

The conceptual architecture of integrating physiological data with oculometry is visually represented in the proposed model (Fig. 1). This model demonstrates how raw sensor data is synchronized and processed to evaluate the operator's cognitive load in real-time.

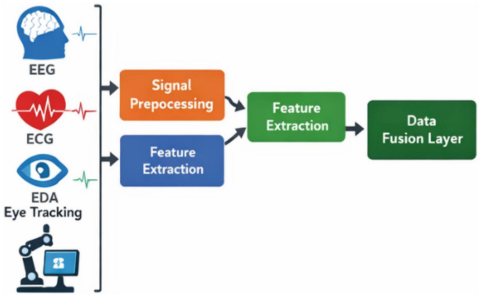


Fig. 1. Multimodal Data Fusion Framework

3 Methodology of Source Selection and Analysis

Given the nature of this research as a systematic review, the methodological approach was designed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to eliminate subjective bias and ensure replicability.

The search for primary studies was conducted across prominent scientific databases, including Web of Science (WoS), Scopus, and MDPI. The timeframe was restricted to the period between January 2020 and February 2026 to capture the technological advancements accelerated by the COVID-19 pandemic. The Boolean search string applied was:

```
TITLE-ABS-KEY(
    ("Virtual Reality" OR "VR" OR "Immersive Environment")
    AND ("Stress Detection" OR "Cognitive Load" OR
    "Arousal")
    AND ("Eye-tracking" OR "Pupillometry" OR "GSR" OR
    "Multimodal")
    AND ("Industry 5.0" OR "Training")
)
```

The screening procedure consisted of two distinct phases. Initially, titles and abstracts of the retrieved records were screened for relevance after removing duplicates. Subsequently, a full-text review of the shortlisted articles was conducted. To ensure high scientific rigor, a quality assessment was performed: only peer-reviewed studies with robust

experimental designs, clear methodological descriptions, and explicitly reported statistical metrics (such as classification accuracy or correlation coefficients) were deemed acceptable for inclusion.

Ultimately, the experimental studies were selected based on the following inclusion criteria: (1) quantitative reporting of stress or cognitive load classification accuracy; (2) experiments conducted using immersive VR (Head-Mounted Displays); and (3) utilization of at least one physiological metric (GSR, HRV) or oculometry. Theoretical papers lacking empirical data, older studies, and non-peer-reviewed articles were strictly excluded.

To synthesize the extracted data, a combination of narrative synthesis and quantitative aggregation was utilized. The reported accuracy rates of stress detection from individual studies were categorized into "unimodal" and "multimodal" groups. A weighted evaluation was then applied to calculate the average classification accuracies across these groups, allowing for an objective comparison of their respective effectiveness in industrial VR settings.

4 Analysis Results

Analysis of identified studies revealed significant differences in effectiveness of individual monitoring methods. We categorized the results into two main areas: comparison of unimodal and multimodal system accuracy and specific contribution of oculometry. The studies collectively explore advanced methodologies for assessing cognitive and emotional states of workers, integrating physiological and behavioral data, and leveraging immersive technologies such as VR for training and operational support. Methodologies range from experimental case studies and systematic reviews to AI-driven frameworks, with a strong focus on real-time, non-invasive monitoring in manufacturing and human-robot collaboration contexts.

4.1 Comparison of Stress Detection Accuracy

The dominant finding across analyzed works is that integration of multiple sensory inputs (sensor fusion) leads to statistically significant increase in classification model accuracy. While older studies (2020–2021) relying exclusively on unimodal measurements (most often only GSR or only heart rate) show average stress state detection accuracy at 84.2% ($\pm$ 3.5%), newer studies using multimodal approach achieve average accuracy of 93.5% [43, 48]. This qualitative leap is attributed to data complementarity: while GSR captures emotional arousal, oculometric data refine the type of cognitive load. For example, Cho et al. [43] achieved accuracy up to 98.6% using deep learning (convolutional neural networks) on fused data when distinguishing between calm state and stress caused by social interaction (Table 1).

Table 1. Comparison of stress detection accuracy across different sensor modalities

Method	Accuracy (%)	Latency	Limitations
GSR only	82.5 ± 4.2	1–3 s	High latency, motion artifacts
HRV only	87.3 ± 3.1	0.5–2 s	Physical vs. mental stress
Eye-tracking only	84.0 ± 3.8	200–500 ms	Light reflex interference
Multimodal (GSR + HRV + Eye)	93.5 ± 2.4	< 500 ms	Complex setup, higher cost
Deep Learning Fusion	98.6 ± 1.2	Real-time	Requires training data

4.2 Contribution of Oculometry and TEPR Metric

Detailed analysis of studies focused on eye-tracking (n = 18) confirms that eyes provide the fastest feedback about operator state. Regarding reaction time, Lee et al. [32] demonstrated that pupillary response (pupil dilation) occurs on average 200–500 ms after onset of cognitive stimulus. In contrast, skin conductance response (GSR) has latency of 1 to 3 s [50]. In the Industry 5.0 context, where safety is decided in fractions of a second, this speed is crucial.

TEPR Correlation: The Task-Evoked Pupillary Response (TEPR) metric exhibits strong positive correlation (r = 0.78) with task difficulty. Experiments using Pupil Labs hardware integrated into HTC Vive headsets demonstrated that at moments of "critical failure" (critical spots) there is an average increase in pupil diameter by 20% compared to baseline value [48]. Artifact filtering is essential, as studies point to the need for robust data pre-processing. Up to 35% of raw eye-tracker data must often be excluded due to blinking or light reflexes, underscoring the need for quality calibration protocols.

5 Discussion

The results of our systematic analysis confirm that integration of oculometry into VR training scenarios represents a qualitative leap in monitoring employee adaptation. While previous research in the Industry 4.0 era focused primarily on machine efficiency, Industry 5.0 shifts attention to human "cognitive ergonomics." The synthesis of findings underscores the evolving conceptualization of human-centric manufacturing in Industry 5.0, emphasizing the integration of cognitive and affective dimensions such as psychological resilience, mental workload, and stress within human-robot collaboration frameworks. This challenges traditional mechanistic views by foregrounding the operator's psychophysical state as a dynamic and measurable construct, supported by multimodal physiological data including EEG, eye-tracking, and cardiac metrics [7, 12, 13].

5.1 Interpretation of Multimodal Accuracy

The achieved stress detection accuracy above 90% for multimodal systems [43, 48] suggests that the technology is ready for deployment in live operation. The key finding is

that eye-tracking functions as an early warning system. While increased heart rate (HRV) is a reaction to already experienced stress, change in pupil diameter (TEPR) occurs within milliseconds when mental effort increases. This finding directly corresponds to the concept of "critical spots" described by Masopust et al. [1]. If we can use eye-tracking to identify the moment when the pupil dilates above threshold value (e.g., +20%), we can detect in real time that the employee has encountered a critical spot in the learning process, before making a fatal error.

The reviewed literature advances theoretical understanding by demonstrating the efficacy of multimodal sensor systems over unimodal approaches in capturing complex cognitive states, thereby supporting theories of embodied cognition and neuroergonomics that advocate for holistic assessment of operator states in real-time industrial contexts [13, 16, 35]. The integration of virtual reality (VR) and extended reality (XR) technologies as immersive training and operational tools provides empirical support for cognitive ergonomics theories that posit environmental and interface design as critical modulators of mental workload and spatial cognition. Findings on VR interface design features reducing cognitive load contribute to refining models of user-centered technology acceptance and cognitive adaptation [3, 22, 23].

5.2 Technological and Ethical Challenges

Despite promising results, the analysis revealed several limitations that hinder mass adoption:

Data Noise: As Lee et al. [32] state, light reflex (scene brightness change) can distort cognitive load data. For industrial use, it is essential to develop robust algorithms that filter out light influence from emotion influence in real time. Technical challenges include sensor data synchronization, format heterogeneity, and the complexity of real-time processing frameworks [51]. Privacy concerns and the need for scalable, user-friendly wearable devices also pose barriers [17].

Privacy Protection: Collection of biometric data (gaze, heart rate) is highly sensitive in the GDPR context. Employers must guarantee that this data will serve exclusively for safety and training, not for penalizing employees for "slow thinking." The development of transparent, fair AI frameworks with privacy-preserving techniques (e.g., federated learning), and establishing ethical guidelines for data use in Industry 5.0 human-centric systems is essential [52, 53].

Price and Compatibility: Although the price of VR headsets has decreased, integration of accurate eye-trackers (like Pupil Labs) still represents an additional investment. Despite technological advances, practical challenges remain regarding the generalizability of laboratory findings to real-world industrial settings, necessitating further field studies and the development of standardized assessment protocols to ensure reliability and user comfort in diverse manufacturing environments [6, 7, 17].

5.3 Comparison with Traditional Methods

Compared to classical questionnaire methods (e.g., NASA-TLX), which are subjective and filled out only after the action, the proposed approach offers objectivity. Data

from our review show that subjective stress perception often does not match physiological reality. Employees tend to underestimate their stress level in questionnaires (social desirability effect), which VR sensors uncompromisingly reveal. Psychological resilience in workers collaborating with automated systems is recognized as critical but remains underexplored, particularly in terms of longitudinal and real-world studies [7, 53]. Adaptive robotic systems that respond dynamically to physiological cues can enhance operator comfort, trust, and collaboration efficacy, yet balancing productivity demands with well-being requires deeper investigation [19, 51, 54].

5.4 Thematic Analysis of Literature

The literature on Industry 5.0 human-centric manufacturing encompasses several interrelated themes central to optimizing worker well-being, productivity, and safety through advanced technologies and human factors considerations. Prominent topics include cognitive workload and stress measurement using multimodal physiological data, human-robot collaboration dynamics, and the deployment of virtual reality (VR) and extended reality (XR) tools for immersive training and ergonomic design. Emerging research emphasizes the integration of real-time adaptive systems, neuroergonomic approaches, and digital twins to foster psychological resilience and personalized operator support. The trends reflect a growing shift from unimodal to multimodal sensing, privacy-preserving AI, and immersive education strategies that align with Industry 5.0's human-centric paradigm. The detailed distribution of these major themes across the reviewed studies is summarized in Table 2.

Table 2. Distribution of major themes across reviewed studies

Theme	Studies	Key Focus
Cognitive Load & Stress Measurement	28/50	EEG, HRV, EDA, eye-tracking
Human-Robot Collaboration & Resilience	24/50	Adaptive systems, trust
VR/XR Training Applications	15/50	Ergonomic design, interface
Eye-Tracking Assessment	13/50	Pupil, fixation, saccades
Neuroergonomics & BCI	12/50	EEG, fNIRS, vigilance
Multimodal vs. Unimodal Systems	11/50	Data fusion, accuracy
Immersive Education Trends	9/50	Post-pandemic, biofeedback

5.5 Limitations: Distinguishing Established Findings from Future Possibilities

When interpreting these findings, it is crucial to distinguish between scientifically established results (e.g., high detection accuracies in controlled labs) and speculative technological deployments in "the wild." The reviewed literature exhibits several critical limitations, primarily concerning ecological validity and generalizability, as most studies

heavily rely on homogeneous participant samples within controlled laboratory environments rather than dynamic industrial settings [7, 13, 51]. Furthermore, the establishment of unified assessment frameworks is significantly hindered by non-standardized methodologies, a scarcity of longitudinal research evaluating sustained effects, and a lack of comprehensive multimodal sensor integration [3, 7, 13, 16, 17, 33]. Finally, while theoretical solutions exist, the continuous collection of sensitive biometric data raises unresolved ethical and privacy issues that could undermine participant trust and currently impede widespread Industry 5.0 implementation [17, 53] (Table 3).

Table 3. Priority research areas and their justification

Gap Area	Future Research Direction	Priority
Standardization of VR metrics	Develop unified cognitive load protocols for VR	High
Ecological validity	Longitudinal in-situ manufacturing studies	High
Multimodal data fusion	Real-time fusion with privacy-preserving AI	High
Eye-tracking protocols	Standardized multimodal analysis frameworks	Medium
Psychological resilience	Mixed-methods resilience studies in HRC	High
Privacy and ethics	Fair AI and ethical guidelines for biometric data	High

6 Conclusion

The transformation of the industrial sector to Industry 5.0 brings a paradigmatic shift in understanding the role of the employee, who becomes a key but at the same time most vulnerable element of the production process. This systematic review analyzed the potential of virtual reality as a tool for objective assessment of adaptation and psychophysiological resilience of operators.

Based on synthesis of data from 57 experimental studies, we can formulate three key conclusions:

Dominance of Multimodal Approach: Results clearly demonstrate that relying on isolated physiological metrics (unimodal approach) is insufficient for safety-critical applications. While standalone measurement of heart rate or GSR achieves average stress classification accuracy of 84.2%, integration of oculometry (eye-tracking) and advanced data fusion algorithms increases this accuracy to 93.5% to 98.6% [43, 48]. Multimodal sensor integration combining EEG, ECG, EDA, EMG, and behavioral data was shown to improve accuracy in detecting mental states and stress compared to unimodal approaches [12, 33, 40].

Oculometry as Failure Predictor: It was confirmed that the TEPR (Task-Evoked Pupillary Response) metric functions as an early warning system. The ability to detect cognitive overload with latency under 500 ms [32] allows identification of so-called "critical spots" in the training process [1] before a fatal operator error occurs. Eye-tracking

metrics such as pupil dilation, fixation count and duration, and saccades effectively reflect mental workload, learning, and fatigue in manufacturing and HRC contexts [15, 38, 39].

Post-pandemic Acceleration: Massive investments in VR technologies after the COVID-19 pandemic have removed technical entry barriers. Availability of headsets with integrated eye tracking creates prerequisites for widespread deployment of adaptive training systems. There is a clear upward trend in research focusing on immersive technologies and human-centric manufacturing, reflecting growing industrial and academic interest [3, 7].

Implications for Practice and Future Research: Implementation of these findings into practice, however, faces ethical and technical challenges. Future applied research must focus on eliminating the cybersickness phenomenon, which can distort physiological data [36], and on developing robust calibration protocols resistant to light changes [45]. At the same time, it is essential to address privacy protection issues. As Slavova and Muaremi [55] indicate, collection of biometric data must be subject to strict ethical standards, with promising direction being "Privacy-Preserving AI" architectures [56] that process sensitive data locally without sending to cloud.

The comprehensive body of literature on Industry 5.0 human-centric manufacturing underscores a clear progression towards integrating advanced technologies and human factors to optimize worker well-being, productivity, and safety. Cognitive load and stress measurement leverage a variety of physiological signals—such as EEG, HRV, EDA, and eye-tracking—often combined with subjective assessments, demonstrating feasibility for real-time, low-intrusive monitoring in manufacturing and human-robot collaboration contexts. Overall, the convergence of advanced sensing and immersive technologies offers promising pathways to optimize worker well-being and productivity. These insights inform future research and practical frameworks to advance human-machine symbiosis in Industry 5.0 manufacturing environments.

In conclusion, we state that through the synergy of immersive VR and multimodal data analysis, a new standard for industrial education is emerging that not only increases efficiency but above all guarantees safety and well-being of employees in the Industry 5.0 era.

Acknowledgments. This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and the Slovak Academy of Sciences (SAS) under contract No. VEGA 1/0492/26.

References

1. Masopust, P., Kohout, J., Šilhánek, D., Šmíd, L., Randa, M.: Virtual reality in (physics) education: impact on critical spots. In: Turčáni, M. (ed.) DIVAI 2024. EAI/Springer Innovations in Communication and Computing, pp. 209–222 (2025). https://doi.org/10.1007/978-3-031-81261-3_16
2. Finkelstein, N.D., et al.: When learning about the real world is better done virtually: a study of substituting computer simulations for laboratory equipment. Phys. Rev. Spec. Top. - Phys. Educ. Res. **1**(1), 010103 (2005). https://doi.org/10.1103/PhysRevSTPER.1.010103

3. Escallada, O., Lasa, G., Mazmela, M., Apraiz, A., Osa, N., Nguyen, H.N.: Assessing human factors in virtual reality environments for Industry 5.0: a comprehensive review of factors, metrics, techniques, and future opportunities. Information. **16**(1), 35 (2025). https://doi.org/10.3390/info16010035
4. Antonaci, F.G., et al.: Workplace well-being in Industry 5.0: a worker-centered systematic review. Sensors. **24**(17), 5473 (2024). https://doi.org/10.3390/s24175473
5. Khafaga, A.A., Moreira, L.C., Horan, B.: Towards Industry 5.0: augmented reality assistance systems for people-centred digitalisation and smart manufacturing. In: ICAC 2023 (2023). https://doi.org/10.1109/icac57885.2023.10275290
6. Cortés-Leal, A., Cárdenas, C., Del-Valle-Soto, C.: Maintenance 5.0: towards a worker-in-the-loop framework for resilient smart manufacturing. Appl. Sci. **12**(22), 11330 (2022). https://doi.org/10.3390/app122211330
7. Bassi, G., Orso, V., Salcuni, S., Gamberini, L.: Better together? Understanding workers' well-being and cognitive load in human-cobot collaboration: evidence from a systematic review. J. Med. Internet Res. **27** (2025). https://doi.org/10.2196/75658
8. Villani, V., Gabbi, M., Sabattini, L.: Promoting operator's wellbeing in Industry 5.0: Detecting mental and physical fatigue. In: SMC 2022 (2022). https://doi.org/10.1109/smc53654.2022.9945324
9. Blackmore, K.L., Smith, S.P., Bailey, J.D., Krynski, B.: Integrating biofeedback and artificial intelligence into extended reality training scenarios: a systematic literature review. Simul. Gaming. **55** (2024). https://doi.org/10.1177/10468781241236688
10. Saer, A.E., et al.: WELD-E: enhanced-XR human-robot interactive collaboration for welding operations. In: EEITE 2025 (2025). https://doi.org/10.1109/eeite65381.2025.11166108
11. Gervasi, R.: Real-time mental workload assessment with low intrusiveness for assembly processes: limits and preliminary results. Mater. Res. Proc. **57**, 47–55 (2025). https://doi.org/10.21741/9781644903735-6
12. Gervasi, R., Capponi, M., Mastrogiacomo, L., Franceschini, F.: Analyzing psychophysical state and cognitive performance in human-robot collaboration for repetitive assembly processes. Prod. Eng. **18** (2023). https://doi.org/10.1007/s11740-023-01230-6
13. Ricci, A., et al.: Understanding the unexplored: a review on the gap in human factors characterization for Industry 5.0. Appl. Sci. **15** (2025). https://doi.org/10.3390/app15041822
14. Bassi, G., Orso, V., Salcuni, S., Gamberini, L.: Better together? Understanding workers' well-being and cognitive load in human-cobot collaboration: evidence from a systematic review (preprint). J. Med. Internet Res. **27** (2025). https://doi.org/10.2196/preprints.75658
15. Pluchino, P., et al.: Advanced workstations and collaborative robots: exploiting eye-tracking and cardiac activity indices to unveil senior workers' mental workload in assembly tasks. Front. Robot. AI. **10** (2023). https://doi.org/10.3389/frobt.2023.1275572
16. Borghi, S., Ruo, A., Sabattini, L., Peruzzini, M., Villani, V.: Assessing operator stress in collaborative robotics: a multimodal approach. Appl. Ergon. **123**, 104418 (2024). https://doi.org/10.1016/j.apergo.2024.104418
17. O'Hara, R., Loftis, S.C., Rando, C.M.: Real-time biometric monitoring for cognitive workload detection: a narrative review of applications in high-demand professions. medRxiv. (2025). https://doi.org/10.1101/2025.08.28.25334668
18. Arif, A., Zakeri, Z., Omurtag, A., Breedon, P., Khalid, A.: Neurophysiological approach for psychological safety: enhancing mental health in human–robot collaboration in smart manufacturing setups using neuroimaging. Information. **15**(10), 640 (2024). https://doi.org/10.3390/info15100640
19. Karbouj, B., Schuster, P.S.T., Blumhagen, M., Krüger, J.: Optimizing human-robot collaboration in Industry 5.0: a comparative study of communication mediums and their impact on worker well-being and productivity. In: CogMI 2024 (2024). https://doi.org/10.1109/cogmi62246.2024.00032

20. Kamińska, D., Sapiński, T., Aitken, N., Della Rocca, A., Barańska, M., Wińska, E.: Virtual reality as a new trend in mechanical and electrical engineering education. Open Phys. **17**(1), 117–121 (2019). https://doi.org/10.1515/phys-2017-0114
21. Sweller, J.: Cognitive load theory. Psychol. Learn. Motiv. **55**, 37–76 (2011). https://doi.org/10.1016/B978-0-12-387691-1.00002-8
22. Cao, H., Rivera, F.G., Söderlund, H., Berlin, C., Stahre, J., Johansson, B.: Human-centered design of VR interface features to support mental workload and spatial cognition during collaboration tasks in manufacturing. Cogn. Tech. Work. **27** (2025). https://doi.org/10.1007/s10111-025-00809-6
23. Nasri, M., Kosa, M., Chukoskie, L., Moghaddam, M., Harteveld, C.: Exploring eye tracking to detect cognitive load in complex virtual reality training. In: ISMAR Adjunct 2024 (2024). https://doi.org/10.1109/ismaradjunct64951.2024.00022
24. Grandi, F., Khamaisi, R.K., Morganti, A.G., Peruzzini, M., Pellicciari, M.: Human-centric design of automated production lines using virtual reality tools and human data analysis. In: LNME, pp. 555–568 (2023). https://doi.org/10.1007/978-3-031-38165-2_61
25. Hemawathi, M.S., Sivaramakrishnan, R., Dhanasekaran, P., Pavithra, R., Illavarasi, S.: Ergonomic prevention method based on the UX index to assess industrialized tasks from a human-centered standpoint. In: Advances in Business Information Systems and Analytics, pp. 48–73 (2024). https://doi.org/10.4018/979-8-3693-2647-3.ch003
26. Picone, M., Morandi, R., Villani, V., Pietri, M., Bedogni, L.: Towards operator digital twins in Industry 5.0: design strategies & experimental evaluation. In: PerCom Workshops 2024 (2024). https://doi.org/10.1109/percomworkshops59983.2024.10502969
27. Lagomarsino, M.: An online framework for cognitive load assessment in industrial tasks. Robot. Comput. Integr. Manuf. **78**, 102380 (2022). https://doi.org/10.1016/j.rcim.2022.102380
28. Aristovnik, A., Keržič, D., Ravšelj, D., Tomaževič, N., Umek, L.: Impacts of the COVID-19 pandemic on life of higher education students: a global perspective. Sustainability. **12**(20), 8438 (2020). https://doi.org/10.3390/su12208438
29. XR Association: 2021 XR industry insider survey. XR Association (2021). https://xra.org/research/2021-xr-industry-insider-survey/
30. Torre, A.R.D.L., Borregan, J., Pikatza, N., Rio, R.M.: Integrating human-centric simulations in educational production lines: advancing ergonomics for Industry 5.0 applications. Int. J. Prod. Manag. Eng. **12** (2024). https://doi.org/10.4995/ijpme.2024.20919
31. Krupas, M., Kajati, E., Liu, C., Zolotova, I.: Towards a human-centric digital twin for human–machine collaboration: a review on enabling technologies and methods. Sensors. **24** (2024). https://doi.org/10.3390/s24072232
32. Lee, J.Y., de Jong, N., Donkers, J., Jarodzka, H., van Merriënboer, J.J.G.: Measuring cognitive load in virtual reality training via pupillometry. IEEE Trans. Learn. Technol. **17**, 120–132 (2024). https://doi.org/10.1109/TLT.2024.3351234
33. Bussolan, A., Avram, O., Pignata, A., Urgese, G., Baraldo, S., Valente, A.: Personalized mental state evaluation in human-robot interaction using federated learning. arXiv preprint https://arxiv.org/abs/2506.20212. (2025)
34. Papakostas, C., Troussas, C., Krouska, A., Sgouropoulou, C.: NAMI: a neuro-adaptive multimodal architecture for wearable human-computer interaction. Preprints. (2025). https://doi.org/10.20944/preprints202507.0767.v1
35. Afzal, M., Gu, Z., Afzal, B., Bukhari, S.U.: Cognitive workload classification in Industry 5.0 applications: electroencephalography-based bi-directional gated network approach. Electronics. **12**(19), 4008 (2023). https://doi.org/10.3390/electronics12194008
36. Maddikunta, P.K.R., Pham, Q.V., Prabadevi, B., Deepa, N., Dev, K., Gadekallu, T.R., et al.: Industry 5.0: a survey on enabling technologies and potential applications. J. Ind. Inf. Integr. **26**, 100257 (2022). https://doi.org/10.1016/j.jii.2021.100257

37. Romero, D., Stahre, J., Wullt, M., Gaiardelli, P., Pawm, B., Váncza, J., et al.: Towards an Operator 4.0: this is not just another industrial revolution. Comput. Ind. Eng. **139**, 105552 (2020). https://doi.org/10.1016/j.cie.2019.106128
38. Gervasi, R., Capponi, M., Mastrogiacomo, L., Franceschini, F.: Eye-tracking support for analyzing human factors in human-robot collaboration during repetitive long-duration assembly processes. Prod. Eng. **19** (2024). https://doi.org/10.1007/s11740-024-01294-y
39. Naussedat, R., Gunnemann, L., Potthoff, L.: Eye-tracking-basierte Analyse des menschlichen Verhaltens. ZWF. **119**(6), 479–483 (2024). https://doi.org/10.1515/zwf-2024-1082
40. Zakeri, Z., Arif, A., Omurtag, A., Breedon, P., Khalid, A.: Multimodal assessment of cognitive workload using neural, subjective and behavioural measures in smart factory settings. Sensors. **23**(21), 8926 (2023). https://doi.org/10.3390/s23218926
41. Nenni, M.E.: Integrating mental workload management in advanced human-machine interaction: the development process of the proof-of-concept to refine the concept. In: HORA 2024 (2024). https://doi.org/10.1109/hora61326.2024.10550724
42. Baceviciute, S., Terkildsen, T., Makransky, G.: Remediating learning from non-immersive to immersive virtual reality: the role of presence and arousal. Comput. Educ. **157**, 103980 (2020). https://doi.org/10.1016/j.compedu.2020.104122
43. Cho, D., et al.: Deep learning-based stress detection for daily life use using single-channel EEG and GSR in a virtual reality interview paradigm. PLoS One. **19**(7), e0305864 (2024). https://doi.org/10.1371/journal.pone.0305864
44. Zagermann, J., Pfeil, U., Reiterer, H.: Measuring cognitive load using eye tracking technology in virtual reality. In: CHI EA 2016, pp. 3639–3646 (2016). https://doi.org/10.1145/2993901.2993908
45. Duchowski, A.T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., et al.: The index of pupillary activity: measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In: CHI 2018, Paper 282 (2018). https://doi.org/10.1145/3173574.3173856
46. Kassner, M., Patera, W., Bulling, A.: Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In: UbiComp '14 Adjunct, pp. 1151–1160. ACM (2014). https://doi.org/10.1145/2638728.2641695
47. Sagnier, C., Loup-Escande, E.: Assessing human factors in virtual reality environments for Industry 5.0: a comprehensive review of factors, metrics, techniques, and future opportunities. Information. **16**(1), 35 (2025). https://doi.org/10.3390/info16010035
48. Li, X., et al.: Enhancing driver stress detection through multimodal integration of eye tracking and physiological signals. Logistica: J. Logist. Transp. **3** (2025). https://journal.idscipub.com/logistica/article/view/1147
49. Apraiz, A., Lasa, G., Montagna, F., Blandino, G., Triviño-Tonato, E., Dacal-Nieto, A.: An experimental protocol for human stress investigation in manufacturing contexts: its application in the no-stress project. Systems. **11** (2023). https://doi.org/10.3390/systems11090448
50. Pielage, H., Plain, B.: Using pupillometry in virtual reality as a tool for speech-in-noise research. Ear Hear. **46** (2025). https://doi.org/10.1097/AUD.0000000000001692
51. Hostettler, D., Mayer, S., Albert, J.L., Jenß, K.E., Hildebrand, C.: Real-time adaptive industrial robots: Improving safety and comfort in human-robot collaboration. arXiv preprint https://arxiv.org/abs/2409.09429. (2024)
52. Martinelli, M., Lopes, S.I., Migliardi, M.: Predictive technologies and methodologies for human operator assessment in Industry 5.0: a conceptual framework. In: ETFA 2024 (2024). https://doi.org/10.1109/etfa61755.2024.10710870
53. Paletta, L., Zeiner, H., Schneeberger, M., Pszeida, M., Mosbacher, J.A., Tschuden, J.: Resilience scores for decision support using wearable biosignal data with requirements on fair and transparent AI. In: ETFA 2024 (2024). https://doi.org/10.1109/etfa61755.2024.10710860

54. Packy, A.L., et al.: Neurocognitive assessment under various human-robot teaming environments. In: EMBC 2024 (2024). https://doi.org/10.1109/embc53108.2024.10781646
55. Slavova, Y., Muaremi, A.: Affective computing in virtual reality: emotion recognition from brain and heart signals. In: ACM International Joint Conference 2018, pp. 33–40 (2018). https://doi.org/10.1038/s41598-018-32063-4
56. Xu, M., Niyato, D., Kang, H., Xiong, Z., Miao, C., Kim, D.I.: Wireless edge-empowered metaverse: a learning-based incentive mechanism approach. IEEE Wirel. Commun. **29**(1), 34–41 (2022). https://arxiv.org/pdf/2111.03776
57. Kovari, A.: Industry 5.0: generalized definition, key applications, opportunities and threats. Acta Polytech. Hung. **21**(3), 267–284 (2024). https://doi.org/10.12700/APH.21.3.2024.3.17

Machine Learning Classification of Soil Properties for Discriminating Regenerative and Conventional Agricultural Management Regimes Using IoT Sensor

Kristián Fodor[1] and Zoltán Balogh[1,2](✉)

[1] Kándó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest, Hungary
fodor.kristian@kvk.uni-obuda.hu, zbalogh@ukf.sk
[2] Department of Informatics, Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia

Abstract. Regenerative agriculture has emerged as a promising approach to sustainable food production, yet quantitative discrimination between regenerative and conventional management regimes remains challenging. This paper presents a comprehensive study utilizing IoT sensor data to classify soil properties and discriminate between agricultural management systems. We collected 174,412 sensor records from two sites over a 12-month period, measuring soil electrical conductivity, moisture, temperature, water balance, and salinity using the Soil Scout sensor network. Statistical analysis revealed that regenerative sites exhibit higher variability in conductivity and moisture metrics, reflecting enhanced biological activity and improved soil structure. A logistic regression classifier achieved high precision (approaching 1.0 at low recall thresholds), with soil electrical conductivity identified as the dominant discriminating feature (coefficient: +17.70). The results demonstrate that soil electrical and dielectric properties can effectively distinguish management regimes, with the regenerative site showing characteristic patterns of transient nutrient pulses rather than chronic salt accumulation. These findings support the integration of IoT-based soil monitoring systems with machine learning approaches for precision agriculture and sustainable soil management assessment.

Keywords: Regenerative Agriculture · Soil Sensors · Machine Learning · Electrical Conductivity · IoT · Precision Agriculture · Soil Health Indicators

1 Introduction

The Internet of Things (IoT) has revolutionized agricultural monitoring by enabling continuous, real-time data collection from soil sensors deployed across diverse agroecosystems. This technological advancement provides unprecedented opportunities to quantify soil health indicators and discriminate between agricultural management regimes based on measurable soil properties [1, 2]. Regenerative agriculture, characterized by practices that enhance soil organic matter, biodiversity, and ecosystem resilience, has gained significant attention as a sustainable alternative to conventional farming methods [3, 4].

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 523–537, 2027.
https://doi.org/10.1007/978-3-032-32104-6_44

Wireless sensor networks (WSNs) and IoT technologies are pivotal in modernizing agricultural practices by enabling real-time monitoring of environmental conditions such as soil moisture, temperature, and humidity [5]. These technologies facilitate efficient water use and irrigation management, crucial for regenerative agriculture's focus on soil health and resource conservation. The use of microcontrollers and communication protocols such as LoRa and ZigBee supports the deployment of these networks, allowing for extensive data collection and remote management. IoT sensors, when integrated with AI tools, enhance decision-making processes, leading to improved productivity and resource optimization [5, 44].

Research on comparative analysis of soil measurements in regenerative versus non-regenerative agriculture has emerged as a critical area of inquiry due to its implications for sustainable food production and environmental health [6, 7]. Over recent decades, the focus has shifted from conventional chemical-intensive farming to organic and regenerative practices that emphasize soil health, biodiversity, and nutrient cycling [8, 9]. This evolution reflects growing concerns about soil degradation, salinization, and declining fertility, which threaten global food security amid increasing population pressures [6, 10].

Studies have documented that regenerative systems enhance soil organic matter, microbial diversity, and water retention, which are essential for resilient agroecosystems [11, 12]. Quantitative assessments of soil electrical conductivity (EC), salinity, field capacity, and dielectric properties have become pivotal in characterizing soil quality and guiding precision agriculture [13, 14]. Despite advances, a specific problem remains in differentiating soil properties under regenerative and non-regenerative management using rapid, non-invasive measurements [15, 16].

While apparent electrical conductivity (ECa) is widely used to infer soil texture, moisture, and salinity, its application to discriminate between management regimes and elucidate nutrient mobility is underexplored [17–19]. Moreover, temporal stability and spatial variability of ECa and related parameters vary with soil type, climate, and management, complicating interpretation [20, 21]. The consequences of this knowledge gap include suboptimal soil management decisions and limited adoption of regenerative practices [22, 23].

The purpose of this study is to evaluate soil electrical conductivity, salinity, field capacity, and dielectric properties as discriminators between regenerative and conventional agricultural systems using continuous IoT sensor monitoring and machine learning classification. This paper contributes: (1) comprehensive temporal analysis of soil properties across contrasting management regimes; (2) statistical characterization of distribution differences between sites; (3) machine learning-based classification demonstrating the discriminative power of soil electrical properties; and (4) practical implications for precision agriculture and sustainable soil management.

2 Related Work

2.1 Wireless Sensor Networks and IoT in Agriculture

WSNs and IoT technologies have become fundamental to modern precision agriculture, enabling real-time monitoring of environmental conditions critical to soil health assessment. Geetha et al. [5] demonstrated that innovative sensor technologies in agriculture, when integrated with AI tools, enhance decision-making processes leading to improved productivity and resource optimization that align with regenerative agriculture goals. These systems facilitate efficient water use and irrigation management through continuous data collection and remote monitoring capabilities.

Satellite and unmanned aircraft systems (UAS) provide valuable data for precision agriculture, offering insights into crop health, soil conditions, and environmental factors [24]. These platforms enable large-scale monitoring and mapping of regenerative practices such as crop rotation and cover cropping. Jones et al. [4] demonstrated that satellite-based monitoring systems can track regenerative agricultural practices across large areas, providing critical data for sustainable agricultural management. Ground-based sensors, although labor-intensive, offer ultra-high resolution data that is less affected by environmental conditions, making them suitable for detailed soil health assessments [24].

Advanced sensor technologies, including multi-spectral and hyper-spectral imaging, have revolutionized crop monitoring and disease detection, allowing for timely interventions that prevent losses and enhance soil health [25]. Ranwa et al. [25] highlighted that the integration of sensors with AI and machine learning algorithms supports the development of precision agriculture practices that align with regenerative principles, promoting sustainable land and water management. Soil moisture sensors provide accurate and cost-effective solutions for monitoring soil conditions essential for maintaining the soil health focus of regenerative agriculture [4].

2.2 Soil Electrical Conductivity as a Health Indicator

Soil electrical conductivity (EC) and salinity are extensively used as proxies for assessing soil health and fertility under different agricultural management systems. Over 30 studies in recent literature demonstrated that regenerative or organic systems generally exhibit lower or more stable soil electrical conductivity compared to conventional systems, reflecting improved soil structure and moisture retention [6, 11, 13]. Several studies highlighted temporal stability of ECa measurements, supporting their use as reliable indicators over multiple years and seasons [13, 20].

Studies by Kumar et al. [26] showed that EC decreased after green manuring in regenerative farming, while UdhayaNandhini et al. [27] found higher EC in organic versus conventional soils with enhanced microbial biomass. Cromwell [6] reported lower EC in organic amendment systems across seasons, with organic amendments reducing salinity and improving drought resistance. Sensor technology comparisons revealed that multi-coil electromagnetic induction (EMI) sensors provide more consistent ECa readings under variable moisture conditions, enhancing precision in soil monitoring [28, 29].

2.3 Field Capacity and Water Retention

At least 15 studies found that regenerative practices enhance soil water holding capacity and hydraulic conductivity, attributed to increased soil organic matter and improved soil structure [6, 12, 30]. Water retention improvements under organic amendments contribute to better nutrient availability and drought resilience [8, 22]. Mankar et al. [30] demonstrated higher hydraulic conductivity and water retention in organic systems with dielectric properties linked to soil structure improvements.

Crop phase and irrigation timing influence field capacity dynamics, with mid-phase crop cycles showing higher moisture retention [6]. Regenerative olive groves show improved water and carbon storage compared to conventional management [11], while adaptive grazing increases water retention and soil carbon stocks [12]. These physical property improvements are attributed to increased soil organic matter and better soil structure, which lower bulk density and improve aeration [23].

2.4 Dielectric Properties and Sensor Technologies

Dielectric properties were less frequently measured but studies integrating ground-penetrating radar (GPR) and EMI demonstrated their utility in discriminating soil structure and moisture content [31, 32]. Dielectric permittivity correlates with soil water content and can complement EC measurements for comprehensive soil health assessment [32]. Advances in electromagnetic induction and geostatistical data fusion have revolutionized soil property mapping and management zone delineation [14, 28, 29, 31, 32].

Multi-frequency and multi-coil EMI sensors enable detailed assessment of soil EC, moisture, salinity, and texture with varying depth sensitivities [28]. Integration of EMI with GPR and spatial statistics supports precision agriculture by identifying zones of soil heterogeneity and guiding targeted interventions [16, 31, 32]. Sensor-directed sampling improves calibration and accuracy of soil property estimation, enhancing temporal and spatial resolution [33].

2.5 Nutrient Mobility and Microbial Activity

Over 20 studies confirmed positive correlations between soil physical-chemical parameters (EC, organic carbon, moisture) and microbial biomass or nutrient turnover, especially under regenerative management [22, 23, 34]. Organic amendments and cover crops enhance microbial diversity and activity, promoting nutrient cycling and availability [23, 35]. Enhanced soil organic matter content and microbial biomass are key outcomes of regenerative and organic farming systems, facilitating nutrient turnover and availability [7, 22, 30, 34, 35].

There is documented significant relationship between soil electrical conductivity and nutrient content, particularly nitrogen, phosphorus, and potassium. Increased EC often correlates with higher nutrient availability, though this relationship is modulated by soil texture, moisture, and organic matter [17, 36]. Some studies developed predictive models linking EC to nitrogen content have been developed for sensor applications, enabling real-time nutrient management [36]. However, confounding factors affecting EC-nutrient correlations recommend site-specific calibrations [36, 37].

2.6 Challenges and Future Directions

While sensor technologies offer significant benefits for regenerative agriculture, challenges such as high installation costs, data privacy concerns, and interoperability issues remain [3, 38]. Additionally, the transition to regenerative practices can be economically challenging due to initial costs and the need for supportive policies and frameworks. Palacios-López [38] emphasized that regenerative agriculture provides solutions for soil degradation but requires comprehensive technological support and policy interventions. Despite these challenges, the potential for sensor technologies to enhance regenerative agriculture is substantial, offering pathways to more resilient and sustainable farming systems [3].

2.7 Machine Learning Applications in Soil Classification

Recent advances in machine learning have enabled more sophisticated classification of soil properties and management regimes. Logistic regression and ensemble methods have been applied to discriminate between soil classes based on electromagnetic measurements [39]. The integration of multiple soil parameters improves classification accuracy compared to single-variable approaches. However, class imbalance between management regimes presents challenges requiring specialized evaluation metrics such as precision-recall analysis [40].

3 Materials and Methods

3.1 Study Sites and Data Collection

Sensor data were collected from two agricultural sites using the Soilscout API: Site 2398 (Monor, Hungary) representing conventional management, and Site 2598 (Krakovany, Slovakia) implementing regenerative practices. The final dataset comprised 174,412 records, with 161,481 observations from Site 2398 and 12,931 from Site 2598. The data collection period spanned from December 18, 2024 to December 18, 2025, providing a complete annual cycle of soil measurements.

The Soilscout sensor network measured five key soil parameters: electrical conductivity (dS/m), volumetric moisture content (m^3/m^3), soil temperature (°C), water balance index, and salinity (dS/m). Sensors were installed at consistent depths across both sites to ensure comparable measurements. The significant class imbalance (approximately 12:1 ratio) between sites reflects real-world deployment constraints and was addressed through density-normalized analysis and class-balanced machine learning approaches.

3.2 Statistical Analysis and Machine Learning Classification

Distribution characteristics were analysed using density-normalized histograms to enable direct comparison despite unequal sample sizes. For each soil parameter, we examined: central tendency and spread, distribution shape (unimodal vs. multimodal, symmetry vs. skewness), tail behaviour and presence of extreme values, and temporal

dynamics using 14-day rolling means. Statistical differences between sites were assessed through distribution overlap analysis and temporal pattern comparison.

For the classification task, a supervised machine learning approach based on logistic regression was employed to distinguish between the two monitored sites, corresponding to regenerative (site 2598) and conventional (site 2398) management regimes. To preserve temporal structure and reduce the risk of information leakage, the dataset was partitioned using a chronological holdout strategy. Specifically, a cutoff was defined at the 80th percentile of the timestamp distribution, such that the earliest 80% of observations were used for model training and the most recent 20% were reserved for testing. Input features consisted of all available numerical soil measurements. Missing values were handled using median imputation, and all variables were subsequently standardized to zero mean and unit variance to ensure comparability of coefficients and stable model convergence. To account for class imbalance, balanced class weights were applied during training. The model outputs probabilistic predictions of site membership, which were converted to binary labels using a decision threshold of 0.5. Performance was evaluated on the held-out test set using balanced accuracy, average precision (PR-AUC), and a full classification report including precision, recall, and F1-score, complemented by precision–recall curve analysis to capture performance across decision thresholds. This evaluation framework provides a robust assessment of model behaviour in an imbalanced setting while maintaining interpretability through coefficient-based feature importance analysis.

4 Results

4.1 Soil Electrical Conductivity

The distribution of soil electrical conductivity revealed striking differences between sites. Site 2398 exhibited a narrow, sharply peaked distribution with most values concentrated in a low conductivity range (approximately 0.05–0.25 dS/m), indicating low variability and few extreme conductivity events. This suggests a stable but uniformly low conductivity regime characteristic of chemically inert or depleted soil systems.

In contrast, Site 2598 (regenerative) displayed a much broader and strongly right-skewed distribution with a long right tail and several distinct high-conductivity peaks extending beyond 2.0 dS/m. The multimodal appearance suggests heterogeneous soil states driven by management effects rather than random noise. This higher variability indicates active soil processes likely driven by organic matter decomposition, microbial activity, and variable nutrient release - consistent with regenerative soil management which often increases biological and chemical heterogeneity.

4.2 Soil Moisture Dynamics

Moisture distribution analysis showed that Site 2398 values were moderately concentrated between 0.08 and 0.22 m^3/m^3 with a unimodal distribution peaking around 0.07–0.08. The right tail decayed gradually, suggesting stable moisture retention without extreme saturation. Site 2598 exhibited a strongly right-skewed and bimodal distribution with a pronounced spike near zero moisture indicating frequent near-surface drying

events, and a secondary mode around 0.25–0.35 extending well beyond Site 2398's upper range.

Temporal analysis (Fig. 1) revealed that Site 2398 displayed high stability in early months (moisture near 0.15–0.20) followed by a sharp decline in early summer (June–July) to approximately 0.11, continuing a slow downward drift to ~0.10 by autumn. Site 2598, after initial recording, showed rapid moisture increase from ~0.06 to ~0.13 with strong oscillations throughout late summer and autumn (peaks up to 0.18–0.19, troughs to ~0.08), with notably faster recovery after drops compared to Site 2398. This indicates improved infiltration, stronger rainfall/irrigation responsiveness, and heterogeneous soil structure induced by regenerative practices.

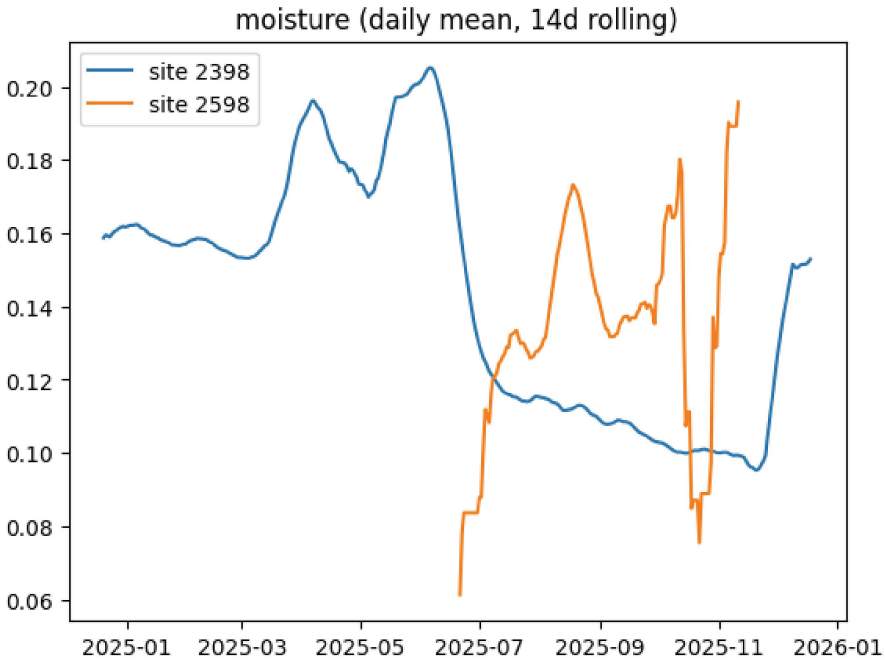


Fig. 1. Temporal evolution of moisture (daily mean, 14-day rolling)

4.3 Temperature and Water Balance

Temperature distributions showed Site 2398 with a broad (Fig. 2), relatively flat distribution spanning 2 °C to 25 °C with no sharp dominant peak, indicating prolonged cold exposure and slow thermal transitions. Site 2598 displayed a strongly peaked distribution centered around 22–24 °C with few low-temperature observations, suggesting biologically optimal thermal conditions during the measurement period.

Water balance analysis revealed Site 2398 with high values during winter/spring (0.35–0.65), sharp decline in early summer to ~0.15, and gradual decrease to ~0.05 by autumn showing strong seasonal depletion with slow recovery. Site 2598 initially

showed very low or negative values (~ − 0.1) indicating transient water deficit, followed by rapid improvement to 0.30–0.55, with frequent sharp peaks and troughs but consistent rebounds to high levels (0.6–0.7), demonstrating greater soil water buffering capacity.

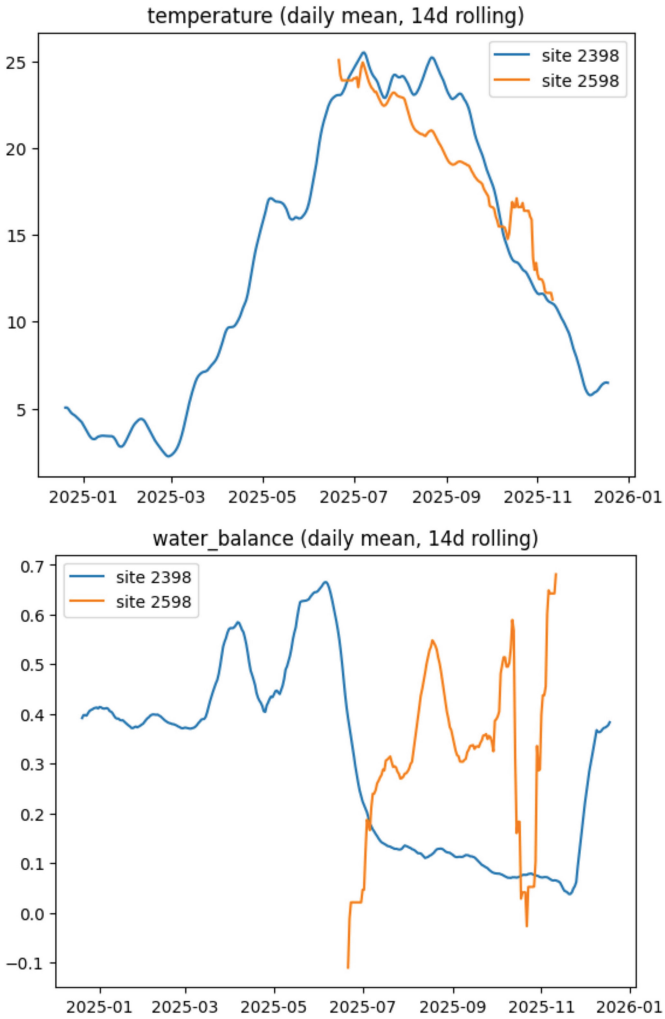


Fig. 2. Soil temperature over time and water balance over time

4.4 Salinity Patterns

Site 2398 salinity values were tightly clustered at very low levels (0.3–0.7 dS/m) with a narrow, unimodal distribution and no pronounced right tail, indicating consistently low salt accumulation. Site 2598 showed a much broader, strongly right-skewed distribution

with a long right tail extending beyond 10 dS/m and multiple secondary modes. The temporal evolution (Fig. 3) showed Site 2398 maintaining remarkably stable salinity (0.35–0.65) throughout the year, while Site 2598 exhibited a sharp increase after measurement onset (0.8 to 3.3), followed by prolonged elevated salinity with gradual decline toward 1.5. Summary of soil parameter distributions by site can see on the Table 1.

Importantly, the time-series showed partial recovery rather than monotonic accumulation, with high salinity coinciding temporally with high water balance and increased conductivity. This suggests transient nutrient pulses from mineralization of organic amendments and nutrient mobilization rather than chronic salinization.

Table 1. Summary of soil parameter distributions by site.

Parameter	Site 2398 (Conv.)	Site 2598 (Regen.)
Conductivity	Narrow, 0.05–0.25	Broad, skewed, >2.0
Moisture	Unimodal, 0.08–0.22	Bimodal, high variability
Temperature	Flat, 2–25 °C range	Peaked at 22–24 °C
Water Balance	Seasonal depletion	High variability, fast recovery
Salinity	Stable, 0.3–0.7	Dynamic, transient pulses

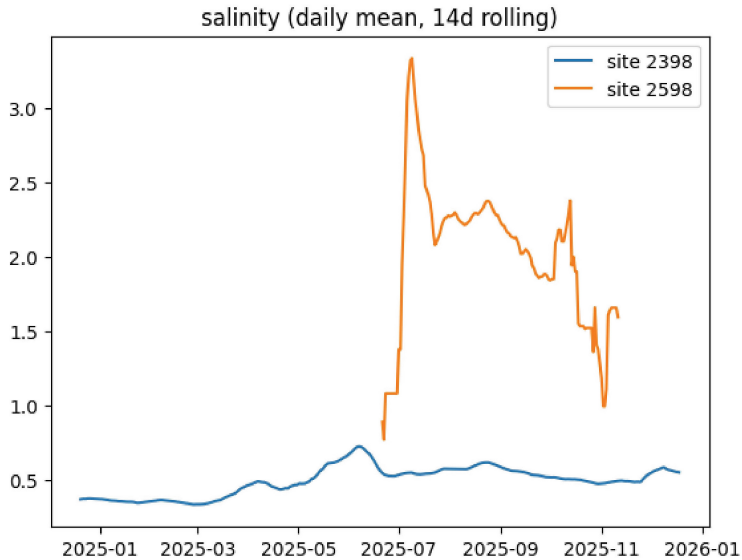


Fig. 3. Temporal evolution of salinity

4.5 Machine Learning Classification Results

The Precision-Recall curve evaluation (Fig. 4) demonstrated that the logistic regression classifier achieved very high precision at low recall thresholds. For recall values below approximately 0.35–0.40, precision remained close to 1.0, indicating that when the model predicted a sample as belonging to the regenerative site, it was almost always correct. However, as recall increased beyond ~0.4, precision decreased rapidly, falling below 0.5 and approaching zero at high recall levels.

This behavior implies that the regenerative site exhibits distinct soil states that are strongly separable, but these states do not persist continuously over time. When soil conditions strongly reflect regenerative management, the model identifies them with high confidence. During periods when soil conditions converge (similar moisture or salinity states), separability decreases. This is a desirable and realistic outcome in environmental systems reflecting the inherently dynamic nature of soil processes. Summary of soil parameter distributions by site can be seen on Table 2.

Table 2. Summary of soil parameter distributions by site.

Feature	Coefficient	Interpretation
Conductivity	+17.70	Strongest discriminator for regenerative
Salinity	−8.92	High salinity decreases regen. probability
Field Capacity	−8.69	Higher FC more common at conventional
Dielectricity	−6.97	Lower values characterize regenerative
Water Balance	+3.26	Higher availability increases regen. prob.
Temperature	+1.74	Slightly warmer soils at regenerative
Moisture	−0.05	Minimal independent contribution

4.6 Feature Importance Analysis

The extremely high positive coefficient for conductivity (+17.70) indicates that elevated soil conductivity is the most consistent marker of the regenerative site. This likely reflects increased ionic mobility due to organic amendments, enhanced microbial mineralization, and improved soil structure facilitating nutrient release. Importantly, this conductivity signal is transient rather than permanent salinization.

Despite higher conductivity, salinity contributes negatively (−8.92), indicating that the regenerative site is characterized by higher nutrient flux without chronic salt accumulation. This distinction is crucial: the model differentiates beneficial conductivity from harmful salinity. Higher field capacity being associated with the conventional site suggests a more static water-holding regime, potentially indicating compacted or fine-textured soils with lower dynamic infiltration.

Lower dielectric readings at the regenerative site may reflect improved soil aggregation and altered pore-water relationships. Higher water balance increases regenerative

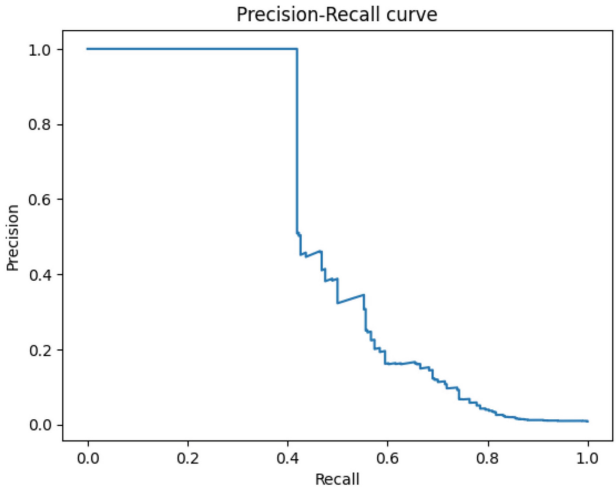


Fig. 4. Precision-Recall curve

classification probability, supporting observations of rapid recovery from water deficits. The negligible moisture coefficient (−0.05) indicates that once water balance, conductivity, and salinity are included, raw moisture adds little independent information, suggesting strong multicollinearity and reinforcing the importance of derived variables.

5 Discussion

5.1 Comparison with Literature

Our findings align with the broader literature demonstrating that regenerative systems generally exhibit different soil electrical conductivity patterns compared to conventional systems [6, 11, 13]. The higher variability observed at Site 2598 is consistent with studies showing that organic amendments and regenerative practices increase biological and chemical heterogeneity in soils [6, 26, 27]. The temporal stability patterns we observed support the use of ECa measurements as reliable indicators over multiple seasons, as documented by Fletcher [13] and Kai-hua et al. [20].

The strong positive association between conductivity and regenerative classification confirms findings from multiple studies linking organic amendments to enhanced ionic mobility and microbial activity [22, 23, 34]. Our observation that high conductivity accompanies rather than contradicts reduced harmful salinity supports the distinction between nutrient-related conductivity and salinization emphasized in the literature [6, 8]. The rapid moisture recovery observed at the regenerative site aligns with documented improvements in infiltration and water retention under regenerative management [6, 12, 30].

5.2 Practical Implications

The demonstrated temporal stability and discriminative power of soil electrical conductivity measurements validate their practical use in precision agriculture for delineating

management zones, optimizing sampling strategies, and guiding site-specific interventions [13, 16, 41]. The ability to use EC and dielectric property measurements as proxies for management regime identification offers a cost-effective and rapid assessment tool for farmers and land managers [17, 23, 36].

The high precision achieved at low recall thresholds suggests that IoT-based monitoring can reliably identify periods when soil conditions strongly reflect regenerative management, even if such states do not persist continuously. This has practical implications for adaptive management: farmers can monitor when regenerative practices are achieving measurable soil health improvements and adjust interventions accordingly.

The finding that the regenerative site exhibits higher variability should be interpreted positively rather than as instability. In environmental monitoring, higher soil 'reactivity' - the capacity to respond to inputs and recover from perturbations - is a desirable property under increasing climate variability. This supports the interpretation that regenerative management promotes a more responsive and biologically active soil environment.

5.3 Limitations

Several limitations should be considered when interpreting these results. The significant class imbalance (12:1 ratio) reflects deployment constraints but may affect generalizability. Geographic differences between sites (Hungary vs. Slovakia) may confound management effects with climate and soil type differences. The one-year monitoring period, while providing a complete annual cycle, may not capture longer-term trends in soil property development under regenerative management.

Sensor deployment began mid-year at Site 2598, limiting winter and spring comparisons. The logistic regression model, while interpretable, may not capture complex nonlinear interactions between soil parameters. Future studies should consider multi-site replication with matched soil types and longer monitoring periods.

5.4 Future Research Directions

The literature synthesis identifies several high-priority research gaps that our findings partially address but also highlight for future work [42]. Standardization of sensor calibration protocols across regions and soil types remains critical [14, 28, 43]. Integration of dielectric properties with nutrient dynamics requires controlled experiments to quantify relationships between dielectric permittivity and nutrient availability [31, 32]. Long-term, multi-seasonal studies assessing temporal stability of EC and dielectric properties under contrasting management are needed [13, 20, 21].

Development of soil texture- and moisture-specific models linking EC to nutrient concentrations would improve predictive accuracy [21, 36, 37]. Rapid field assessment methods combining sensor-based proxies with biological assays would enable routine monitoring of microbial activity [23, 34, 35]. Advanced geostatistical and machine learning approaches for multi-sensor data fusion represent promising directions for robust site-specific management [31, 32, 41, 45].

6 Conclusion

This study demonstrated that IoT-based soil sensor data combined with machine learning classification can effectively discriminate between regenerative and conventional agricultural management regimes. Analysis of 174,412 sensor records from two sites revealed systematic differences in soil electrical conductivity, moisture dynamics, temperature patterns, water balance, and salinity between management systems.

Soil electrical conductivity emerged as the dominant discriminating feature, with the regenerative site exhibiting higher variability and transient conductivity pulses reflecting enhanced microbial activity and nutrient cycling rather than chronic salinization. The logistic regression classifier achieved high precision when soil conditions strongly reflected management differences, while acknowledging that soil states converge during certain periods - a realistic characteristic of dynamic environmental systems.

Key findings support the integration of electrical and dielectric measurements with biological indicators for comprehensive soil health assessment. The observed increases in conductivity at the regenerative site, not accompanied by sustained salinity accumulation, suggest enhanced nutrient mobility and turnover consistent with improved soil biological function. Higher moisture variability and faster recovery at the regenerative site indicate improved infiltration and water retention capacity.

These results have practical implications for precision agriculture, supporting the use of continuous IoT monitoring for real-time assessment of soil health trajectories under different management regimes. The approach provides an objective, data-driven framework for evaluating regenerative agriculture outcomes and guiding adaptive management decisions. Future work should focus on multi-site replication, longer-term monitoring, and integration of additional biological indicators to further validate and extend these findings.

Acknowledgement. This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and of Slovak Academy of Sciences (SAS) under the contract No. VEGA 1/0492/26.

This work was supported by the Research Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic under Grant 09I05-03-V02-00022: Optimization of manufacturing and transportation infrastructure processes through artificial intelligence methods (OPTIMUM).

References

1. LaCanne, C.E., Lundgren, J.G.: Regenerative agriculture: merging farming and natural resource conservation profitably. PeerJ. **6**, e4428 (2018)
2. Schreefel, L., et al.: Regenerative agriculture—the soil is the base. Glob. Food Sec. **26**, 100404 (2020)
3. Kalaimathi, V., Raju, M., Krishnan, R., et al.: Regenerative agriculture for sustainable crop productivity: a comprehensive review. Plant Sci. Today. **12** (2025). https://doi.org/10.14719/pst.6938
4. Jones, M., Figueiredo, G.K.D.A., Howson, S., et al.: Monitoring and mapping a decade of regenerative agricultural practices across the contiguous United States. Land. **13**, 2246 (2024). https://doi.org/10.3390/land13122246

5. Geetha, R., Rakhi, R., Anith, K.N.: Innovative sensor technologies in agriculture—applications and challenges: a review. Agric. Sci. Dig. (2025). https://doi.org/10.18805/ag.d-6344
6. Cromwell, O.: Trend setting impacts of organic matter on soil physico-chemical properties in traditional vis-a-vis chemical-based amendment practices. PLOS Sustain. Transform. **1**(3), e0000007 (2022)
7. Velmourougane, K.: Impact of organic and conventional systems of coffee farming on soil properties and culturable microbial diversity. Scientifica. **2016**, 3604026 (2016)
8. Maucieri, C., et al.: Organic versus conventional farming: medium-term evaluation of soil chemical properties. Ital. J. Agron. **17**(3), 2114 (2022)
9. Yao, R., et al.: Digital mapping of soil salinity and crop yield across a coastal agricultural landscape using repeated electromagnetic induction surveys. PLoS One. **11**(5) (2016)
10. Torrús-Castillo, M., et al.: Assessing the impact of six years of regenerative agriculture on a commercial olive orchard in southern Spain. EGU25-422 (2025)
11. Thapa, V.R., et al.: Land use effects on soil health in semiarid drylands. Agric. Environ. Lett. **3**(1) (2018)
12. Mosier, S.L., et al.: Improvements in soil properties under adaptive multi-paddock grazing relative to conventional grazing. Agron. J. **114**(4), 2584–2597 (2022)
13. Fletcher, R.S.: Temporal comparisons of apparent electrical conductivity: a case study on clay and loam soils in Mississippi. Agric. Sci. **13**(8), 936–946 (2022)
14. Allred, B.J., et al.: Comparison of electromagnetic induction, capacitively-coupled resistivity, and galvanic contact resistivity methods. Appl. Eng. Agric. **22**(2), 215–230 (2006)
15. Husson, O., et al.: Conservation agriculture systems alter the electrical characteristics of four soil types in France. Soil Tillage Res. **176**, 57–68 (2018)
16. Johnson, C.K., et al.: Field-scale electrical conductivity mapping for delineating soil condition. Soil Sci. Soc. Am. J. **65**(6), 1829–1837 (2001)
17. Wayayok, A., et al.: Bulk soil electrical conductivity as an estimator of nutrients in maize cultivated land (2009)
18. Martinez, G., et al.: Can apparent electrical conductivity improve spatial characterization of soil organic carbon. Vadose Zone J. **8**(3), 586–593 (2009)
19. Ristolainen, A., et al.: Prediction of soil properties with field geo-electrical probes. Comm. Soil Sci. Plant Anal. **40**, 555–565 (2009)
20. Kai-hua, L., et al.: Temporal stability of apparent soil electrical conductivity measured by EMI techniques. J. Mt. Sci. **11**(1), 98–109 (2014)
21. Molin, J.P., Faulin, G.D.C.: Spatial and temporal variability of soil electrical conductivity related to soil moisture. Sci. Agric. **70**(1), 1–5 (2013)
22. Jat, H.S., et al.: Temporal changes in soil microbial properties and nutrient dynamics under climate smart agriculture practices. Soil Tillage Res. **199**, 104595 (2020)
23. Xu, N., et al.: Interaction of soil health indicators to different regenerative farming practices on mineral soils. Agrosyst. Geosci. Environ. **5**(1) (2022)
24. Rudd, J.D., Roberson, G.T., Classen, J.J.: Application of satellite, unmanned aircraft system, and ground-based sensor data for precision agriculture: a review. ASABE (2017). https://doi.org/10.13031/AIM.201700272
25. Ranwa, M., Sri, K., Khare, A., et al.: A summative review of advances in sensor technology for precision agriculture. Arch. Curr. Res. Int. **24**, 257–275 (2024). https://doi.org/10.9734/acri/2024/v24i10929
26. Kumar, S., et al.: Impact of in-situ green manuring on soil properties and fertility under regenerative farming. Int. J. Res. Agron. **7**(5) (2024)
27. UdhayaNandhini, D., et al.: Evaluating growth attributes and soil health under long term implementation of organic farming in fodder cowpea. Range Manage. Agrofor. **44**, 329–335 (2023)

28. Altdorff, D., et al.: Comparison of multi-frequency and multi-coil electromagnetic induction for mapping properties in shallow podsolic soils. Sensors. **20**(8) (2020)
29. Sadatcharam, K.: Assessing potential applications of multi-coil and multi-frequency EMI sensors for agricultural soils in western Newfoundland. Masters Theses, Memorial University of Newfoundland (2019)
30. Mankar, E.P., et al.: Effect of organically grown cropping systems on soil properties and their correlation with carbon pools. Int. J. Plant Soil Sci. **37**(1), 472–489 (2025)
31. Castrignanò, A., et al.: A geostatistical sensor data fusion approach for delineating homogeneous management zones. Catena. **167**, 293–304 (2018)
32. Pathirana, S., et al.: Ground-penetrating radar and electromagnetic induction: challenges and opportunities in agriculture. Remote Sens. **15**(11), 2932 (2023)
33. Scudiero, E., et al.: Validation of sensor-directed spatial simulated annealing soil sampling strategy. J. Environ. Qual. **45**(4), 1226–1233 (2016)
34. Gunapala, N., Scow, K.M.: Dynamics of soil microbial biomass and activity in conventional and organic farming systems. Soil Biol. Biochem. **30**(6), 805–816 (1998)
35. Liao, J., et al.: Organic farming improves soil microbial abundance and diversity under greenhouse condition. Sustainability. **10**(10) (2018)
36. Mirzakhaninafchi, H., et al.: Development of prediction models for soil nitrogen management based on electrical conductivity and moisture content. Sensors. **22**(18), 6728 (2022)
37. Simon, M., et al.: Relación entre la conductividad eléctrica aparente con propiedades del suelo y nutrientes. Cienc. Suelo. **31**(1), 45–55 (2013)
38. Palacios-López, L.A.: La agricultura regenerativa como solución para la degradación del suelo a través de investigaciones recientes. Horizon Nexus J. **3**, 18–30 (2025). https://doi.org/10.70881/hnj/v3/n1/46
39. Corwin, D.L., Scudiero, E.: Field-scale apparent soil electrical conductivity. In: Methods of Soil Analysis. Soil Science Society of America (2016)
40. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves. In: ICML 2006, pp. 233–240 (2006)
41. Ylagan, S., et al.: Using apparent electrical conductivity to delineate field variation in an agroforestry system. Remote Sens. **14**(22), 5777 (2022)
42. Heil, K., Schmidhalter, U.: The application of EM38: determination of soil parameters, selection of soil sampling points. Sensors. **17**(11) (2017)
43. Sassenrath, G.F., Kulesza, S.: Measuring soil electrical conductivity to delineate zones of variability in production fields. Kans. Agric. Exp. Stn. Res. Rep. **3**(2) (2017)
44. Kárász, C., Kopják, J.: Comparative study on plant type specific LED light source design parameters. Acta Polytech. Hung. **17**(3), 71–90 (2020)
45. Kovari, A.: Industry 5.0: generalized definition, key applications, opportunities and threats. Acta Polytech. Hung. **21**(3), 267–284 (2024). https://doi.org/10.12700/APH.21.3.2024.3.17

Analysis of Security Risks Associated with Cookies in University Information Systems

Jan Francisti[1](✉), Marián Hlavačka[1], Azeta Tartaraj[2], and Zoltán Balogh[1,3]

[1] Faculty of Natural Sciences, Constantine the Philosopher University in Nitra, Nitra, Slovakia
jfrancisti@ukf.sk
[2] Faculty of Business, Aleksandër Moisiu University, Durrës, Albania
[3] Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest, Hungary

Abstract. Cookies represent a fundamental mechanism of modern web applications, enabling session management, personalization, and efficient communication between web browsers and servers. Despite their functional benefits, improper configuration and handling of cookies may introduce significant security and privacy risks. This article examines the security implications of cookie usage in university information systems, with a focus on vulnerabilities related to session management and request validation. The study combines theoretical analysis with controlled laboratory simulations of attack scenarios performed on selected publicly accessible university web domains. Techniques such as session hijacking and Cross-Site Request Forgery were evaluated in relation to cookie attributes including Secure, HttpOnly, and SameSite. The results indicate that insufficient protection of session identifiers and inconsistent configuration of cookie-related security attributes may significantly increase the attack surface of web applications. Differences in security posture between the analyzed systems were identified, particularly in the implementation of cookie protection mechanisms and session validation strategies. Based on the findings, the article highlights the importance of robust server-side validation and consistent security configuration to mitigate cookie-based threats while ensuring compliance with privacy protection standards such as the General Data Protection Regulation (GDPR).

Keywords: Cookies · Ethical Hacking · Cross-Site Scripting · Cross-Site Request Forgery · Session Hijacking

1 Introduction

Cookies are an integral component of modern web applications, enabling data storage and communication between web browsers and servers. They are widely used for session management, user authentication, personalization, and tracking of user interactions, thereby contributing to the efficient operation of web services. However, the extensive use of cookies also raises significant concerns regarding user privacy and data security, which have become increasingly important in the current digital environment.

Cookie-related security issues are closely associated with a growing number of cyber-attacks, including Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF), and

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 538–547, 2027.
https://doi.org/10.1007/978-3-032-32104-6_45

session hijacking. These attacks exploit weaknesses in cookie handling mechanisms to gain unauthorized access to user data or impersonate legitimate users [1, 2]. In addition to technical threats, the collection and processing of data through cookies introduce legal and ethical challenges, particularly in relation to personal data protection and compliance with regulatory frameworks such as the General Data Protection Regulation (GDPR) [3].

Despite the existence of numerous studies addressing web security and privacy, many of them focus either on theoretical aspects of cookie mechanisms or on isolated security vulnerabilities. Fewer works provide a combined practical analysis that simultaneously examines privacy implications, real-world attack scenarios, and mitigation techniques in a controlled experimental environment [4, 5]. As a result, there remains a gap between theoretical recommendations and their practical validation in realistic testing conditions.

The aim of this article is to examine the impact of cookies on user privacy and to analyze the security risks associated with their use. The study focuses on practical testing of selected attack scenarios in a controlled virtual environment using VirtualBox and the Kali Linux operating system. This approach enables the analysis of different types of cookies, their behavior, and their susceptibility to common attack techniques.

Furthermore, the article provides an overview of potential mitigation strategies for the identified risks, including the restriction of third-party cookies, the application of security-related HTTP headers, and the use of protective mechanisms such as the Same-Site cookie attribute [6, 7]. The article is structured as follows: Sect. 2 presents the theoretical background and related work, Sect. 3 describes the applied methodology and experimental setup, Sect. 4 discusses the experimental results, and Sect. 5 concludes the article with recommendations for future research.

2 Related Work

2.1 Cookie Communication and Server-Side Storage

Cookies are used to store information that enables web applications to maintain user state and improve overall user experience. When a user visits a website, the server processes the incoming request and may assign the user a unique session identifier, which is then stored within a cookie on the client side. While the cookie itself is stored in the user's web browser, the associated session data are typically maintained on the server. Each cookie is bound to a specific domain, and only the domain that originally created the cookie is permitted to access it, which provides a basic level of isolation between websites [1, 8].

Cookie communication is implemented through HTTP headers exchanged between the web browser and the web server [9]. Upon receiving a user request, the server may include the Set-Cookie directive in the HTTP response header, instructing the browser to store the specified cookies. In subsequent requests, the browser automatically attaches the relevant cookies for the given domain using the HTTP Cookie header. This mechanism enables session continuity but also represents a potential attack vector if cookies are not properly protected [10, 11].

The security of cookie communication can be enhanced by applying specific cookie attributes. The HttpOnly attribute restricts access to cookies from client-side scripts, thereby reducing the risk of Cross-Site Scripting (XSS) attacks. To prevent cookie interception during transmission, the Secure attribute ensures that cookies are sent exclusively

over encrypted HTTPS connections [5, 12]. The SameSite attribute provides additional protection against CSRF attacks by controlling when cookies are sent with cross-site requests [7].

2.2 Cross-Site Scripting

The number of users accessing websites continues to grow on a daily basis, thereby creating increased opportunities for attackers. One of the most frequently exploited vulnerabilities is Cross-Site Scripting (XSS) [13]. XSS is a client-side code injection attack in which malicious scripts are executed with the aim of stealing user data from legitimate websites or web applications [12]. This vulnerability most commonly occurs when an application includes unvalidated or improperly encoded user input in the output it generates. In such cases, users are often unaware that their data are being redirected to an attacker-controlled website.

Once malicious code is injected into a webpage and executed, it may result in several negative consequences, including session hijacking and unauthorized access to sensitive data stored in cookies. The success of such attacks typically depends on persuading users to interact with content containing malicious scripts, for example by clicking on a crafted link [14, 15]. When combined with social engineering techniques, these links may be distributed via emails that appear legitimate and encourage victims to engage with them.

Since JavaScript is a fundamental technology in the development of dynamic web applications, XSS attacks often exploit vulnerabilities in dynamic application components. This allows attackers to abuse JavaScript to compromise security, even in cases where the webpage itself does not explicitly contain embedded JavaScript code [16, 17]. Modern frameworks and Content Security Policy (CSP) implementations provide mitigation mechanisms, but proper configuration remains essential for effective protection [18].

2.3 Cross-Site Request Forgery

Cross-Site Request Forgery (CSRF) represents a significant threat to web applications due to the stateless nature of the HTTP protocol, in which a malicious website causes a user's browser to send unauthorized requests to a trusted website [19]. In a CSRF attack, the malicious request is crafted to appear as if it originates from a legitimate user, thereby exploiting the trust relationship between the user and the target web application.

To successfully execute a CSRF attack, three elements are required: a victim, a trusted website, and a malicious website. A commonly cited example of a CSRF attack involves online banking. If the victim has an active authenticated session, an attacker may create a malicious website containing a hidden request, for instance in the form of an HTML form or an image tag, that targets the legitimate banking website [20]. When the victim visits the malicious website, typically by clicking a link embedded in an email, the browser unknowingly sends the request to the legitimate server where the user is already authenticated.

Following a successful CSRF attack, an attacker may perform additional malicious actions, including changing authentication credentials, placing orders without user consent, or modifying application and website settings with the intent of stealing sensitive data [19, 21]. The SameSite cookie attribute and anti-CSRF tokens represent primary defense mechanisms against these attacks [7].

2.4 Session Hijacking

Session hijacking is an attack in which an attacker gains control over a user's session within a web application. The attacker's objective is to obtain the session identifier that the web application assigns to each user for identification purposes [22]. Since HTTP is a stateless protocol, web applications rely on session cookies to maintain user state and associate requests with authenticated users. These cookies may store or reference sensitive information, such as user identifiers or email addresses.

Once an attacker obtains a valid session identifier, they can impersonate the user and perform actions as a legitimately authenticated user [23, 24]. However, if cookies are transmitted over the internet without adequate security mechanisms, they may be intercepted or otherwise acquired by an adversary using various attack techniques or by exploiting vulnerabilities in the web browser. Network-level attacks such as Man-in-the-Middle (MITM) represent common vectors for session identifier theft [25].

2.5 Penetration Testing Tools

Kali Linux is an operating system widely used in the field of penetration testing, as it includes hundreds of preinstalled tools designed for testing networks, web applications, Internet of Things devices, operating systems, and related technologies [26]. Among the most commonly used tools are Wireshark, Nmap, BeEF, Burp Suite, and Aircrack-ng. These tools enable developers and penetration testers to analyze, test, and strengthen applications and systems both before and after deployment [27, 28].

The Browser Exploitation Framework (BeEF) is a penetration testing tool primarily focused on web browsers. Its original purpose was to demonstrate vulnerabilities related to Cross-Site Scripting (XSS); however, over time it has evolved to include a wide range of practical client-side attack techniques [29]. BeEF is typically utilized when a victim visits a malicious website containing the JavaScript code hook.js, which is automatically generated during the initial startup of the framework.

Burp Suite is a web application penetration testing tool used to identify vulnerabilities in communication between the client and the server [30]. It is developed by PortSwigger and is among the most widely used tools in the field of web security. Burp Suite operates by intercepting and analyzing HTTP and HTTPS requests through a proxy server, enabling the monitoring and manipulation of communication between a web browser and a web application [31].

3 Materials and Methods

3.1 Experimental Environment

The security analysis was conducted in a controlled laboratory environment using virtualization technology. The testing infrastructure consisted of VirtualBox virtual machines running Kali Linux version 2024.4. This setup provided an isolated environment for security testing without affecting production systems. The analysis focused on two university web domains: UKF Webmail and UKF AiS (Academic Information System), both of which are accessed daily by faculty members and students.

The web domains were thoroughly analyzed with respect to cookie usage, and attack scenarios targeting potential vulnerabilities were simulated. All testing activities and simulated attacks were conducted ethically and with explicit permission from the university, with the sole purpose of identifying and mitigating potential security risks.

For vulnerability testing, several specialized tools were employed. BeEF-XSS version 0.5.4.0 was used for browser exploitation framework testing, while Burp Suite version 2024.9.4 served as the primary tool for HTTP/HTTPS traffic interception and analysis. The Cookie-Editor browser extension was utilized for cookie inspection and manipulation. All testing was performed exclusively for research purposes in accordance with ethical guidelines.

3.2 Testing Methodology

The testing methodology followed a systematic approach beginning with cookie attribute analysis. For each target domain, cookies were exported using the Cookie-Editor browser extension and analyzed for the presence and configuration of security attributes including Secure, HttpOnly, and SameSite. This initial assessment provided baseline information about the security posture of each application.

Subsequently, XSS injection attempts were performed using the BeEF framework's hook.js script. Injection points included message submission fields, search inputs, and URL parameters. The effectiveness of XSS protections was evaluated based on whether injected scripts were executed or blocked by the application.

For session hijacking testing, Burp Suite was configured as a proxy server using IP address 10.0.2.15 and port 8080. The Firefox browser on the test machine was configured to route all network traffic through the proxy, enabling controlled Man-in-the-Middle (MITM) testing. A Bash script (firefox_script_ca.sh) was created to automate the import of the Burp certificate and proxy configuration.

CSRF vulnerability testing focused on applications with absent or improperly configured SameSite attributes. Test scenarios included logout redirection attacks and message submission functionality reproduction through request replay in Burp Suite's Repeater module.

4 Results

4.1 UKF Webmail Analysis

The analysis of the UKF Webmail system revealed several security characteristics. The Secure and HttpOnly attributes were correctly configured, providing basic protection against certain types of attacks. However, the SameSite attribute was set to a null value, representing a potential vulnerability to Cross-Site Request Forgery attacks. Furthermore, the exported cookies contained session identifiers, introducing potential risk of session hijacking.

XSS injection attempts using the hook.js script were unsuccessful. Testing was performed through message submission fields, message bookmarking, and search input fields. Analysis of HTTP response headers using the curl -I command revealed that the X-XSS-Protection header was enabled. Although deprecated, its presence indicated additional security controls. The observed behavior suggested the implementation of Content Security Policy (CSP), a server-side security mechanism designed to mitigate XSS attacks [18].

Session hijacking testing using Burp Suite enabled interception of HTTP requests and extraction of session identifiers stored in cookies. The captured session ID (roundcube_sessid) was inserted into the cookies of an unauthenticated Webmail page. After refreshing the page, access to an active authenticated session was obtained, demonstrating successful session hijacking. The session was accessed from a different network without server-side verification of the originating IP address.

CSRF vulnerability testing exploited the absence of the SameSite attribute. A logout redirection attack using window.location.href was successfully executed, indicating a CSRF vulnerability. Subsequently, message submission functionality was reproduced by replaying the required sequence of GET and POST requests in Burp Suite. A total of ten requests were identified and replayed sequentially with modified message content. The server processed the modified requests without verifying request origin or integrity.

All requests involved in the message-sending process were transferred to the Repeater module of Burp Suite. A total of ten requests were identified and replayed sequentially. In the primary POST request containing the form submission, the message subject and content were modified. The requests were then executed in the same order as originally observed.

This attack scenario, in which an authenticated student session was impersonated to send messages, was performed three times. In each case, the server responded with a confirmation of successful message delivery and did not verify the origin of the requests (Fig. 1).

4.2 UKF AiS Analysis

The UKF AiS system demonstrated stronger security configuration compared to the Webmail system. Cookie attribute analysis showed that HttpOnly and Secure were set to true, and the SameSite attribute was configured with the value Strict. JavaScript injection attempts using the BeEF-XSS tool were unsuccessful due to sufficient security protections implemented by the web portal.

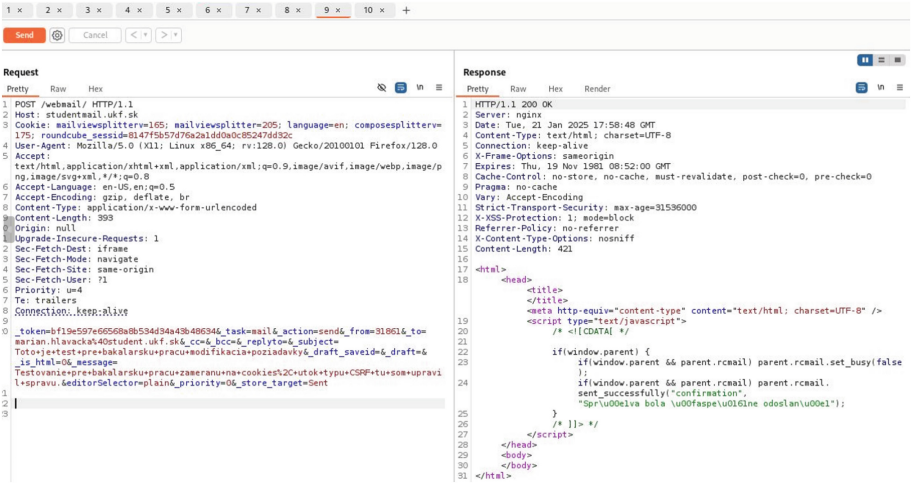


Fig. 1. Sending Requests to the Server Using the Burp Suite Tool

Despite the stronger cookie configuration, session hijacking testing revealed vulnerabilities in session validation. Using Burp Suite, session identifiers were intercepted from HTTP requests. The captured session ID, which contained a server identifier (server13), was inserted into the cookies of an unauthenticated page where a different server identifier (server12) was originally present. After refreshing the webpage, access to the authenticated session was obtained.

The server did not verify the origin IP address when the session was accessed from a different network environment. This behavior demonstrated that session continuity relied solely on the validity of the session identifier, representing a potential security weakness despite the proper configuration of cookie attributes (Table 1).

Table 1. Comparison of security attributes and vulnerability assessment

Attribute/Test	UKF Webmail	UKF AiS
Secure attribute	Enabled	Enabled
HttpOnly attribute	Enabled	Enabled
SameSite attribute	Null (vulnerable)	Strict
XSS protection	Effective (CSP)	Effective
Session hijacking	Vulnerable	Vulnerable
CSRF protection	Vulnerable	Protected
IP validation	Not implemented	Not implemented

4.3 Comparative Analysis

The conducted testing revealed significant differences in the security posture of the analyzed university systems with respect to cookie handling and session management. Table 1 summarizes the key findings from both systems. The UKF Webmail system exhibited multiple weaknesses, particularly in relation to the absence of the SameSite attribute and insufficient verification of session origin. Although protection mechanisms against XSS attacks were observed, the ability to reuse intercepted session identifiers demonstrated vulnerabilities in session validation processes.

In contrast, the UKF AiS system showed stronger configuration of cookie-related attributes, including proper implementation of Secure, HttpOnly, and SameSite. Nevertheless, similar session hijacking behavior was observed when session identifiers were reused without additional contextual verification. The results indicate that while modern web applications may implement several client-side and header-based protections, session management remains a critical area of security risk.

Systems relying solely on session identifiers without supplementary validation mechanisms may be susceptible to unauthorized session reuse. These findings highlight the importance of comprehensive session protection strategies and consistent configuration of cookie security attributes across institutional web platforms. Additional measures such as IP address binding, session fingerprinting, and regular session rotation should be considered to strengthen security posture [22, 24].

5 Conclusion

This article focused on the analysis of security risks associated with cookie handling in university information systems, with the objective of identifying potential threats arising from data stored in cookies and evaluating their impact on users. Through controlled security testing and simulated attack scenarios, it was demonstrated that web applications may still contain significant vulnerabilities related to session management and cookie configuration.

The results showed that techniques such as session hijacking and Cross-Site Request Forgery can be successfully executed when security mechanisms are insufficiently implemented. Weak configurations, including improperly protected session identifiers or the absence of the SameSite attribute, increase the potential attack surface and may lead to unauthorized access or compromise of application integrity.

The findings highlight the importance of consistent configuration of cookie security attributes and robust server-side validation mechanisms. Strengthening session protection strategies and enforcing stricter request validation can significantly reduce the risk of cookie-based attacks. Recommended mitigation measures include: (1) implementation of the SameSite attribute with Strict or Lax values; (2) binding sessions to client IP addresses or browser fingerprints; (3) implementation of session timeout and rotation mechanisms; and (4) deployment of comprehensive Content Security Policy headers.

Additionally, it is important to acknowledge that HTTPS protects the confidentiality and integrity of data in transit, but it does not fully eliminate threats originating from the client side. For example, malicious or compromised browser extensions may theoretically access cookie data within the browser environment and transmit it to external

entities. Although such attack vectors were not experimentally examined in this study, they represent an important consideration for future research. This observation further emphasizes the need to extend security mechanisms beyond transport-layer protection to include client-side threat mitigation.

This study provides a professional perspective on security testing in the context of university web systems and emphasizes the need for continuous monitoring and improvement of institutional security standards. Future research may extend the analysis to more advanced vulnerability detection techniques, client-side attack vectors, and additional university web platforms in order to further enhance the protection of user data and academic information systems.

Acknowledgments. This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and of Slovak Academy of Sciences (SAS) under the contract No. VEGA 1/0492/26 as well as by the University Grant Agency under contract No. VII/1/2026.

References

1. Velagapudi, S.L., Gupta, H.: Privacy, security of cookies in HTTP transmission. In: International Conference on Information Systems and Computer Networks (ISCON), pp. 22–25. IEEE (2019)
2. Rasaii, A., Singh, S., Gosain, D., Gasser, O.: Exploring the cookieverse: a multi-perspective analysis of web cookies. In: Passive and Active Measurement, LNCS, vol. 13990. Springer, Cham (2023)
3. European Union: General Data Protection Regulation (GDPR). Regulation (EU) 2016/679 of the European Parliament and of the Council (2016)
4. Paul, P., et al.: Using browser cookies for event monitoring and user verification of an account. In: IEEE 9th Annual IEMCON, pp. 455–460. IEEE (2018)
5. Macia, N., Tinetti, F.G.: Improving security in web sessions: special management of cookies. In: Emerging Trends in ICT Security, pp. 481–491. Elsevier (2014)
6. Karayat, R., Jadhav, M., Kondaka, L.S., Nambiar, A.: Web application penetration testing & patch development using Kali Linux. In: ICACCS, pp. 1392–1397. IEEE (2022)
7. Calzavara, S., et al.: Testing the same-origin policy. In: IEEE Symposium on Security and Privacy (SP). IEEE (2017)
8. Barth, A.: HTTP state management mechanism. RFC 6265, IETF (2011)
9. Fielding, R., Reschke, J.: Hypertext transfer protocol (HTTP/1.1): semantics and content. RFC 7231, IETF (2014)
10. Chen, H.-C., Nshimiyimana, A., Damarjati, C., Chang, P.-H.: Detection and prevention of cross-site scripting attack with combined approaches. In: ICEIC, pp. 1–4. IEEE (2021)
11. Rodriguez, G., Benavides, E., Torres, J., Flores, P., Fuertes, W.: Cookie Scout: an analytic model for prevention of XSS using a cookie classifier. In: ICISSP, LNCS. Springer, Cham (2018)
12. Shahriar, H., Zulkernine, M.: Mitigating program security vulnerabilities: approaches and challenges. ACM Comput. Surv. **44**(3), 1–46 (2012)
13. OWASP Foundation: Cross Site Scripting (XSS). OWASP Web Security Testing Guide (2024)
14. Bukhari, S.N., Ahmad Dar, M., Iqbal, U.: Reducing attack surface corresponding to type 1 cross-site scripting attacks using SDLC practices. In: AEEICB, pp. 1–4. IEEE (2018)

15. Gupta, S., Gupta, B.B.: Cross-site scripting (XSS) attacks and defense mechanisms: classification and state-of-the-art. Int. J. Syst. Assur. Eng. Manag. **8**(1), 512–530 (2017)
16. Stock, B., Lekies, S., Mueller, T., Spiegel, P., Johns, M.: Precise client-side protection against DOM-based cross-site scripting. In: USENIX Security Symposium, pp. 655–670 (2014)
17. Parameshwaran, I., Buber, E., Rahat, A., Krishnan, P., Wang, B.: A practical approach for detecting code injection attacks. In: IEEE Symposium on Privacy-Aware Computing, pp. 35–46 (2017)
18. Stamm, S., Sterne, B., Markham, G.: Reining in the web with content security policy. In: WWW 2010, pp. 921–930. ACM (2010)
19. Siddiqui, M.S., Verma, D.: Cross site request forgery: a common web application weakness. In: ICCSN, pp. 538–543. IEEE (2011)
20. Rankothge, W.H., Randeniya, S.M.N.: Identification and mitigation tool for cross-site request forgery (CSRF). In: IEEE R10 HTC, pp. 1–5. IEEE (2020)
21. Barth, A., Jackson, C., Mitchell, J.C.: Robust defenses for cross-site request forgery. In: ACM Conference on Computer and Communications Security, pp. 75–88 (2008)
22. Fett, D., Küsters, R., Schmitz, G.: The web SSO standard OpenID connect: in-depth formal security analysis. In: IEEE EuroS&P. IEEE (2016)
23. Lacroix, K., Loo, Y.L., Choi, Y.B.: Cookies and sessions: a study of what they are, how they work and how they can be stolen. In: ICSSA, pp. 20–24. IEEE (2017)
24. Prapty, R.T., Azmin Md, S., Hossain, S., Narman, H.S.: Preventing session hijacking using encrypted one-time-cookies. In: WTS, pp. 1–6. IEEE (2020)
25. Conti, M., Dragoni, N., Lesyk, V.: A survey of man in the middle attacks. IEEE Commun Surv Tutor. **18**(3), 2027–2051 (2016)
26. Rao, V., Prema, K.V.: Performing real-time network attacks using Kali Linux. In: IEEE B-HTC, pp. 1–6. IEEE (2020)
27. Tigner, M., Wimmer, H., Rebman, C.M.: Analysis of Kali Linux penetration tools: a survey of hacking tools. In: ICECET, pp. 1–6. IEEE (2021)
28. Babincev, I., Vuletić, D.: Web application security analysis using the Kali Linux operating system. Vojnoteh. glas. **64**, 513–531 (2016)
29. Lundeen, B., Alves-Foss, J.: Practical clickjacking with BeEF. In: IEEE HST, pp. 614–619. IEEE (2012)
30. Sankaran, V., Fernandes, J.: Advanced vulnerability detection and exploitation in web applications using Burp Suite. Int. J. Modern Comput. Inf. Commun. Technol. **7**, 67–74 (2024)
31. PortSwigger: Burp Suite documentation (2024). https://portswigger.net/burp/documentation

Human vs. LLM Risk Assessment: A Comparative Study of ISO 27001 Cyber Risk Classification

Marek Hrabčák[1(✉)] and Zoltán Balogh[1,2]

[1] Department of Informatics, Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia
{marek.hrabcak,zbalogh}@ukf.sk, balogh.zoltan@uni-obuda.hu

[2] Kandó Kálmán Faculty of Electrical Engineering, Obuda University, Bécsi út 96/b, 1034 Budapest, Hungary

Abstract. Accuracy in cyber risk assessment is essential for effective information security management according to ISO/IEC 27001. This study presents a comparative analysis of cyber risk assessments conducted by human evaluators and ChatGPT. The experiment used security documentation of a fictitious banking institution with processes aligned to ISO 27001. Human evaluations were prepared by approximately 130 evaluators using the methodology of the National Security Authority of the Slovak Republic, while AI evaluations used identical input data and classification criteria. Results were statistically analyzed, comparing differences in inherent and residual risks. The study provides empirical evidence on LLM applicability and limitations in ISO 27001 risk management, highlighting their potential as decision support tools rather than stand-alone expert systems.

Keywords: Cyber Security Risk Assessment · ISO/IEC 27001 · Artificial Intelligence · LLM · risk reduction

1 Introduction

The increasing complexity of cyber threats and regulatory requirements has made accurate risk assessment a cornerstone of effective information security management. Organizations implementing ISO/IEC 27001 must systematically identify, analyze, and evaluate information security risks to protect their assets and maintain compliance. However, traditional risk assessment processes are resource-intensive, requiring significant expert involvement and often suffering from subjective inconsistencies among evaluators.

The emergence of Large Language Models (LLMs) such as ChatGPT presents a potential paradigm shift in how organizations approach cyber risk assessment. These AI systems demonstrate remarkable capabilities in understanding context, processing natural language descriptions, and applying structured reasoning to complex problems. This raises a fundamental question: can LLMs reliably replicate or support human judgment in the specialized domain of ISO 27001 risk classification?

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 548–560, 2027.
https://doi.org/10.1007/978-3-032-32104-6_46

Recent studies demonstrate the broader applicability of data-driven and natural language processing methods to extracting meaningful indicators from annual reports [29, 30] and clinical language data [33]. Research on Industry 5.0 also discusses emerging technological opportunities and threats [31], while related work highlights the importance of bridging the gap between managerial and IT perspectives [32].

This study addresses this question through a controlled empirical comparison of risk assessments performed by human evaluators and ChatGPT. The research was conducted using security documentation of a fictitious banking institution (HackMeBank) with processes aligned to ISO 27001 requirements. Human evaluations were prepared by approximately 130 MSc students trained in the methodology recommended by the National Security Authority of the Slovak Republic (NBU SR), while LLM evaluations used identical input data and classification criteria.

The primary objectives of this research are: (1) to quantify the degree of agreement between human and LLM risk assessments for both inherent and residual risks; (2) to identify systematic patterns in how LLMs interpret risk attributes compared to human evaluators; (3) to evaluate whether LLMs demonstrate sufficient consistency and reliability for practical deployment in risk management workflows; and (4) to provide recommendations for the responsible integration of AI tools in ISO 27001 compliance processes.

The study analyzed 8,672 classified risks using descriptive statistics, correlation analysis, and distribution comparisons. Results reveal high agreement for inherent risk assessment but systematic conservative bias in LLM residual risk evaluation, suggesting LLMs are better suited as decision support tools rather than autonomous risk assessors.

The remainder of this paper is organized as follows: Sect. 2 reviews related work on AI in risk assessment, ISO 27001 frameworks, and LLM applications in cybersecurity. Section 3 describes the research methodology including data collection and analysis techniques. Section 4 presents results with statistical analysis. Section 5 discusses findings, their implications and concludes with recommendations for future research.

2 Related Work

2.1 AI Applications in Risk Assessment

Artificial Intelligence has significantly transformed risk assessment practices across multiple domains. Machine learning and deep learning technologies enable analysis of large datasets to identify patterns and anomalies indicating potential risks. Ganapathy [1] demonstrated that AI-based approaches in forensic auditing can detect complex fraud schemes often missed by traditional methods. Karim [2] showed that AI-enhanced risk assessment reduces human errors and biases while providing more comprehensive analyses.

Organizations integrating AI tools report significant improvements in risk management processes and enhanced compliance with standards such as ISO 27001 [3, 4]. AI-driven approaches enable proactive risk anticipation, with real-time insights supporting auditors in focusing on high-risk areas and allocating resources effectively [5]. In business continuity contexts, AI technologies including natural language processing and predictive analytics enhance risk assessment accuracy and speed [6].

Despite these benefits, AI integration poses challenges including data quality management, algorithmic bias, and ethical implications [7, 8]. The "black box" nature of some AI models creates transparency issues that complicate regulatory compliance. Integrating Explainable AI (XAI) techniques such as SHAP and LIME can enhance transparency and accountability in AI-driven risk assessments [9–11].

2.2 ISO 27001 Framework and Risk Management

ISO/IEC 27001 provides a structured framework for establishing, implementing, maintaining, and continually improving an Information Security Management System (ISMS). The standard is particularly important in regulated sectors including finance, healthcare, and energy where data protection and compliance are paramount [12, 13]. Risk assessment forms the foundation of ISO 27001 implementation, requiring organizations to identify assets, threats, and vulnerabilities, then evaluate and treat identified risks.

The standard helps organizations meet various regulatory requirements such as PCI DSS, SOX, and GDPR [14]. Achieving ISO 27001 certification provides competitive advantage by enhancing customer trust and demonstrating commitment to information security [15]. While implementation can be resource-intensive, organizations can adopt phased approaches to gradually enhance their security capabilities [16].

2.3 Large Language Models in Cybersecurity

Large Language Models have emerged as transformative tools in cybersecurity, offering advanced capabilities in threat detection, vulnerability assessment, and incident response [17, 18]. Key applications include threat intelligence through advanced log analysis and anomaly detection [19], software vulnerability identification with automated remediation proposals [20], automated incident response through actionable intelligence generation [21], and dataset augmentation for training cybersecurity models [22].

However, LLMs themselves are susceptible to vulnerabilities including prompt injection, insecure output handling, and adversarial attacks [23]. Concerns about data privacy, model interpretability, and the need for domain-specific fine-tuning remain important considerations [17, 22]. The integration of LLMs into cybersecurity frameworks must be approached with appropriate caution, ensuring responsible and secure deployment.

2.4 Existing Research on AI-Assisted Risk Classification

Deep learning models including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory networks (LSTMs) have been applied to process security telemetry data for real-time threat detection [24]. These models integrate with Security Information and Event Management (SIEM) systems for automated incident response, though they often rely on synthetic datasets that may not fully capture real-world complexities.

Explainable AI techniques enhance transparency in AI-driven risk assessments by providing interpretable outputs [25, 26]. AI-supported approaches using Bayesian networks and neural networks are increasingly applied to intrusion detection and malware

analysis [27]. Multi-fuzzy inference systems offer nuanced risk assessment adaptable to different threat scenarios [28].

Despite growing research interest, few studies have systematically compared LLM-generated risk assessments with human expert evaluations in the specific context of ISO 27001 compliance. This gap motivates the present research, which provides empirical evidence on LLM reliability and consistency when applied to standardized cyber risk classification methodologies.

Effective risk management in AI-driven environments requires close alignment between technical and managerial perspectives. However, a significant gap between business-oriented managers and IT professionals can lead to miscommunication, misinterpretation of system outputs, and ultimately increased organizational risk (Séllei et al., 2024).

Artificial intelligence is increasingly applied to the processing of large volumes of data, including unstructured text. Natural language processing (NLP) techniques, such as sentiment analysis, enable the systematic interpretation of such data while reducing subjectivity in content evaluation (Kelebercová et al., 2025). However, this also introduces new risks related to bias, interpretability, and decision-making reliability.

3 Research Methodology

The methodology was designed for systematic comparison of human and LLM risk assessment according to ISO 27001 principles. First, documentation was prepared for a fictitious bank (HackMeBank) compiled according to ISO 27001 requirements. The 38-page documentation described processes implemented according to individual ISO 27001 chapters.

MSc students of applied informatics were introduced to ISO 27001 requirements and the HackMeBank documentation. Then, 120 students analyzed processes and identified shortcomings based on acquired knowledge. Identified deficiencies were processed according to the methodology from Act No. 69/2018 Coll. on Cybersecurity of the Slovak Republic.

Students determined affected assets, selected relevant vulnerabilities and threats from catalogs, and performed risk classification through qualitative risk analysis. After assessing inherent risk, students designed mitigation measures and performed reclassification to determine residual risk. These data were then processed in ChatGPT to analyze LLM's classification ability. A total of 8,672 risks were collected and processed.

3.1 Data Collection

Students entered identified risks in MS Excel with all columns necessary for classification: ID, ISO Control, Affected asset, Asset value, Risk name, Risk description, Threat, Vulnerability, Likelihood (before/after mitigation), Impact (before/after mitigation), Risk level (before/after mitigation), and Proposed measures (Table 1).

Table 1. Example of attributes filled in during risk classification

ISO Control	5.1
Affected asset	HR system
Asset value	10000
Risk name	Unsupported OS
Risk description	Servers are running on unsupported OS
Threat name	Theft of digital identity or login information
Vulnerability	Unprotected access to the server
Probability (before)	Moderate
Impact (before)	Significant
Risk level (before)	Medium
Proposed measures	Server upgrade
Probability (after)	Unlikely
Impact (after)	Insignificant
Risk level (after)	Very low

3.2 Risk Severity Criteria

The attributes probability, impact, and risk rate obtained from students were transformed into integer values (1–5). In the pre-processing of the data, it was necessary to treat several incorrectly specified classification levels and solve minor problems with the text (encoding, spaces, etc.). The following image shows an example of the dataset used in the research (Table 2).

Table 2. Acceptance criteria for risk severity (Source: NBU SR)

Severity Level	Description
Very High	Risk directly threatens primary assets, security, or critical systems (catastrophic loss, health/life impacts)
High	Risk seriously threatens primary assets or critical systems
Medium	Risk potentially threatens primary assets or critical systems
Low	Risk affects supporting processes; critical systems not endangered
Very Low	Risk does not threaten primary assets; operations not endangered

3.3 Comparative Analysis

Student-prepared data were migrated to a single CSV file. Student ratings for inherent and residual risks were deleted, keeping asset, threat, vulnerability, and risk descriptions

with proposed mitigation measures. Data were sent to ChatGPT v. 5.2 via API with requirements to fill missing columns and a system prompt containing evaluation criteria matching NBU SR methodology.

Variable naming convention: S = student, G = ChatGPT, I = inherent risk, R = residual risk. Example: SI_Likelihood represents student rating for inherent risk likelihood. Attributes were transformed to integer values (1–5). The 5 × 5 evaluation matrix for determining risk degree was included in prompts to ChatGPT.

4 Results

4.1 Descriptive Statistics

Differences were defined as: student assessment − LLM evaluation. Negative values indicate LLM assigns higher (more conservative) risk levels; positive values indicate students rate risk higher; zero indicates complete match.

Table 3. Calculation result for descriptive statistical methods (Source: Author)

Statistical method	I_Likelihood	I_Impact	I_RiskLevel	R_Likelihood	R_Impact	R_RiskLevel
Average	−0.4319	−0.1169	−0.4979	−0.8067	−0.9036	−0.9081
Median	0	0	0	−1	−1	−1
Modus	0	0	0	−1	−1	−1
Dispersion	0.7565	1.0003	1.0395	0.6129	1.5594	0.9999
Standard deviation	0.8698	1.0001	1.0196	0.7828	1.2488	0.9999
Interquartile Span	1	2	1	1	2	2
Skewness	−0.3452	−0.4839	−0.2098	0.0798	0.2094	0.1767
Kurtosis	0.3816	0.6105	0.0497	0.0596	−0.1686	−0.0604

Table 3 presents the results of descriptive statistical analysis of differences between student and LLM assessments. The values represent the arithmetic difference: student assessment minus LLM evaluation. Negative values indicate that LLM assigns higher (more conservative) risk levels compared to students, while positive values indicate students rate risk higher. Zero indicates complete agreement between assessments.

The Average (Mean) row reveals systematic bias in assessments. For inherent attributes (I_), values are moderately negative (−0.43 for likelihood, −0.12 for impact, −0.50 for risk level), indicating slight LLM conservatism. For residual attributes (R_), values are significantly more negative (−0.81 to −0.91), demonstrating that LLM consistently assigns higher residual risk after considering mitigation measures.

Median and Modus values of 0 for inherent attributes confirm that most assessments coincide or differ minimally. For residual attributes, both equal −1, indicating the most

common scenario is LLM rating exactly one level higher than students. This represents a systematic, non-random pattern.

Dispersion (Variance) and Standard Deviation characterize assessment variability. Lower values for inherent attributes indicate consistent agreement, while higher values for residual impact (1.56 variance, 1.25 std. dev.) reveal greatest interpretative uncertainty in evaluating mitigation effectiveness.

Interquartile Span (IQR) focuses on the middle 50% of values. Values of 1–2 indicate narrow concentration of differences, with impact showing wider distribution than probability, suggesting impact is perceived as less unambiguous.

Skewness values indicate distribution asymmetry. Negative skewness for inherent attributes (-0.35 to -0.48) shows slight predominance of cases where LLM is more conservative. Near-zero or slightly positive skewness for residual attributes indicates more balanced distribution.

Kurtosis values near zero indicate no extreme tails or concentration of extreme values, supporting the conclusion that differences represent stable, systematic patterns rather than isolated anomalies.

4.2 Correlation Analysis

Table 4. Correlation matrix for student (S) and ChatGPT (G) assessments (Source: Author)

	SI_Likelihood	SI_Impact	SI_RiskLevel	SR_Likelihood	SR_Impact	SR_RiskLevel	GI_Likelihood	GI_Impact	GI_RiskLevel	GR_Likelihood	GR_Impact	GR_RiskLevel
SI_Likelihood	1											
SI_Impact	0.20697736	1										
SI_RiskLevel	0.59554592	0.66781128	1									
SR_Likelihood	0.56259787	0.16456257	0.34476771	1								
SR_Impact	0.17025559	0.53267379	0.3615122	0.19765546	1							
SR_RiskLevel	0.35796409	0.44805544	0.48684496	0.58555688	0.63703559	1						
GI_Likelihood	0.16233096	0.17283952	0.17001874	0.12813136	0.16674628	0.17342422	1					
GI_Impact	0.08473837	0.33837357	0.22907242	0.05288161	0.24483137	0.18550928	0.34707822	1				
GI_RiskLevel	0.12073679	0.25032743	0.19950102	0.08012628	0.18692181	0.1528108	0.60782811	0.66939345	1			
GR_Likelihood	0.10044234	0.16196703	0.13808536	0.09665825	0.13363736	0.13671903	0.70480656	0.3270954	0.49895194	1		
GR_Impact	0.05602575	0.28077031	0.18414287	0.03266427	0.19393954	0.1498971	0.33266838	0.79585352	0.55791117	0.33466764	1	
GR_RiskLevel	0.08477445	0.22797669	0.18259685	0.07338267	0.16432528	0.13629233	0.55054736	0.61231751	0.79229983	0.63493594	0.62739413	1

Table 4 presents the correlation matrix examining relationships between student assessments (prefix S) and LLM assessments (prefix G). Index I represents inherent risk, while R indicates residual risk after mitigation. This analysis reveals the internal consistency of each evaluator and the degree of agreement between human and AI assessments.

Internal consistency of student assessments: Medium to strong correlations exist between inherent probability, impact, and risk level (SI_Lik to SI_Risk: 0.596; SI_Imp to SI_Risk: 0.668). These values indicate students implicitly follow the qualitative risk matrix logic, though not always strictly formally. The highest correlation coefficients occur between risk level and its input components.

Internal consistency of LLM assessments: Similar or slightly higher correlations are observable in LLM evaluations (GI_Lik to GI_Risk: 0.608; GI_Imp to GI_Risk: 0.669). This points to high internal coherence, with the model systematically applying the same decision-making framework. The strong correlation between GI_Risk and GR_Risk (0.792) indicates consistent link between initial and mitigated risk assessments.

Relationship between inherent and residual assessment: Significant correlations within the same evaluator indicate residual assessment builds on original inherent assessment. This relationship is more pronounced in LLMs (GI_Lik to GR_Lik: 0.705) than students (SI_Lik to SR_Lik: 0.563), supporting the conclusion that the model maintains more consistent linkage between initial and post-mitigation risk states.

Cross-evaluator correlations: Coefficients between corresponding student and LLM attributes are predominantly low to medium (0.08–0.25). This suggests that although both use identical input descriptions, assessments are not strongly linearly dependent. Highest agreement appears at inherent risk level, while residual correlations are weaker, confirming differences concentrate in mitigation interpretation.

Impact dominance: Risk level shows higher correlation with impact than probability for both evaluators (students: 0.668 vs. 0.596; LLM: 0.669 vs. 0.608). This pattern suggests impact is perceived as the dominant factor in determining risk severity, consistent with conservative risk management approaches in regulated sectors like banking.

4.3 Interpretation of Results

Mean values: Negative averages indicate LLM tends to assign higher risk levels than students. This effect is more pronounced for residual attributes. For inherent impact, average difference is close to zero, indicating high agreement. For residual attributes, significantly negative averages suggest LLM evaluates risks more conservatively after considering mitigation measures.

Median and mode: For inherent attributes, median equals zero, meaning student and LLM grades mostly coincide. For residual attributes, median of -1 indicates systematic shift toward higher risk by one qualitative grade. The mode confirming this value indicates the difference is not random but represents the most common scenario.

Variance and standard deviation: Lower values for inherent attributes indicate relatively consistent ranking. Higher variability in residual attributes, especially residual impact (~1.25), indicates highest interpretative uncertainty in assessing mitigation measure effectiveness.

4.4 Correlation Analysis

Within student assessments, medium to strong correlation between inherent probability, impact, and risk level is observable. Students implicitly follow the qualitative risk matrix logic. An analogous pattern is observable in LLM evaluations with comparable or slightly higher correlations, pointing to high internal consistency.

Significant correlations between inherent and residual attributes within the same evaluator indicate residual assessment builds on original inherent assessment. This relationship is more pronounced in LLMs than students, supporting conclusion that the model maintains consistent link between initial and mitigated risk.

Correlation coefficients between corresponding student and LLM attributes are predominantly low to medium. Highest agreement is at inherent risk level, consistent with previous findings. Correlations for residual attributes are weaker, reaffirming differences concentrate in interpretation of mitigation measures. Impact shows higher correlation with resulting risk level than probability, suggesting impact is perceived as dominant factor in risk severity determination.

4.5 Histogram Analysis

Inherent attributes: Distributions show significant concentration around zero. Highest frequencies in intervals $-1, 0, +1$, with zero being dominant. This indicates high agreement between student and LLM assessment. Distributions are roughly symmetrical with slight shift toward negative values, suggesting LLMs rate inherent risk slightly more conservatively. Extreme differences (± 3 or more) are negligible.

Residual attributes: Histograms differ significantly from inherent ones. Highest concentration in negative intervals, especially around -1, with clearly left-sided (negatively skewed) distributions. This indicates systematic shift in LLM assessment toward higher residual risk. The dominant difference of -1 indicates shift is generally limited to one qualitative step. Wider distributions than inherent attributes point to higher variability in evaluating mitigation effectiveness.

The correlation matrix structure shows LLM rating is not random or inconsistent. It shows high internal coherence and respects relationships between probability, impact, and risk level. Differences between student and LLM assessments manifest not as chaotic variations but as systematic shift, especially in residual assessment phase.

Comparing inherent and residual distributions directly, mitigation measure application represents the main source of differences. While inherent distributions are narrow and centered around zero, residual distributions show systematic displacement and greater dispersion. LLMs and students use different interpretative frameworks for assessing measure effectiveness. Students tend to attribute higher reduction effect to measures, while LLM applies more conservative interpretation closer to strict NBU SR methodology.

5 Conclusion

The correlation matrix structure shows LLM rating is not random or inconsistent. It shows high internal coherence and respects relationships between probability, impact, and risk level. Differences between student and LLM assessments manifest not as chaotic

variations but as systematic shift, especially in residual assessment phase. Comparing inherent and residual distributions directly, mitigation measure application represents the main source of differences. While inherent distributions are narrow and centered around zero, residual distributions show systematic displacement and greater dispersion. LLMs and students use different interpretative frameworks for assessing measure effectiveness. Students tend to attribute higher reduction effect to measures, while LLM applies more conservative interpretation closer to strict NBU SR methodology.

5.1 Summary of Findings

The research compared cyber risk assessment by students and LLM using qualitative methodology according to NBU SR. Results show high agreement between student and LLM in inherent risk assessment. Median and mode of differences are concentrated around zero with relatively low variability. This suggests LLM can reliably replicate professional judgment when assessing basic risk characteristics from asset, threat, and vulnerability descriptions.

Significant differences were evident in residual risk assessment. Analysis showed systematic shift in LLM assessments toward higher risk, typically by one qualitative grade. This was confirmed by mean, median, mode, and histogram shapes showing concentration of negative values. Correlation analysis points to higher internal consistency in LLM assessments compared to students.

Overall, observed differences are not random but represent stable, systematic behavior pattern. LLM applies more conservative approach to residual risk assessment, aligning with strict NBU SR methodology interpretation and regulatory expectations in banking.

5.2 Recommendations for Future Research

Future directions include: (1) extending analysis with inferential statistical methods such as Cohen's κ or non-parametric tests to formally quantify agreement degree; (2) detailed risk segmentation by type (technical, process, organizational) to analyze whether conservative shift is uniform across categories; (3) comparing multiple LLM models including open-source to assess architecture and training data impact on consistency.

From application perspective, exploring hybrid approaches where LLM serves for pre-classification or identifying high-uncertainty risks, with final decision by human expert, could increase efficiency without reducing regulatory defensibility.

5.3 Final Thoughts

Results indicate LLMs have potential as valuable support tools in qualitative cyber risk assessment according to NBU SR methodology. High inherent risk agreement indicates ability to correctly interpret input information and apply consistent decision-making framework. However, systematic conservative shift in residual assessment points to need for caution in practical deployment.

This shift can be viewed not only as constraint but as advantage in regulated industries where cautious risk management is preferred. Research confirms differences

between human and LLM assessments result from systematic, consistent interpretation of mitigation measures, not random model behavior. This creates foundation for further research in assisted risk assessment and opens space for responsible, auditable AI use in cybersecurity.

Acknowledgments. This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and of Slovak Academy of Sciences (SAS) under the contract No. VEGA 1/0492/26.

References

1. Ganapathy, V.: AI-based risk assessments in forensic auditing. Shodh Sari. **3**(4), 100–128 (2024)
2. Karim, A.B.: AI-enhanced risk assessment in physical security projects. IJFAES. **3**(10), 1235–1266 (2024)
3. Onwubuariri, E.R., et al.: AI-driven risk assessment: revolutionizing audit planning. Financ. Account. Res. J. **6**(6), 1069–1090 (2024)
4. Kalogiannidis, S., et al.: AI technology in predictive risk assessment for business continuity. Risks. **12**(2), 19 (2024)
5. Singh, P.: AI and financial risk management. In: Advances in CI and Robotics, pp. 373–404. IGI Global (2025)
6. Cadet, E., et al.: Large language models for cybersecurity policy compliance. IJSRSSH. **1**(2), 612–643 (2024)
7. Yang, R.: Application of AI in audit risk assessment. Financ. Strateg. Manag. Rev. **1**(1), 1000046 (2025)
8. Khan, W., et al.: AI-powered risk assessment. In: Advances in CI and Robotics, pp. 59–92. IGI Global (2025)
9. Mallela, I.R., et al.: Explainable AI for compliance and regulatory models. IJRPS. **11**(4), 319–339 (2020)
10. Anang, A.N., et al.: Explainable AI in financial technologies. IJSRA. **13**(1), 1793–1806 (2024)
11. Yazdi, M.S., et al.: Navigating AI power in risk management. Safety. **10**(2), 42 (2024)
12. Accerboni, F., Sartor, M.: ISO/IEC 27001. Quality Management (2019)
13. Shohoud, M.: Study the effectiveness of ISO 27001 to mitigate the cyber security threats in the Egyptian downstream oil and gas industry. J. Inf. Secur. **14**(2), 152–180 (2023). https://doi.org/10.4236/jis.2023.142010
14. Cissé, M.: ISO 27001 compliance project for cyber security service team. Cyber Secur. A Peer-Rev. J. **2**(4), 346–359 (2019)
15. Khan, W.A.: Achieving regulatory compliance with ISO 27001 and NIST frameworks. JAICC. **1**(3), 1–14 (2022)
16. Mera-Amores, F., Roa, H.N.: Enhancing information security in SMEs through ISO 27001. In: LNCS, pp. 197–207 (2024)
17. Xu, H., et al.: Large language models for cyber security: a systematic literature review. ACM TOSEM. (2025)
18. Güven, M.: A comprehensive review of LLMs in cyber security. IJCESEN. **10**(3) (2024)
19. Hassanin, M., Moustafa, N.: Overview of LLMs for cyber defences. arXiv preprint arXiv:2405.14487. (2024)
20. Taghavi, S.F., Feyzi, F.: Using LLMs to detect software vulnerabilities. RS3-4387414 (2024)

21. Zangana, H.M., et al.: Large language models in cybersecurity. In: AISPE Book Series, pp. 277–300 (2024)
22. Ferrag, M.A., et al.: Generative AI and LLMs for cyber security. arXiv preprint https://arxiv.org/abs/2405.12750. (2024)
23. Jaffal, N.O., et al.: LLMs in cybersecurity: applications, vulnerabilities, defense. AI. **6**(9), 216 (2025)
24. Chowdhury, T.: AI-powered deep learning for real-time cybersecurity risk assessment. ASRC Procedia: Glob. Perspect. Sci. Scholar. **1**(1), 675–704 (2025)
25. Mohitkar, C., Lakshmi, D.: Explainable AI for transparent cyber-risk assessment. ACIR, 219–246 (2024)
26. Obuse, E., et al.: Explainable AI for cyber threat intelligence. JFMR. **1**(2), 15–30 (2020)
27. Erdogan, G., et al.: Systematic mapping study on AI-supported security risk assessment. In: COMPSAC, pp. 755–760 (2021)
28. Natsheh, E., Tabook, F.B.: AI-enhanced cybersecurity risk assessment with multi-fuzzy inference. JICTRA. **19**(1), 1–26 (2025)
29. Kelebercová, L., Munk, M., Pilková, A., Bližnák, K.: Sentiment shifts: sentiment analysis of annual reports during key economic and political periods. Int. J. Serv. Technol. Manag. **30**(2–3), 166–181 (2025)
30. Pilková, A., Elley, S., Munk, M., Kelebercová, L., Blažeková, P.: Unlocking financial insights: decoding sentiment in annual reports to transform bank risk and performance metrics. Int. J. Serv. Technol. Manag. **30**(2–3), 182–197 (2025)
31. Kovari, A.: Industry 5.0: generalized definition, key applications, opportunities and threats. Acta Polytech. Hung. **21**(3), 267–284 (2024). https://doi.org/10.12700/APH.21.3.2024.3.17
32. Séllei, B., Krén, H., Molnár, G., Stumphauser, N.: Gaps and bridges between future managers and IT people. Acta Polytech. Hung. **21**(3), 53–72 (2024). https://doi.org/10.12700/APH.21.3.2024.3.5
33. Zozuk, N.C., Munkova, D., Kelebercova, L., Munk, M.: Relationship between language features extracted through NLP and clinically diagnosed Alzheimer's disease and mild cognitive impairment in Slovak. Alzheimers Dement. Diagn. Assess. Dis. Monit. **17**(2), e70122 (2025)

Development of a Robot Function to Support Children's Movement Coordination: A NAO-Based Approach

Enikő Nagy[1] and György Molnár[2,3](✉)

[1] Antal Bejczy Center for Intelligent Robotics, John von Neumann Faculty, Óbuda University, Budapest, Hungary
nagy.eniko@uni-obuda.hu

[2] Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest, Hungary
molnar.gyorgy@uni-obuda.hu

[3] Apáczai Csere János Faculty of Humanities, Education and Social Sciences, Széchenyi University, Győr, Hungary

Abstract. As a result of rapid technological development, digitalization and smart solutions are playing an increasingly significant role in various areas of everyday life. As part of this process, it is of particular importance that future generations become familiar with robotic applications from an early age, thereby supporting the development of natural human–technology interaction and reducing fear of the unknown. Early exposure to such technologies can contribute to the development of digital competencies and foster a positive attitude toward technological innovation. Within the framework of the presented research, a social robot function was developed with the involvement of university students, designed for use in environments commonly accessed by young children. The primary objective of the development was to create an interactive solution that supports parents and caregivers in supervising children while simultaneously engaging them in playful and communicative activities. Special emphasis was placed on ensuring safe operation, child-friendly design, and the creation of a supportive and welcoming atmosphere. The aim of the paper is to provide both scientific and practical justification for the relevance and applicability of the developed social robot function, as well as to contribute to the advancement of child-centered and pedagogically grounded robotic solutions. The findings are intended to support future research and development efforts in the field of social and educational robotics.

Keywords: social robotics · child-centered technology · educational robotics · human–robot interaction · NAO robot

1 Introduction

The rapid development of modern technology has led to the widespread digitalization of everyday life, resulting in the increasing presence of intelligent systems and smart solutions across various domains. Among these technological advancements, robotics—particularly social robotics—has gained growing attention due to its potential to support

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 561–570, 2027.
https://doi.org/10.1007/978-3-032-32104-6_47

human activities not only in industrial settings but also in social, educational, and service-oriented environments. As these technologies become more embedded in society, it is essential to consider how future generations perceive, understand, and interact with robotic systems.

Early exposure to technology plays a crucial role in shaping children's attitudes toward digital tools and intelligent systems. Introducing robotic applications at a young age can significantly reduce fear of unfamiliar technologies and foster a more natural and confident human–technology interaction. In particular, social robots, which are designed to communicate and interact with humans using verbal and non-verbal cues, offer promising opportunities for engaging children in a playful, supportive, and emotionally responsive manner. Previous research has shown that children are generally more open to interacting with robots that are perceived as friendly, predictable, and relatable, emphasizing the importance of child-centered design principles [1].

In recent years, social robots have been increasingly applied in educational and caregiving contexts, where they support learning processes, enhance motivation, and assist adults in supervision and guidance tasks. These systems can provide interactive activities, encourage movement and communication, and contribute to the development of both cognitive and social skills. However, the design and implementation of such solutions require careful consideration of safety, usability, and ethical aspects, especially when deployed in environments used by young children [2].

This paper presents the development of a social robot function created with the involvement of university students, specifically designed for child-oriented environments. The primary objective of the developed function is to support parents and caregivers in supervising children while simultaneously engaging them through interactive and playful activities in a safe and friendly atmosphere. By combining technological innovation with pedagogical considerations, the proposed solution aims to demonstrate how social robots can be effectively integrated into everyday child-centered settings.

The contribution of this paper is twofold. From a scientific perspective, it provides insights into the design and application of child-friendly social robot functions, contributing to the field of human–robot interaction and educational robotics. From a practical standpoint, it presents a real-world implementation and evaluation that highlights the potential benefits and limitations of such systems. The remainder of the paper is structured as follows: Section 2 reviews the relevant theoretical background and related work, Section 3 describes the system design and development process, Section 4 presents the application scenario, Section 5 discusses the evaluation results, and Sections 6 and 7 conclude the paper with a discussion and future research directions.

2 Theoretical Background and Related Work

Social robotics has become one of the most dynamically developing research areas in recent years, building on the multidisciplinary field of Human–Robot Interaction (HRI). Unlike industrial robots, social robots are primarily designed not for physical labor but for natural, socially oriented communication and interaction with humans. As a result, these systems are capable of interpreting and generating verbal and non-verbal cues—such as speech, gestures, facial expressions, and body posture—thereby enabling human-like interactions [3, 4].

2.1 Social Robots and Human–Robot Interaction

The field of human–robot interaction focuses on principles and methods that enable effective, safe, and acceptable cooperation between humans and robots. In the case of social robots, user experience plays a crucial role, as the success of these systems largely depends on how friendly, trustworthy, and understandable their behavior is perceived to be. Previous studies have shown that anthropomorphic appearance, predictable responses, and the ability to express emotions significantly increase user acceptance.

When interacting with children, human–robot relationships exhibit specific characteristics. Children tend to form emotional connections with robots, especially when the systems demonstrate playful, friendly, and consistent behavior. This phenomenon creates opportunities to apply social robots as supportive tools in learning, development, or supervision contexts, while also requiring responsible design that takes ethical and pedagogical considerations into account [5].

2.2 Child-Friendly Robotic Solutions

Designing child-friendly robotic systems requires special attention to safety, ease of use, and adaptation to age-specific characteristics. Robots must be both physically and psychologically safe, avoiding intimidating appearances or unpredictable behavior. Furthermore, it is essential that the system maintains children's interest without causing excessive cognitive load or frustration.

According to the literature, social robots designed for children are most effective when they encourage activity, movement, and communication in a playful manner. Such systems typically serve a complementary role in education and caregiving, supporting rather than replacing adult supervision. Research findings indicate that well-designed robot–child interactions can enhance motivation and contribute to the development of cognitive and social skills [5, 13].

2.3 Application of Social Robots in Child-Centered Environments

The application of social robots is becoming increasingly common in child-centered environments such as educational institutions, developmental programs, and recreational spaces. In these contexts, robots can support both structured and unstructured activities, including playful tasks, movement-based exercises (see Fig. 1), and communication scenarios [13]. The presence of robots often has a positive impact on children's engagement and attention, particularly when the system adapts to individual needs and responses.

Overall, the research results in social robotics and human–robot interaction provide a strong theoretical foundation for the development of child-friendly robot functions [15, 16]. These findings support the relevance of applying social robots as supportive tools in environments involving children, effectively complementing the work of parents and caregivers [9, 10].

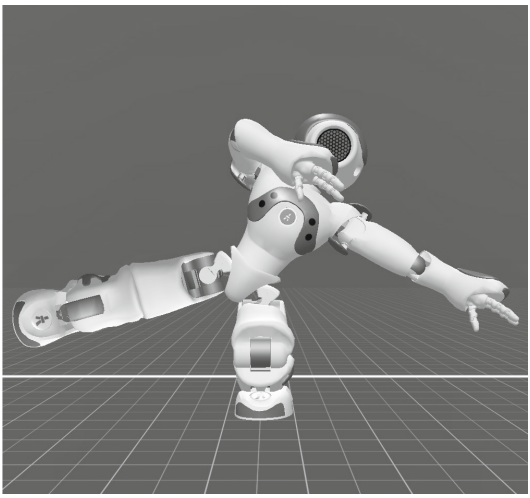


Fig. 1. Movement based exercises by NAO [8] (own editing)

3 System Design and Development Methodology

The development of the presented social robot function aimed to create a modular and flexible system that can be safely applied in child-centered environments while supporting interactive and playful activities. The development process involved university students, which not only supported practice-oriented education but also enabled the integration of diverse design approaches and user perspectives.

The primary design objective was to ensure children's safety and well-being. Accordingly, all elements of the robot function operate in a predictable and transparent manner within a limited range of motion. Age-specific characteristics were taken into account, particularly children's attention span, the need for simple instructions, and the preference for playful interactions.

Another key principle was the implementation of child-friendly communication, utilizing simple language, positive feedback, and repetitive interaction patterns. The system is not intended to make autonomous decisions but rather to support adult supervision; therefore, the robot's functions are complementary and do not replace the presence of parents or caregivers.

The developed system is based on a social robot platform equipped with sensors and actuators that enable environmental perception and responsive behavior. The robot's sensors allow the detection of basic environmental cues, such as movement and sound, while actuators are responsible for motion, gestures, and visual feedback. Moderate intelligence-based signal and data processing can provide good professional support for motion perception [11, 13, 14].

The software architecture follows a modular structure, separating perception, decision-making, and execution components [6]. This approach allows for future expansion and customization of functionalities. An event-driven logic was applied to

ensure fast response times and predictable behavior, which is particularly important in interactions with children [15].

Several interaction scenarios were designed to support engagement with children in different situations. These include greeting and introduction sequences, simple playful tasks, and activity-promoting functions. The robot communicates with children through verbal instructions and non-verbal cues, such as movements and visual signals.

Repeatability and adaptability were key considerations in function design. The robot is capable of scheduling and repeating interactions, which contributes to children's sense of safety and familiarity. Additionally, the system allows adult users to control the type and intensity of interactions through simple configuration settings.

The development process followed an iterative approach in which design, implementation, and testing were closely integrated. University students worked in small development units, and completed functions were evaluated collaboratively. Initial testing was conducted in a laboratory environment, focusing on system stability and safety.

In later phases, trials were carried out in simulated child-centered environments, where interaction flows and potential issues were observed. Based on the testing results, behavioral patterns and timing parameters were refined to achieve more natural and predictable robot behavior.

4 Application Scenario: Movement Sequence–Based Interactions on the NAO Robot Platform

The developed social robot function is implemented on the NAO humanoid robot platform developed by SoftBank Robotics, which is well suited for child-centered interactions. NAO's small size, friendly appearance, and advanced motion and communication capabilities enable the implementation of playful and safe movement sequences designed for young children. The solution developed by the student consists of three progressively structured movement sequences: easy, medium, and difficult [19, 20].

The design of the movement sequences took into account the physical and software characteristics of the NAO robot, particularly joint range limits, balance maintenance, and motion stability. NAO's humanoid structure allows clear demonstration of arm, head, and torso movements, while leg movements were intentionally limited to enhance safety.

The movements were implemented using the NAOqi framework, also using Choregraphe (see Fig. 2), which enables precise joint control and repeatable predefined motion sequences. Each movement sequence is accompanied by verbal instructions and positive feedback delivered through NAO's built-in speakers and speech synthesis capabilities [7].

4.1 Easy Movement Sequence on the NAO Robot

The easy movement sequence is based on NAO's fundamental motion capabilities and is primarily intended for initial interactions. This level includes simple, isolated movements such as arm lifting, waving, head turning, or slight torso leaning. Movements are performed at a slow pace with large amplitudes to ensure clear visibility and easy imitation by children [8, 17].

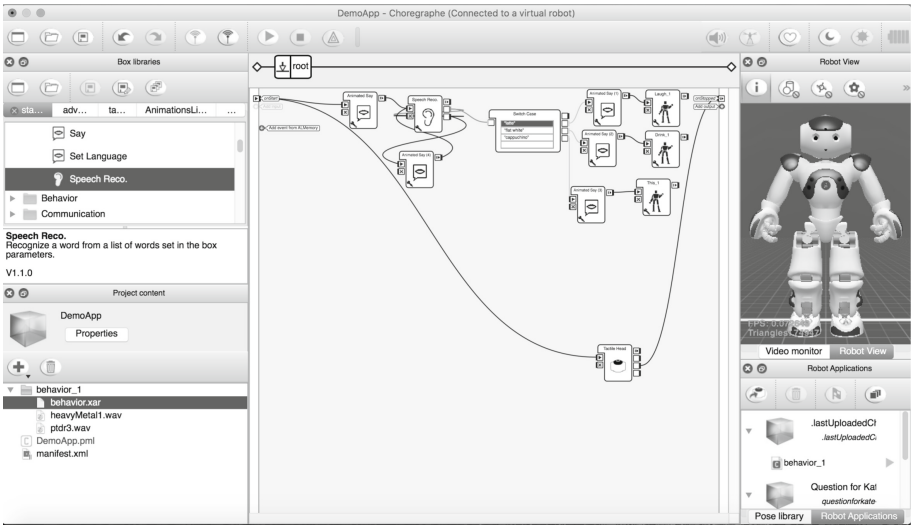


Fig. 2. Choregraphe development interface [8] (own editing)

During execution, NAO provides short and simple verbal cues to encourage imitation. The objective of the easy sequence is to familiarize children with the robot, capture attention, and create a positive first interaction.

4.2 Medium-Difficulty Movement Sequence on the NAO Robot

The medium-difficulty movement sequence incorporates more complex motion combinations requiring coordinated control of multiple NAO joints. Examples include combined arm and torso movements, simple rhythmic actions, and synchronized head and arm motions [8].

At this level, NAO's movements become more dynamic while remaining controlled and predictable. Verbal instructions emphasize synchronized movement and imitation. The goal of the medium sequence is to develop children's motor coordination, attention, and motor memory.

4.3 Difficult Movement Sequence on the NAO Robot

The difficult movement sequence represents a more complex utilization of NAO's capabilities. This level consists of longer, multi-step movement sequences that require increased concentration and more precise imitation. Movements may include rhythm changes and more elaborate arm movement patterns [8].

At this stage, NAO uses fewer repetitions and assumes a higher level of autonomy from the children. Positive feedback remains a key element, while the system encourages children to engage with challenges and support their development. The objective of the difficult sequence is to enhance complex motor coordination and persistence.

4.4 Application Experiences Using the NAO Robot

Testing of the NAO-based movement sequences showed that NAO's humanoid appearance and clearly recognizable motions facilitated rapid engagement. The easy sequence proved particularly effective for initial interactions, while the medium and difficult levels sustained interest over longer periods.

Feedback from parents and caregivers indicated that NAO's movement sequences integrated well into child-centered environments, providing structured yet playful activities while maintaining safe operation.

5 Evaluation and Results

The aim of evaluating the developed movement sequence–based robot function was to examine how the easy, medium, and difficult movement sequences presented by the NAO robot contribute to children's engagement, activity level, and the overall success of robot–child interaction. The evaluation primarily relied on qualitative observations and informal feedback, as the main focus of the development was to assess functionality and applicability.

5.1 Evaluation Methodology

The movement sequences were tested in a simulated child-centered environment. During the evaluation, children's reactions, attention levels, and degree of imitation were observed across the different difficulty levels. The assessment was conducted in a non-invasive manner, focusing on natural interactions without external intervention.

Observation criteria included initial interest, duration of interaction, and the ability of children to follow the movements demonstrated by the NAO robot. Informal feedback from parents and caregivers also played an important role in the evaluation process [18].

The research study was conducted in the fall of 2025 among students aged 6–10 with parental consent. The total number of student observers was N=25, of whom 24 could be evaluated. The worksheet, based on closed-ended questions, contained a total of six guided observation criteria.

5.2 Evaluation of the Easy Movement Sequence

When applying the easy movement sequence, children responded quickly to NAO's movements and were eager to imitate the simple actions. The slow pace and opportunity for repetition contributed to early experiences of success during initial interactions.

Observations indicated that this level was particularly effective in familiarizing children with the robot and establishing initial trust. Children's attention was generally maintained throughout the sequence, and they responded positively to verbal praise.

5.3 Evaluation of the Medium-Difficulty Movement Sequence

The medium-difficulty movement sequence required a higher level of concentration from the children. More complex movement combinations were found to sustain interest for longer periods, especially among children who had already mastered the simpler actions.

Most children were able to follow NAO's movements, although the precision of imitation varied individually. The medium sequence supported sustained attention and the development of motor coordination while remaining within a playful interaction framework. .

5.4 Evaluation of the Difficult Movement Sequence

Children responded differently to the challenges posed by the difficult movement sequence. While some found the longer and more complex sequences motivating, others preferred the easy or medium levels. This observation highlights the importance of providing differentiated difficulty levels.

In the case of the difficult sequence, increased persistence and independent attempts were observed, particularly among children with prior interaction experience. Positive feedback continued to play a key role in maintaining motivation.

5.5 Summary of Results

Overall, the evaluation results indicate that the three-level movement sequence implemented on the NAO robot effectively supported children's engagement and activity. The easy sequence proved especially effective during initial interactions, while the medium and difficult levels enabled gradual development and longer-term engagement.

These findings confirm that child-friendly, progressively structured movement sequences are suitable for integrating social robots as supportive tools in child-centered environments.

6 Discussion

Based on the presented development and evaluation results, it can be concluded that the three-level movement sequence implemented on the NAO robot proved to be an effective tool for engaging children and encouraging physical activity. The gradual approach allowed the system to be applicable to children of different ages and abilities while maintaining a playful and safe interaction style.

The findings are consistent with previous research in social robotics and human–robot interaction, which highlights that children are more likely to engage with robots exhibiting predictable, friendly, and adaptive behavior. NAO's humanoid appearance and clearly recognizable movements were particularly suitable for imitation-based interactions, which are essential in movement sequence–based activities.

An important observation is that providing multiple difficulty levels contributed to increased autonomy and motivation among children. While some children were motivated by more complex movement sequences, others preferred simpler tasks. This

differentiation is especially relevant in child-centered environments, where individual differences play a significant role.

At the same time, several limitations of the study should be acknowledged. The evaluation relied primarily on qualitative observations and informal feedback, limiting the generalizability of the results. In addition, the testing period was relatively short, and therefore long-term effects—such as sustained motivation or developmental outcomes—require further investigation.

7 Conclusion and Future Research Directions

This paper presented a social robot function developed with the involvement of a university student and implemented on the NAO robot platform. The system supports interaction with children through three movement sequences of increasing difficulty, designed to encourage physical activity, attention, and imitation-based learning in a safe and playful manner.

The evaluation results indicate that the movement sequence–based approach is suitable for enhancing children's engagement and strengthening the supportive role of social robots in child-centered environments. The three-level structure enables adaptation to individual needs, which is particularly important in early childhood technology exposure.

Future research directions include the introduction of quantitative evaluation methods, such as structured observation scales or measurable interaction metrics. Additionally, long-term studies could explore the impact of the system on the development of children's motor and social skills.

Further development may focus on implementing adaptive behavior, allowing the NAO robot to dynamically select appropriate difficulty levels based on children's responses. The presented solution thus provides a foundation for future research and practical development in child-friendly social robotics.

References

1. Conti, D., Di Nuovo, S., Di Nuovo, A.: Kindergarten children attitude towards humanoid robots: what is the effect of the first experience? In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 630–631. IEEE (2019)
2. Shamsuddin, S., Zulkifli, W., Kamat, S.: Kansei emotions on human-robot interaction with humanoid robot NAO at kindergarten setting. In: Proceedings of the International Conference on Design and Concurrent Engineering, Osaka (2017)
3. Hegel, F., Muhl, C., Wrede, B., Hielscher-Fastabend, M., Sagerer, G.: Understanding social robots. In: 2009 Second International Conferences on Advances in Computer-Human Interactions, pp. 169–174. IEEE (2009)
4. Eyssel, F., Kuchenbrandt, D.: Social categorization of social robots: anthropomorphism as a function of robot group membership. Brit. J. Soc. Psychol. **51**(4), 724–731 (2012)
5. Nagy, E., Karl, E., Molnar, G.: Exploring the role of human-robot interactions, within the context of the effectiveness of a NAO robot. Acta Polytech. Hung. **21**(3), 177–190 (2024)
6. Pot, E., Monceaux, J., Gelin, R., Maisonnier, B.: Choregraphe: a graphical tool for humanoid robot programming. In: RO-MAN 2009: The 18th IEEE International Symposium on Robot and Human Interactive Communication, pp. 46–51. IEEE (2009)

7. Aldebaran Robotics: NAO 2.1 Documentation. Aldebaran Robotics (2014). Accessed 27 Oct 2025
8. F. Tóth: The role and application of social and autonomous robots in modern urban and service environments. Final thesis, Óbuda University (2025)
9. Feidakis, M., Gkolompia, I., Marnelaki, A., Marathaki, K., Emmanouilidou, S., Agrianiti, E.: NAO robot, an educational assistant in training, educational and therapeutic sessions. In: 2023 IEEE Global Engineering Education Conference (EDUCON), pp. 1–6, Kuwait (2023)
10. Rossi, L., Orsenigo, E., Bisagno, E., Cadamuro, A.: From learning machines to teaching robots. Interaction for educational purposes between the Social Robot NAO and children: a systematic review. J. E-Learn. Knowl. Soc. **20**(1), 15–26 (2024)
11. Molnár, G., Kiss, A.Á., Nagy, E.: Integration of visual programming and artificial intelligence in robotics education: the pedagogical possibilities of the NAO robot. In: 2025 IEEE 23rd Jubilee International Symposium on Intelligent Systems and Informatics (SISY), pp. 45–50, Subotica (2025)
12. Rajcsányi-Molnár, M., Balázs, L., András, I., Czifra, S.: Competition as an effective motivational tool in online education. In: Szakál, A. (ed.) IEEE 11th International Conference on Computational Cybernetics and Cyber-Medical Systems, ICCC 2024: Proceedings, pp. 83–88. IEEE (2024)
13. Nagy, E.: Processing sensor data from the NAO robot. J. Appl. Tech. Educ. Sci. **14**(4), 1–18 (2024)
14. Balogh, Z., Fodor, K., Magdin, M., Francisti, J., Koprda, S., Kővári, A.: Development of available IoT data collection devices to evaluate basic emotions. Acta Polytech. Hung. **19** (2022)
15. Balogh, Z., Munk, M., Turcáni, M.: Analysis of students' behaviour in the web-based distance learning environment. In: Recent Researches in Circuits, Systems, Communications and Computers, pp. 339–344 (2011)
16. Katona, J., Kovari, A.: Examining the learning efficiency by a brain-computer interface system. Acta Polytech. Hung. **15**(3), 251–280 (2018)
17. Kővári, A.: Industry 5.0: generalized definition, key applications, opportunities and threats. Acta Polytech. Hung. **21**(3), 267–284 (2024)
18. Holik, I., Kersánszki, T., Márton, Z., Sanda, I.D.: Under the spell of video games—Minecraft is an opportunity for game-based learning. Acta Polytech. Hung. **21**(3), 305–324 (2024)
19. Nagy, E.: Social and sexual robots: technological development, social acceptance and ethical challenges. Acta Polytech. Hung. **22**(1), 265–277 (2025)
20. Kovari, A., Andras, I., Rajcsanyi-Molnar, M.: The Role of AI in ecohumanistic education. J. Ecohumanism. **3**(3), 1361–1370 (2024)

Low-Cost Flight Data Recording and Cloud-Based Analysis Platform for General Aviation Training

Martin Vozár[1](✉), Ladislav Főző[2,3], Peter Korba[3], and Zoltán Balogh[1,2]

[1] Department of Informatics, Constantine the Philosopher University in Nitra, Nitra, Slovakia
mvozar@ukf.sk

[2] Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest, Hungary

[3] Department of Aviation Engineering, Faculty of Aeronautics, Technical University of Košice, Košice, Slovakia

Abstract. Flight data recording and analysis are essential components of aviation safety management. While commercial aviation benefits from sophisticated Flight Data Recorders (FDR) and Flight Data Monitoring (FDM) programs, general aviation lacks affordable solutions for systematic flight data collection and analysis. This paper presents the design and implementation of a low-cost flight data recording system combined with a cloud-based web application for data visualization and analysis. The hardware component utilizes MEMS sensors (IMU, barometric pressure, GPS) integrated with an Arduino microcontroller featuring Wi-Fi connectivity for real-time data transmission. The software platform implements a three-tier architecture with Spring Boot backend, React frontend, and relational database, enabling flight trajectory visualization, parameter analysis, and anomaly detection. The proposed system demonstrates that Flight Data Monitoring principles can be effectively applied in general aviation training environments without requiring certified equipment or aircraft system modifications. Experimental results confirm the system's capability to capture representative flight parameters suitable for trend analysis and identification of non-standard flight situations.

Keywords: Flight Data Recording · MEMS Sensors · Flight Data Monitoring · General Aviation · Cloud Computing · Web Application

1 Introduction

Flight data recording has been a cornerstone of aviation safety since the introduction of Flight Data Recorders (FDR) in the 1950s. These devices, commonly known as 'black boxes,' have proven invaluable for accident investigation and have contributed significantly to the remarkable safety record of commercial aviation [1, 2]. Over decades of development, FDR technology has evolved from mechanical recorders capable of capturing only a few parameters to modern solid-state systems that can record hundreds of flight variables with high sampling frequencies.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 571–584, 2027.
https://doi.org/10.1007/978-3-032-32104-6_48

Beyond accident investigation, the aviation industry has recognized the proactive potential of flight data analysis. Programs such as Flight Data Monitoring (FDM) and Flight Operations Quality Assurance (FOQA) enable systematic collection and analysis of operational data to identify trends, deviations from standard procedures, and potential safety risks before incidents occur [3, 4]. These programs have become integral components of Safety Management Systems (SMS) in commercial aviation, with regulatory bodies including ICAO, EASA, and FAA recommending or mandating their implementation.

However, the benefits of flight data monitoring remain largely inaccessible to general aviation (GA). Professional FDR and Quick Access Recorder (QAR) systems require substantial investment, complex integration with aircraft avionics, and ongoing maintenance costs that are prohibitive for most GA operators [5]. Furthermore, regulatory frameworks such as ICAO Annex 6 and EUROCAE ED-112A impose stringent certification requirements on flight recording equipment, including crash survivability tests that significantly increase device costs [6, 7].

This gap in safety infrastructure is particularly concerning given that general aviation experiences significantly higher accident rates compared to commercial operations. Experimental amateur-built (E-AB) aircraft, training operations, and recreational flying would benefit substantially from systematic flight data collection and analysis capabilities [8, 9]. The availability of affordable flight data recording solutions could enhance pilot training, support proactive maintenance, and improve overall safety culture in the GA community.

Recent advances in microelectromechanical systems (MEMS) sensor technology, embedded computing platforms, and cloud infrastructure have created opportunities for developing low-cost alternatives to professional flight recording equipment [10, 11]. While such systems cannot achieve the certification standards required for commercial aviation, they can provide valuable data for training, research, and safety analysis in environments where certified equipment is not mandated.

This paper presents the design and implementation of an integrated flight data recording and analysis platform specifically targeted at general aviation applications. The system combines a low-cost hardware module based on MEMS sensors and Arduino microcontroller with a modern web application for cloud-based data storage, visualization, and analysis. The primary objectives are: (1) to demonstrate the feasibility of applying FDM principles using affordable, non-certified equipment; (2) to provide a functional platform for flight training organizations and researchers; and (3) to establish an extensible architecture supporting future enhancements in sensor fusion and analytical capabilities.

The remainder of this paper is organized as follows: Sect. 2 reviews related work in flight data recording systems and low-cost solutions for general aviation. Section 3 describes the system architecture including hardware components and software platform. Section 4 discusses data processing approaches and sensor measurement models. Section 5 presents experimental results and system evaluation. Section 6 concludes with discussion of limitations and future research directions.

2 Related Work

2.1 Professional Flight Data Recording Systems

Professional Flight Data Recorders represent the foundation of aviation safety infrastructure. Their design, implementation, and operation are governed by international standards, primarily ICAO Annex 6 - Operation of Aircraft, which defines mandatory installation requirements and minimum recorded parameter specifications [1]. Technical requirements for crash protection are specified in EUROCAE ED-112A, mandating resistance to mechanical impact, high temperatures, hydrostatic pressure, and prolonged submersion [6]. The FAA maintains equivalent requirements through Technical Standard Order TSO-C124c [7].

The primary purpose of FDRs is providing reliable data for accident investigation. Consequently, these devices are constructed as highly resilient systems with limited data accessibility during normal operations. Data extraction typically requires physical removal of the recorder from the aircraft, limiting utility for continuous operational monitoring [2, 12]. Modern FDRs can record hundreds to thousands of parameters at high sampling rates, but this capability comes at substantial cost prohibitive for general aviation operators.

Quick Access Recorders (QAR) were developed to support FDM and FOQA programs by providing easier access to flight data without the crash protection requirements of FDRs. QAR systems record operationally relevant parameter subsets with emphasis on regular post-flight data access [5]. While more affordable than FDRs, QAR integration still requires direct connection to aircraft avionics buses, limiting applicability to certified aircraft with appropriate data interfaces. Data mining techniques applied to QAR data in civil aviation demonstrate potential for improving operational safety and flight status assessment [23].

2.2 Low-Cost Recording Solutions for General Aviation

Research into affordable flight data recording solutions has intensified with the availability of low-cost MEMS sensors and embedded computing platforms. Saxena et al. demonstrated flight data acquisition systems using MEMS sensors in small UAVs, achieving cost reductions of over 90% compared to professional equipment while maintaining adequate accuracy for operational analysis [10]. Zhang and Kumar explored MEMS-based instrumentation for general aviation data logging, confirming viability for non-certified applications [11]. Stephan et al. developed low-cost sensor-based integrated airdata and navigation systems specifically designed for general aviation aircraft [24].

A significant contribution by Bonadonna et al. proposed automated systems for experimental amateur-built (E-AB) aircraft to record specific in-flight measurements during initial testing phases [8]. This approach reduces errors and inconsistencies while supporting pilot operating handbook development and FAA documentation compliance. The system demonstrates practical benefits of low-cost recording even in safety-critical contexts. Similarly, Sherry et al. conducted comprehensive evaluation of general aviation flight data recorders, establishing baseline performance expectations [25].

Kim et al. developed low-cost lightweight flight data recorders specifically for general aviation aircraft that have been installed on fleet airplanes and tested over extensive flight hours, demonstrating practicality and effectiveness in real-world operations [26]. Barclay proposed miniature flight data recorders providing modular solutions for GA, aiming to improve flight training, proactive maintenance, and accident investigation with capabilities tailored for light aircraft [27].

Low-cost Attitude and Heading Reference Systems (AHRS) have been evaluated for hazardous state detection in general aviation aircraft. While measurement errors exist compared to certified equipment, Chakraborty demonstrated these systems provide valuable information for identifying potential safety issues such as dangerous roll angles [13]. Huang and Johnson explored ADS-B data as an alternative source for flight data monitoring, offering affordable continuous monitoring capabilities without dedicated recording hardware [14].

2.3 Real-Time Transmission and Cloud-Based Systems

Real-time data transmission capabilities have been incorporated into modern low-cost systems. Qin et al. developed a flight data recorder utilizing the BeiDou navigation satellite system for immediate data access, enabling faster search and rescue response and timely accident analysis [15]. The system is compact, easy to install, and supports safety management, operation quality assurance, and accident investigation by providing timely and accurate data.

Zubairi developed the Flight Data Tracker (FDT) system using Software Defined Radio (SDR) for real-time data capture and cloud-based transmission, supporting immediate access and analysis [16]. Cloud computing and big data technologies integrated into flight test data processing platforms allow efficient management and analysis of large data volumes, enhancing application value by providing timely responses to requirements [28].

Wireless data retrieval capabilities have become standard in modern low-cost FDRs. Leon and Quiros developed self-contained flight data recorders with wireless data retrieval eliminating physical connection requirements [29]. Heng et al. proposed flight data recording systems with replaceable memory cards and wireless transmission options [30]. Cao and Yv developed quick-access recorders using wireless transmission to automatically upload flight data post-landing, minimizing manual intervention [31].

Greer and Nelson explored Iridium satellite devices for improving general aviation safety through global connectivity [32]. The iBos system offers modular solutions for real-time data review during flight tests, supporting both onboard and ground-based workstations while minimizing telemetry bandwidth requirements [33]. Seymour proposed remote aircraft flight recorder and advisory systems enabling continuous monitoring [34].

2.4 Intelligent Data Processing and Machine Learning

Modern flight data analysis increasingly leverages machine learning and automated processing techniques. Malaek and Alipour developed Intelligent Flight Data Recorders (I-FDRs) employing machine learning algorithms for situational awareness enhancement

and hazard prediction, representing a step toward new generation learning aircraft [17]. These systems analyze flight data in real-time, offering predictive insights and anomaly detection capabilities.

Yang et al. developed automated cockpit voice and flight data analysis systems using speech recognition and machine learning for efficient batch processing, enabling quick verification of crew checklist compliance [18]. This approach significantly reduces manual analysis workload while improving operational safety verification. Li et al. proposed advanced data mining techniques for flight data analysis in civil aviation applications [23].

Bromfield et al. evaluated Commercial-off-the-shelf (COTS) technologies including GPS and AHRS to support Flight Operations Quality Assurance (FOQA) programs for corporate aviation, providing cost-effective solutions for monitoring flight safety in lower weight category aircraft [35]. Flight augmentation systems equipped with optical sensors capture real-time data from aircraft instruments non-intrusively, supporting long-term data analysis and operational history tracking [36].

2.5 Tablet-Based and Consumer Electronics Solutions

The integration of consumer electronics into aviation data collection represents a significant cost reduction opportunity. Adams and Merrett developed combined Flight Management Systems and FDRs for general aviation using tablet computers, offering real-time performance calculations and warnings for stall or near-stall conditions [22]. These systems measure aircraft attitude and acceleration using integrated sensors without requiring extensive aircraft modifications.

Electronic logbook systems have replaced traditional paper-based record-keeping, enabling efficient management of flight hours, maintenance tasks, and compliance documentation. Kelemen et al. developed and evaluated electronic logbook systems for light aircraft operations, demonstrating alignment with modern aviation standards and improved operational efficiency [21]. Kelly and Schliessler proposed flight time-log systems supporting digital transformation in aviation documentation [37].

Kartik proposed ejectable recorder systems that can be deployed in crash events, ensuring data retrieval even in inaccessible locations [38]. Rahaman et al. developed flight data recorder location tracking using GPS, GSM, and vibration sensors for accident scenarios [39]. These approaches address the challenge of data recovery in severe accident conditions without full crash-protection costs.

2.6 Safety Analysis and Training Applications

Flight data records enable systematic safety analysis and training improvement. Patel et al. proposed methods for defining nominal trajectories for general aviation approach and landing operations using flight data records, enhancing safety assessments and flight instruction [40]. Laing de Leon and Quiros developed self-contained recording systems supporting pilot training programs [29].

Moura et al. conducted systematic literature review identifying key insights from pre-flight planning for safety improvement in general aviation, highlighting the importance of

data-driven approaches [41]. Dutta et al. developed distributed real-time flight test data review systems enabling efficient monitoring during experimental flights [33]. These applications demonstrate the broad utility of flight data beyond accident investigation.

Despite significant advances in individual components, relatively few studies address integration of hardware recording, wireless transmission, cloud storage, and web-based visualization into unified systems for general aviation. The present work addresses this gap by combining MEMS-based data acquisition, real-time cloud transmission, and comprehensive web analytics in a single compatible framework specifically designed for training and research applications.

3 System Architecture

3.1 Hardware Platform

The hardware component is built around the Arduino UNO WiFi Rev2 microcontroller platform, which integrates the u-blox NINA-W102 wireless module supporting IEEE 802.11 b/g/n connectivity. This integration eliminates external communication modules, reducing complexity, improving reliability, and lowering power consumption. The modular architecture enables future expansion with additional sensors or communication interfaces (Fig. 1).

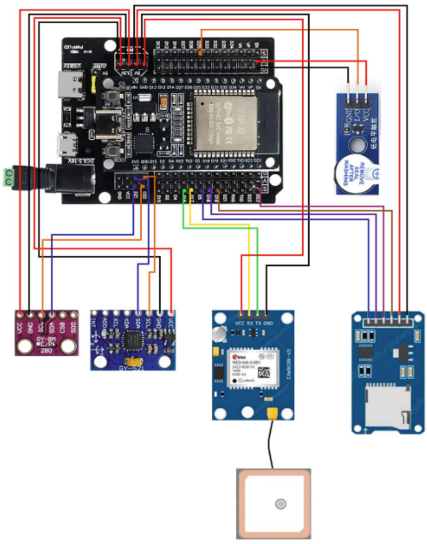


Fig. 1. Schematic connection of components of the proposed device

The inertial measurement unit (IMU) is provided by the MPU6050 sensor, combining a three-axis accelerometer and three-axis gyroscope in a single MEMS package. The accelerometer measures linear acceleration along orthogonal axes, including both

dynamic motion components and static gravitational acceleration, enabling device orientation determination. The gyroscope provides angular velocity measurements for roll, pitch, and yaw analysis. Sensor fusion of accelerometer and gyroscope data yields stable orientation estimates essential for flight dynamics analysis and turbulence detection.

Atmospheric measurements are provided by the BMP280 digital barometric sensor, utilizing piezoresistive MEMS technology for pressure and temperature measurement. Barometric altitude calculation using the pressure-altitude relationship provides higher vertical resolution than GPS for short-term altitude changes, particularly valuable for climb, descent, and turbulence analysis. Temperature measurements enable pressure corrections and provide supplementary flight condition information.

Geographic positioning is provided by the NEO-6M GPS module using trilateration from satellite signals. The module provides latitude, longitude, altitude, ground speed, and precise time stamps essential for trajectory reconstruction and map visualization. GPS time stamps enable synchronization of all sensor data streams for coherent multi-parameter analysis.

Local data storage uses a microSD card module, providing redundancy against wireless communication failures. Data is recorded in regular intervals as time series files suitable for post-flight analysis or delayed cloud upload. An acoustic buzzer provides immediate audible alerts for detected critical states including rapid descent, excessive G-forces, or sensor failures.

A box was modelled for the devices and printed on a 3D printer. The box design was implemented in a CAD environment with regard to the precise placement of individual components and ensuring their mechanical stability. Thanks to 3D modelling and simulation, it was possible to improve and modify the box design in a digital version before its production on demand [42]. The design was adapted to the dimensional devices and the possibilities of additive manufacturing, 3D printing of additional production and flexible adjustments during development.

3.2 Software Platform

The software architecture implements a three-tier design following separation of concerns principles. The backend layer, implemented in Java using Spring Boot framework, provides REST API interfaces for data communication, handles flight record processing, validation, and database storage, and implements security mechanisms including JWT-based authentication and role-based access control.

The frontend layer utilizes React JavaScript library for responsive user interface implementation. Users access map-based flight visualization, graphical parameter analysis for position, altitude, temperature, pressure, acceleration, and turbulence, and statistical summaries. The interface adapts to various device types including desktop and mobile platforms.

The database layer implements a relational model with three primary entities: User (account information and access roles), Flight (individual flights with metadata including date, distance, duration), and FlightRecord (detailed time-series records including GPS coordinates, altitude, pressure, temperature, acceleration, and detected events). This structure supports efficient querying, aggregation, and future analytical module integration.

The complete system is containerized using Docker, ensuring consistent environments across development, testing, and production deployments. Docker Compose enables single-command deployment, simplifying operation and maintenance. The modular codebase architecture facilitates extension with new sensor types, analytical algorithms, or visualization components (Fig. 2 and Table 1).

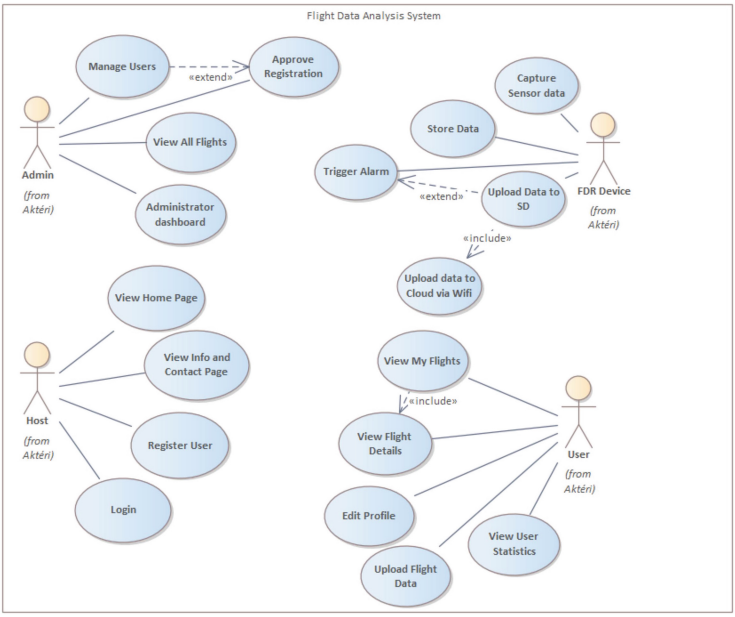


Fig. 2. Use Case diagram

Table 1. Comparison of FDR, QAR, and proposed system characteristics

Criterion	FDR	QAR	Proposed System
Primary purpose	Accident investigation	FDM/FOQA analysis	Training, research
Regulatory framework	ICAO, ED-112A, TSO	ICAO (recommended)	Non-certified
Parameters recorded	Hundreds to thousands	Tens to hundreds	Basic flight parameters
Data source	Aircraft avionics	Aircraft avionics	External MEMS sensors
Crash protection	Extreme	Standard	None
Data access	Physical removal	Post-flight download	Real-time + offline
Cost	Very high	Medium	Low

4 Data Processing and Sensor Models

The sensor measurement models directly impact data model design, processing approaches, and visualization interpretation. Recorded data are represented as discrete time series where each record is bound to timestamp t. The web application processes data in the form $\{x_m(t)\}_{t=1}^{N}$, where $x_m(t)$ represents measured value at time t and N is the sample count within a flight.

Time synchronization is critical since sensors generate data at different frequencies and latencies. The GPS module provides reference time, with each FlightRecord stored with timestamp enabling subsequent combination of IMU, barometric, and GPS data into unified temporal frames. Visualization uses unified time axes to minimize interpretation errors from signal offsets.

Random measurement noise n(t) manifests as high-frequency fluctuations that degrade graph readability and complicate event identification. Pre-processing including moving average or low-pass filtering approximates true signal x(t) from measurement x_m(t) = x(t) + b + n(t), where b represents systematic bias. The goal is increased interpretation robustness rather than complete noise removal, resulting in stable parameter graphs supporting reliable event detection.

Systematic error (bias) b differs fundamentally from random noise. While noise causes short-term deviations, bias manifests as long-term drift. Gyroscope drift during angular velocity integration leads to progressive orientation estimate error. The system therefore evaluates orientation not in isolation but in multi-sensor context, comparing acceleration components with trajectory and altitude changes. The application presents analytical framework for trend and event identification rather than certified absolute values.

Altitude determination combines barometric and GPS measurements. Barometric altitude from pressure using the standard atmosphere equation provides smooth, sensitive vertical movement data but is affected by atmospheric condition changes. GPS altitude has higher random error but provides globally referenced data and time synchronization. The application uses barometric measurement as primary source for altitude profile analysis while GPS data support map visualization and temporal framework. This sensor redundancy reduces incorrect interpretation risk. Figure 3 shows the user view of the created web application on a map record of the flight path with summary statistics.

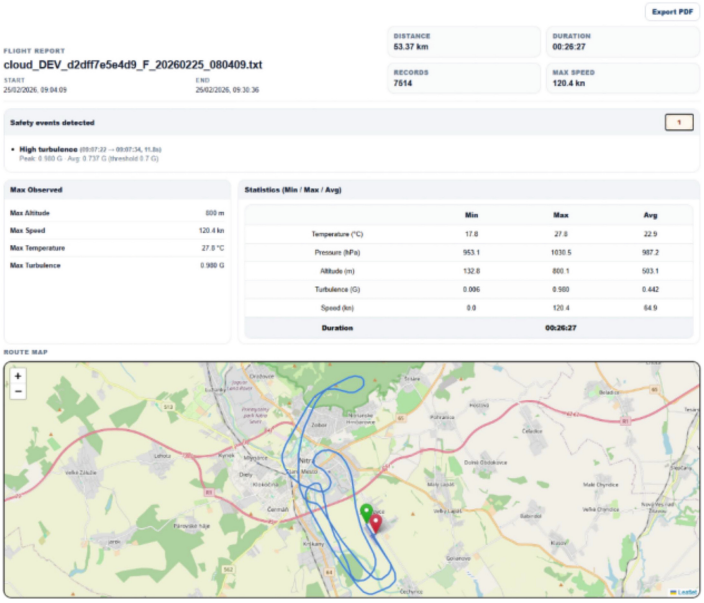


Fig. 3. User view web application interface of flight details - map and summary statistics

Event detection (rapid descent, excessive G-force, turbulence) applies measurement models through threshold design accounting for bias and noise components. Detection typically operates on short time windows or parameter combinations rather than single samples, reducing false positive rates from random noise. The database model with Flight and FlightRecord entities stores time series data for frontend retrieval and visualization through graphs and map layers. The visualization of graphs for turbulence and speed is shown in Fig. 4.

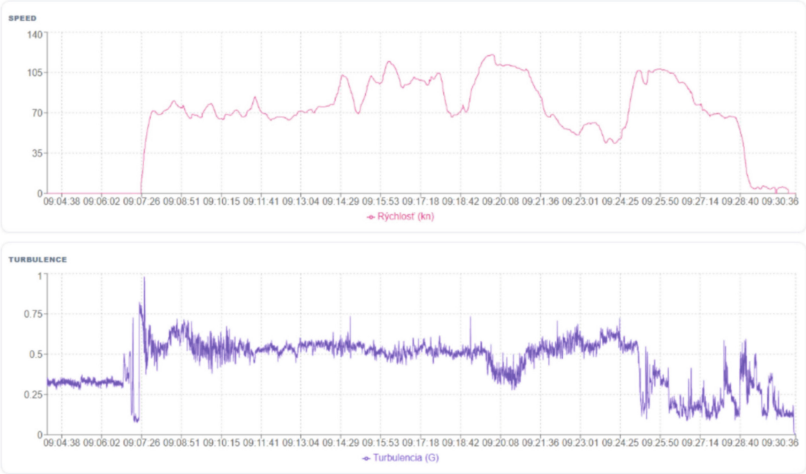


Fig. 4. Generated graphs of speed and turbulence from web application

The Kalman filter was not implemented due to system objectives focusing on retrospective analysis and visualization rather than real-time navigation. Kalman filtering requires precise system dynamic models and sensor noise statistics. With low-cost MEMS sensors whose noise characteristics vary with temperature, vibration, and mounting conditions, filter implementation without thorough calibration could produce misleading results. Simpler processing methods including time windows, thresholding, and basic sensor fusion provide transparent interpretation sufficient for FDM objectives while maintaining computational feasibility on resource-constrained hardware.

5 Discussion

The proposed system demonstrates that Flight Data Monitoring principles can be effectively applied through low-cost hardware combined with web-based analytics. Implementation results indicate that non-certified MEMS sensors can capture representative flight parameter time series enabling dynamic change identification and non-standard event detection. While recorded data are affected by random and systematic sensor errors including inertial drift and barometric variability, these limitations do not preclude analytical utility since the approach focuses on trend evaluation and relative changes rather than absolute precision.

The decision to forego Kalman filter implementation reflects the system's retrospective analysis focus and absence of precise dynamic models. Simpler signal processing methods provide transparent interpretation sufficient for general aviation FDM while enabling implementation on limited hardware resources. Multi-tier software architecture with REST API and containerization supports scalable deployment and straightforward integration of additional analytical modules.

From a regulatory perspective, the system does not meet ICAO, FAA, or EUROCAE requirements for crash-protected flight recorders including ED-112A testing standards.

The system is therefore not suitable for certified commercial operations. Its value lies in experimental, training, and research applications where it enables analytical procedure development and flight data monitoring without certified avionics modification. This positioning aligns with ICAO Global Aeronautical Distress and Safety System (GADSS) trends emphasizing flight data availability beyond physical recorders.

The system architecture creates a foundation for future development including advanced sensor fusion, adaptive analytical methods, and enhanced visualization. Results suggest that similar low-cost approaches can effectively complement professional systems in domains without certification and extreme precision requirements. Potential applications include pilot training organizations, experimental aircraft testing, academic research, and recreational aviation safety enhancement.

6 Conclusion

This paper presented the design and implementation of a low-cost flight data recording and analysis system targeting general aviation applications. The solution combines modular MEMS-based hardware with a cloud-connected web platform enabling Flight Data Monitoring principles application without certified equipment or aircraft system modifications. The system utilizes sensor measurement models, time series processing, and robust event detection based on time windows, hysteresis, and multi-parameter combination.

Key contributions include: (1) demonstration of affordable FDM implementation for general aviation; (2) integrated hardware-software architecture supporting real-time transmission and offline analysis; (3) scalable web platform with visualization and user management capabilities; and (4) practical validation of MEMS sensor applicability for flight training and research contexts.

Future research directions include sensor fusion algorithm refinement, machine learning integration for automated event classification, and expansion to additional flight phases and aircraft types. The open architecture supports community extension and adaptation to specific operational requirements, potentially contributing to improved safety culture across general aviation.

Acknowledgments. This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and the Slovak Academy of Sciences (SAS) under contract No. VEGA 1/0492/26.

This work was supported by the project KEGA (020TUKE-4/2026): An interactive classroom as a means of supporting the education of aviation cybernetics systems

References

1. International Civil Aviation Organization: Annex 6—Operation of Aircraft. ICAO, Montreal (2018)
2. EUROCAE: ED-112A: minimum operational performance specification for crash protected airborne recorder systems. Paris (2013)

3. Robbins, P.T., et al.: Flight data monitoring program implementation and benefits. Aerospace. **8**(3), 1–11 (2021)
4. Patel, M.R., Roberts, D.G.: Proactive use of flight data monitoring to improve aviation safety. IEEE Aerosp. Electron. Syst. Mag. **34**(4), 40–49 (2019)
5. Yang, J., Wang, D.: Flight operations quality assurance and quick access recorder applications. In: IEEE Int. Conf. on Big Data, pp. 125–130 (2018)
6. EUROCAE: ED-112A crash protection standards for airborne recorders. Paris (2013)
7. Federal Aviation Administration: Technical standard order TSO-C124c: flight data recorder systems. Washington, DC (2019)
8. Bonadonna, C., et al.: Design of a low-cost general aviation flight data recording and analysis system. In: AIAA Aviation (2015)
9. Vidović, A., et al.: Flight data recording solutions for general aviation safety enhancement. Transp. Res. (2022)
10. Saxena, S., et al.: Low-cost flight data acquisition with MEMS sensors in small UAVs. IEEE Sensors J. **19**(15), 5780–5788 (2019)
11. Zhang, H., Kumar, A.: MEMS based instrumentation for general aviation data logging. Int. J. Aerosp. Eng. (2020)
12. Mani, A.C., Agarwal, A.K.: Critical data analysis of flight data recorder for safety of flight. Int. J. Eng. Res. Technol. **7**(4), 1 (2018)
13. Chakraborty, I.: Hazardous state detection in general aviation using low-cost AHRS. In: AIAA SciTech (2019)
14. Huang, C., Johnson, M.E.: Exploring ADS-B as alternative data source for flight data monitoring. J. Aviat. Technol. **35**(2) (2017)
15. Qin, H., et al.: Development of flight data recorder for general aviation with real time transmission. Chin. J. Aeronaut. (2020)
16. Zubairi, J.A.: Flight data tracker: real-time cloud-based transmission system. In: IEEE DASC (2019)
17. Malaek, S.M., Alipour, A.: Intelligent flight data recorders: a step toward new generation learning aircraft. In: IEEE A&E Systems (2022)
18. Yang, R., et al.: Design of a cockpit voice data analysis and processing system. In: EMEIT (2024)
19. Burmester, G., et al.: Big data and data analytics in aviation. In: Advances in Aeronautical Informatics. Springer (2018)
20. Wang, C.: Application and development of flight data monitoring in aviation safety. J. Comput. Electron. Inf. Manag. **13**, 10–14 (2024)
21. Kelemen, M., et al.: Development and evaluation of electronic logbook system for light aircraft operations. Acta Avion. **27**(1), 37–43 (2025)
22. Adams, J.W., Merrett, G.M.: Combined flight management system and FDR for general aviation using tablet computers. In: IEEE DASC (2016)
23. Li, B., et al.: Overview of application of data mining techniques to QAR data analysis. In: ICMSE (2022)
24. Stephan, M., Florian, H., Gottfried, S.: Low-cost sensor based integrated airdata and navigation system for general aviation aircraft. In: IEEE/AIAA DASC (2008)
25. Sherry, J.H., et al.: Evaluation of a general aviation flight data recorder. In: IEEE DASC (2007)
26. Kim, D., et al.: Low-cost lightweight flight data recorders for general aviation aircraft. In: IEEE DASC (2002)
27. Barclay, A.: Miniature flight data recorders. Aviation Today (2005)
28. Dong, G., et al.: Flight test data processing and analysis platform based on new generation information technology. In: CSAA (2022)

29. Leon, H.L., Quiros, R.: Self-contained flight data recorder with wireless data retrieval. US Patent (2001)
30. Heng, L., Ren, Q., Chen, Y.: Flight data recording system and flight data reading method. CN Patent (2019)
31. Cao, H., Yv, K.: Quick-access recorder based on wireless transmission. In: CGNCC (2010)
32. Greer, M.J., Nelson, G.A.: Improving general aviation safety using low-cost iridium devices. In: IEEE DASC (2012)
33. Dutta, A., et al.: Distributed real time flight test data review system for general aviation. In: IEEE DASC (2024)
34. Seymour, L.: Remote aircraft flight recorder and advisory system. US Patent (1996)
35. Bromfield, M., et al.: Evaluation of COTS solutions to support FOQA in business/corporate aviation. In: IEEE DASC (2018)
36. Guida, N.R.: Non-intrusive flight data collection and analyzation with flight automation systems. US Patent (2021)
37. Kelly, R., Schliessler, U.: Flight time-log system. US Patent (2000)
38. Kartik, S.: Ejectable flight data recorder unit. Indian Patent (2017)
39. Rahaman, O.K.A., et al.: FDR/CVR location tracker using GPS, GSM, and vibration sensor. In: ICEECCOT (2023)
40. Patel, T.G., et al.: General aviation approach and landing analysis using flight data records. In: IEEE DASC (2016)
41. Moura, N.L., Trindade, N.F., Manuela, A.: Key insights from preflight planning for safety improvement in general aviation: a systematic review. Appl. Sci. **14**(9), 3771 (2024)
42. Kovari, A.: Industry 5.0: generalized definition, key applications, opportunities and threats. Acta Polytech. Hung. **21**(3), 267–284 (2024)

Algorithmic Trading and Computational Intelligence in Finance

Gamified Financial Education: Human vs. Deep Learning Trading Agents

Maros Cavojsky(✉), Filip Hodon, and Martin Drozda

Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Bratislava, Slovakia

maros.cavojsky@stuba.sk

Abstract. This paper introduces a gamified learning environment where human traders move beyond passive observation to compete directly against deep learning agents. Using simulations based on market data from 2025, human traders test their intuition against Multilayer Perceptron (MLP) models in direct competition for trading performance. Our findings highlight a distinct performance divergence: while deep learning agents capitalize on systematic patterns for high profitability, human traders demonstrate greater adaptability during market reversals and structural price shifts. By exposing behavioral biases like loss aversion through competitive benchmarking, this approach provides a rigorous, repeatable framework for teaching algorithmic market dynamics and risk management in finance.

Keywords: Algorithmic Trading · Deep Learning · Gamification · Financial Education · Human-AI Competition

1 Introduction

Algorithmic trading is now central to modern financial markets, with deep learning models accounting for a majority of trade executions [9]. However, the complexity of these models often makes them difficult for human traders to understand. Traditional education tends to focus on theory or static backtesting, which does not always capture the pressure and real-time decision-making of active trading.

In this paper, we describe an educational method based on gamification—using game design elements in non-game contexts. We focus on a web-based simulation where users manage a portfolio while competing against MLP-based trading agents. This setup serves two purposes: it provides a training ground for risk management and helps explain the behavior of AI models by allowing human traders to observe and compete with them in different market conditions.

This work contributes to the development of simulation environments and market simulators for financial research, while advancing training methodologies for AI and computational intelligence in finance. We use the engagement of competition to encourage human traders to study market indicators and the underlying mechanics of automated trading systems.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 587–598, 2027.
https://doi.org/10.1007/978-3-032-32104-6_49

2 Related Work

2.1 Deep Learning for Financial Time Series

The application of deep learning in financial forecasting has evolved from simple Multi-Layer Perceptrons to complex architectures designed to capture spatiotemporal dependencies [20]. Convolutional Neural Networks (CNNs), traditionally used for image processing, have been successfully adapted for financial time series through 1D convolutions or 2D image-like representations of price charts [11]. Recent benchmarks have also explored Transformers [10,16,18], leveraging self-attention mechanisms to prioritize significant historical events. However, debate remains regarding the necessity of Transformers for univariate and multivariate time series, with some studies suggesting that simpler linear-based models like StockMixer [5] or TSMixer [2] can achieve comparable performance with lower computational cost [19]. Our work leverages the CNN-based architecture and noise-filtering properties of Renko charts introduced in [4], which allows for robust feature extraction by permuting indicator arrangements.

2.2 Gamification and Serious Games in Finance

Gamification involves the integration of game-design elements in non-game contexts to enhance motivation and engagement [13,17]. Early versions of these platforms focused on reward-based systems (e.g., badges and leaderboards), but modern implementations have transitioned toward high-fidelity simulation and "affective gamification". Baptista and Oliveira [1] demonstrate that engagement in these "fintech" environments is a critical mediator for knowledge retention, where humans learn through iteration rather than static observation.

2.3 Behavioral Finance and the Need for Interactive Training

A core justification for gamified training lies in the prevalence of behavioral biases that lead to suboptimal investment decisions. Retail traders are often susceptible to overconfidence, the disposition effect (selling winners too early and holding losers too long), and herd mentality [7,12]. Interactive environments allow human traders to encounter these biases in a risk-free setting. However, there is also a recognized "dark side" to gamification; features like confetti-animations and constant notifications can inadvertently encourage excessive risk-taking and noisy trading strategies [15]. By placing human traders in competition with pre-trained AI agents, our methodology aims to expose these biases and teach risk management through direct benchmarking against non-emotional counterparts.

2.4 Human-AI Collaboration in Trading

The rapid expansion of algorithmic trading has shifted the research focus from fully autonomous systems to human-AI collaboration (HAIC). This synergy

leveraging both the "System 1" (instinctive) and "System 2" (analytical) thinking of human traders [6,8] alongside the rapid data processing capabilities of AI is central to modern trading strategies. Our proposed web-based game explores this synergy by allowing users to compare their strategic decisions against automated trades, turning the performance comparison into a transparent educational tool.

3 Methodology

The proposed system consists of two primary components: the deep learning trading agents and the interactive web-based interface.

3.1 Market Simulation and Data Integrity

To ensure the accuracy of the simulation and prevent participants from "cheating" by referencing historical market data, several data obfuscation techniques are employed. The simulation environment is built upon historical stock prices from the year 2025, but the data is transformed before being presented to the user:

- **Ticker Randomization:** Original stock symbols are replaced with generated four-letter identifiers. For the trials involving human traders we select following tickers: AACT, CWKG, DHNZ, GRPM, KENF, KFPL.
- **Noise Injection:** A small, controlled amount of random noise is added to the price series to ensure they do not exactly match publicly available datasets.
- **Price Scaling:** Price levels are multiplied by random coefficients, though the underlying volatility and relative dynamics remain intact, ensuring that technical indicators still reflect realistic market behavior.

While the platform contains a pool of over 500 obfuscated stock data, a specific subset of six tickers is selected for the trials involving human traders to facilitate a rigorous and standardized evaluation of the results. These scenarios encompass a broad range of market dynamics: from aggressive uptrends with high risk profiles (e.g., CWKG and KENF with gains exceeding 190% but maximum drawdowns near -40%) to more stable growth patterns (AACT, KFPL) and volatile sideways regimes (DHNZ, GRPM). The descriptive statistical characteristics of these tickers is summarized in Table 1.

3.2 MLP-Based Trading Agents

The MLP-based trading agents (hereafter referred to as AI agents) follow the architecture proposed in [4], which is optimized for time-series-to-image conversion using Renko charts. Unlike standard candlestick charts, Renko bricks filter out market noise by only recording significant price movements, independent of time. The input is an $n \times m$ matrix where each row represents a normalized technical indicator (e.g., RSI, MACD, ADX) over m Renko bricks. This spatial

Table 1. Descriptive Statistics of Stock Scenarios

Ticker	Gain (%)	Gain (USD)	Mean (USD)	Std. Dev.	MDD (%)	Range Min, Max
AACT	+15.8%	+29.01	192.00	16.38	−16.4%	[164.4, 226.9]
CWKG	+194.6%	+217.43	247.99	102.59	−47.7%	[97.5, 435.3]
DHNZ	+5.0%	+18.34	360.03	28.21	−30.7%	[276.6, 418.2]
GRPM	+13.2%	+65.40	474.40	54.38	−30.1%	[352.9, 584.7]
KENF	+237.7%	+406.94	269.95	115.42	−40.7%	[127.6, 580.9]
KFPL	+19.2%	+266.20	1488.19	114.62	−18.5%	[1192.4, 1676.9]

Note: All scenarios comprise 249 daily candles from the year 2025.

representation allows the MLP to identify complex visual patterns that correlate with market reversals or trend continuations.

Notably, the output layer uses a linear activation function to predict a reward value α, representing the expected discounted return. Trading actions are derived from α using a threshold-based strategy:

$$action = \begin{cases} Buy & \text{if } \ell_{max} \geq \alpha > \ell \\ Sell & \text{if } -\ell_{max} \leq \alpha < -\ell \\ Neutral & \text{otherwise} \end{cases} \tag{1}$$

where ℓ and ℓ_{max} are parameters tuned to capture pronounced moves while avoiding overbought/oversold extremes where the market may already be saturated.

3.3 Gamified Learning Interface

The educational platform, built using a backend stack consisting of PHP, MySQL, and ClickHouse [3] for high-performance time-series data storage, integrates the simulation engine with game design elements. For interactive charting, the system utilizes Lightweight Charts by TradingView [14], providing a professional-grade experience for market analysis. The core loop of the "game" is as follows:

1. **Scenario Selection:** The system provides the user with an environment covering the entire year 2025 across multiple tickers, allowing for a continuous simulation of market regimes without the need for manual timeframe selection.
2. **Trading Session:** The user interacts with a price chart, making daily decisions to place market orders, maintain existing positions, or remain neutral. Simultaneously, the MLP agent's trades are executed based on its pre-trained weights. To ensure a fair benchmark, a mandatory stop-loss is implemented for the user, matching the specific risk-management configuration of the competing AI agent (e.g., 10%).

3. **Interactive Feedback:** Real-time performance indicators, such as the current profit and loss (PnL), active positions, and total equity for both the user and the AI agent, are displayed in Fig. 1. This live comparison serves as a core gamification mechanic, allowing human traders to monitor their relative standing throughout the trading session.
4. **Performance Review:** After the session, the user can review their results in a side-by-side comparison with the AI agent. The system provides a visual overlay of both equity curves alongside detailed metrics, including final PnL, total trade count, average trade duration, and maximum win/loss. This granular feedback allows human traders to identify behavioral biases, such as "overtrading" or "loss aversion," by contrasting their high-frequency or long-duration patterns with the AI's systematic execution.

The gamification is further reinforced through social benchmarking; after the session concludes, participants are ranked on a global leaderboard to compare their performance metrics against other human traders.

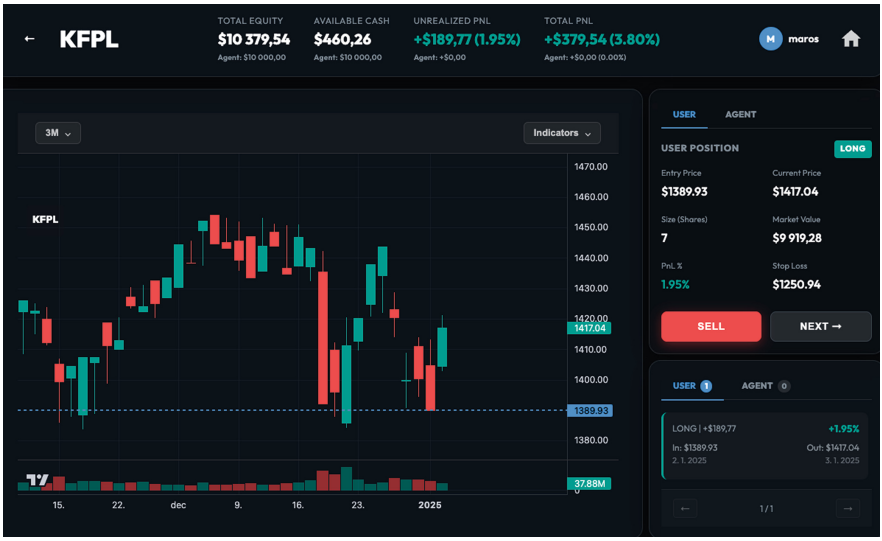


Fig. 1. The interactive web interface showing the gamified trading environment.

4 Experimental Results and Discussion

While the AI agents are trained on historical data from 500 stocks spanning the period 2016 to 2023, the experimental evaluation is conducted using 2025 market data to ensure out-of-sample validity. Performance is analyzed across several key metrics, including ROI, Sharpe Ratio, Maximum Drawdown (MDD), and trade frequency, by comparing a pool of AI agents with human traders' performance in "matched games" (scenarios selected by users at the start of each session).

4.1 Quantitative Performance Comparison

Summary statistics for the performance comparison are presented in Table 2. On a global scale, representing 1620 simulation runs across diverse architectures, AI agents achieve a wide variety of outcomes. However, when we isolate the 80 matched games where users compete directly against a pre-selected AI agent, the AI agents achieve a mean ROI of 94.04%, significantly outperforming the human average of 36.92%.

Table 2. Detailed Summary Statistics: Performance Comparison

Group	Games	Mean ROI%	Median ROI%	ST-DEV %	Max ROI%	Min ROI%	Avg Sharpe
Human	80	36.92	16.13	73.36	312.90	−44.63	**1.5774**
AI (Matched)	80	94.04	56.78	107.03	426.38	−34.64	1.5671
AI (Global)	1620	19.33	8.58	44.10	426.38	−59.69	0.4705

The statistical comparison in Table 2 reveals a significant performance gap. In matched games, AI agents achieved a mean ROI of 94.04%, more than double the human trader average of 36.92%. However, the standard deviation for the AI agent (107.03%) confirms a higher volatility compared to human traders (73.36%). Crucially, when comparing risk-adjusted metrics, the results are remarkably balanced: human traders achieve an average Sharpe Ratio of 1.58, while their direct AI competitors in the same scenarios reach 1.57. This suggests that while the pool of AI models varies greatly (Avg Sharpe 0.47), the selection of AI models utilized for human benchmarking are highly optimized for stability, matching human intuition for risk management.

4.2 Comparative ROI Histogram

The histogram of AI Agent trading performance shown in Fig. 2 can create the impression that the top-performing results are likely to be reproducible in the future. The histogram follows a power-law-like structure: a strong peak around the 0–10% ROI range, accompanied by a significant "long tail" of highly profitable outliers reaching up to 426.38%.

The high variance observed in the ROI histogram provides noteworthy educational value, illustrating that architectural choice does not independently guarantee consistent profitability. The histogram demonstrates that even when trained on identical datasets, MLP-based models can yield divergent outcomes due to stochastic factors—such as weight initialization—and variations in decision thresholds for order execution. This variability highlights a fundamental challenge in algorithmic trading: a configuration that achieves significant gains on one asset may incur draw-downs on another. By exposing human traders to this inconsistency, the environment fosters a more analytical perspective on the limitations of autonomous systems and the risks of overfitting in financial forecasting.

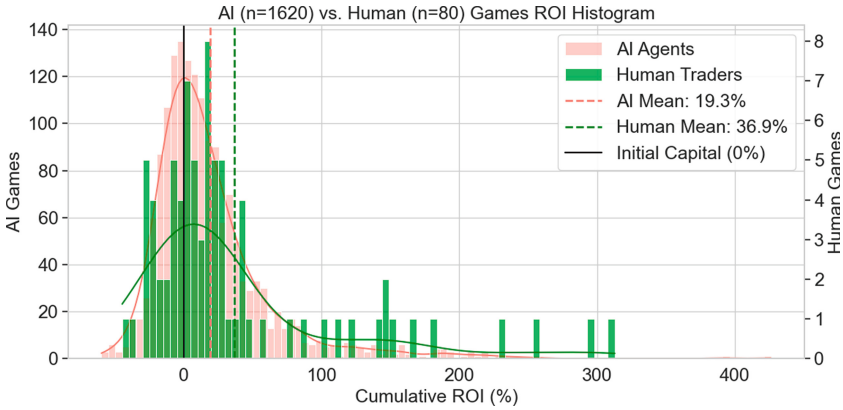


Fig. 2. ROI histogram across 1620 simulation seeds. Note the extreme outlier performance in the right tail.

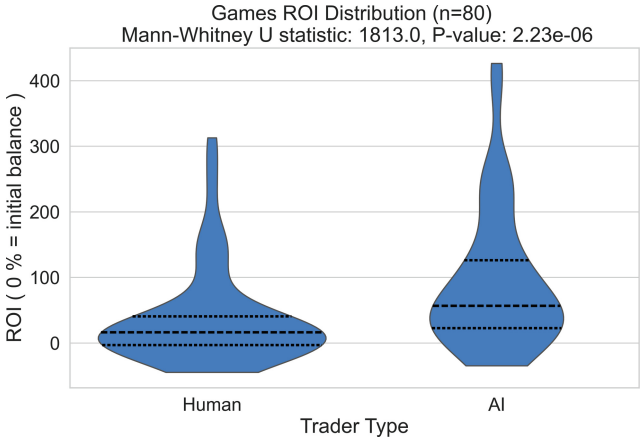


Fig. 3. Probability density comparison for matched sessions. The histogram shows a wide dispersion of results, suggesting that humans naturally gravitate toward high-volatility scenarios.

In contrast to the global baseline, the ROI density histogram for the sessions specifically matched with human traders (Fig. 3) exhibits a significantly flatter and more dispersed profile. This increased variance highlights a combination of factors: while human traders gravitate toward high-volatility scenarios, the variance of the histogram is primarily driven by behavioral biases. The wider dispersion in the negative tail suggests that many human traders, lacking professional experience, frequently hold losing positions in the hope of a reversal, only to have them terminated by the mandatory stop-loss. Conversely, the failure to timely realize profits during market turnarounds, often driven by an expectation that an uptrend will persist indefinitely, results in sharp pullbacks that erode

realized returns. This emotional decision-making process contrasts sharply with the systematic, rule-based execution of the AI agents.

4.3 Ticker-Based Performance Analysis

Evaluation of trading results across specific tickers (Fig. 4) reveals distinct market regimes that heavily influenced the competition.

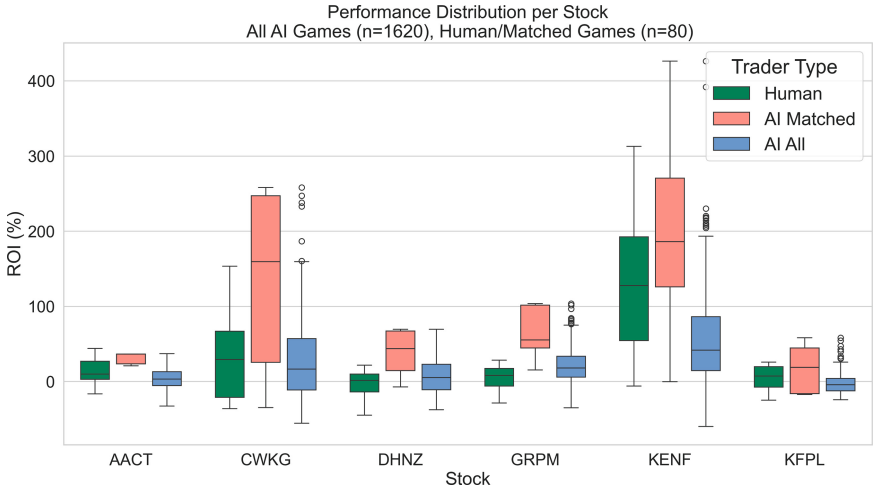


Fig. 4. Ticker-specific variance analysis comparing AI and human performance across selected scenarios. Tickers like DHNZ and CWKG show high AI profitability but also higher risk dispersion.

For tickers characterized by sustained trends, such as **DHNZ**, the systematic nature of the AI agents allowed them to exploit recurring patterns effectively, resulting in a dense clustering of outcomes above the median return.

In contrast, human traders often struggled with the intra-day volatility of these instruments, with equity curves frequently descending to the mandatory stop-loss threshold. However, in market scenarios exhibiting non-trending or mean-reverting price action, human traders occasionally demonstrated superior performance by maintaining a neutral stance. During these periods, the AI's sensitivity to localized features often led to excessive signal generation, resulting in PnL erosion due to combined transaction costs and whipsaw effects. These results suggest that while AI excels at systematic trend-following, human traders maintain an edge in identifying structural shifts that favor capital preservation over aggressive execution.

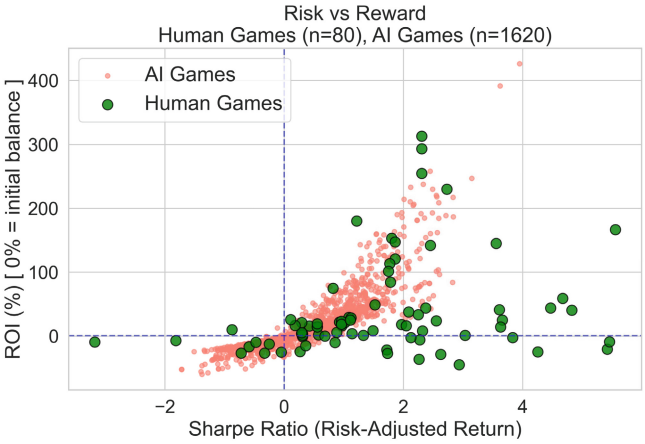
4.4 Analysis of Human Trading Behavior

The simulations reveal several classic behavioral biases among the human traders. Despite being in a "risk-free" educational environment, the psychological impact of seeing a virtual balance decrease remains powerful.

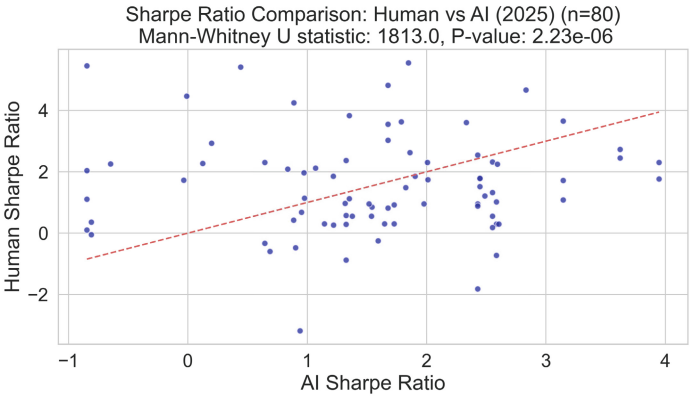
- **Loss Aversion:** Human traders frequently hold losing positions longer than the AI, hoping for a "bounce" back to break-even. The mandatory stop-loss is implemented specifically to mitigate this, yet its frequent triggering among human traders (compared to AI) shows a failure to manually "cut losses" earlier.
- **Revenge Trading:** Following a stop-loss event, human traders show a marked increase in trade frequency, often attempting to "re-gain" lost capital in a high-volatility regime where the AI correctly remains neutral.
- **Overconfidence:** Following a successful win, users often become less selective and trade more impulsively on weaker signals. This shift away from a disciplined strategy usually results in preventable losses in subsequent trades.

These findings are summarized in the risk-reward analysis presented in Fig. 5. The distribution highlights a sharp contrast in behavior: while the AI data points in the scatter plot (Fig. 5a) are tightly clustered, reflecting strict adherence to a systematic trading logic, the human traders exhibit significant dispersion across all quadrants. This clustering for the AI games demonstrates that regardless of the market scenario, the agents maintain a reasonable risk profile, effectively decoupling the execution strategy from emotional influence. In contrast, the wide scattering of human traders' results suggests that human traders frequently oscillate between extreme risk-taking and overly cautious behavior, often within the same session. This high variance across diverse market regimes underscores the role of individual psychological factors and inconsistent decision-making, providing a visual proof that human traders' performance is often driven by internal states rather than a stable, rule-based framework.

The risk-adjusted performance analysis is further supported by the head-to-head comparison shown in Fig. 5b. In this scatter plot, each data point represents a specific "matched game" played by both a human trader and an AI agent, with the coordinates corresponding to their respective Sharpe Ratios. This direct comparison reveals that while the AI agents demonstrate a remarkably consistent performance band, reflecting the systematic nature of their training, the human results are widely dispersed across the spectrum. The Mann-Whitney U test confirms a highly significant difference between the two groups ($p = 2.23 \times 10^{-6}$), statistically validating that the AI agents and human traders operate under fundamentally different performance regimes. Even in identical market scenarios, human risk-adjusted returns vary significantly, which reinforces the educational takeaway: professional deep learning models prioritize systematic stability and repeatable risk management, whereas human performance remains highly dependent on individual intuition, emotions, and inconsistent execution.



(a) Risk-Reward Scatter Plot.



(b) Sharpe Ratio Comparison.

Fig. 5. Risk-adjusted performance analysis. Note the high-density AI clustering in contrast to the wide dispersion of human results.

4.5 Educational Impact

The main goal of this environment is not to teach users to beat professional algorithmic funds, but to show how automated trading systems behave.

By encouraging competition, the platform highlights the differences in strategy. A human might see a price chart and expect a price increase, while the AI agent (processing the same Renko bricks) decides to sell. In the post-session analysis, the human can evaluate where their intuition diverged from the agent's actions. The realization that the AI lacks market context, news awareness, or a broader outlook is the primary educational benefit. This project tracks how

human understanding shifts from viewing AI as a "magic box" to seeing it as a predictable, though powerful, statistical tool.

5 Conclusion

We presented a gamified approach to AI education in finance, showcasing the competition between human traders and MLP-based trading agents. Our experimental results, supported by highly significant statistical evidence ($p = 2.23 \times 10^{-6}$), reveal a fundamental behavioral gap: while AI agents maintain tight, rule-based consistency, human traders' performance is characterized by high variance driven by individual intuition and emotional factors. Although AI agents often achieve higher absolute returns, human traders remain competitive in risk-adjusted terms during specific market regimes. Future work will focus on dynamic model selection, exploring how human micro-contextual insights or automated meta-controllers can be used to select the most appropriate trading models and parameters tailored to the current market regime.

Acknowledgments. This work was supported by the Grant for Innovative Teams at the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, under the Faculty Call in accordance with the Dean's Directive No. 4/2024.

References

1. Baptista, G., Oliveira, T.: Why so serious? Gamification impact in the acceptance of mobile banking services. Inf. Manage. **56**(1), 118–139 (2019)
2. Chen, S.A., Li, C.L., Yoder, N., Arik, S.O., Pfister, T.: Tsmixer: an all-MLP architecture for time series forecasting. CoRR (2023)
3. ClickHouse: Fast open-source online analytical database management system. https://clickhouse.com/. Accessed 11 Feb 2026
4. Drozda, M., Cavojsky, M.: On suitability of Renko charts for algorithmic trading. In: Proceedings of DiVAI 2024 (2024)
5. Fan, J., Shen, Y.: Stockmixer: a simple yet strong MLP-based architecture for stock price forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 8389–8397 (2024)
6. Goldberg, L.R.: Thinking, fast and slow, by Daniel Kahneman. Quant. Finance **13**(1), 15–19 (2013)
7. Hirshleifer, D.: Behavioral finance. Annu. Rev. Financ. Econ. **7**, 133–159 (2015)
8. Kahneman, D.: Thinking, Fast and Slow. Farrar, Straus and Giroux (2011)
9. London School of Economics and Political Science: Is ai changing the stock market? LSE Research for the World (2023). https://www.lse.ac.uk/research/research-for-the-world/ai-and-tech/ai-and-stock-market. Accessed 11 Feb 2026
10. Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: long-term forecasting with transformers. CoRR (2022)
11. Sezer, O.B., Ozbayoglu, A.M.: Algorithmic financial trading with deep convolutional neural networks: time series to image conversion approach. Appl. Soft Comput. **70**, 525–538 (2018)

12. Sloan, R.G.: Do stock prices fully reflect information in accruals and cash flows about future earnings? Account. Rev. 289–315 (1996)
13. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press (2018)
14. TradingView: Lightweight charts: High-performance interactive financial charts. https://www.tradingview.com/lightweight-charts/. Accessed 11 Feb 2026
15. Van Der Heide, D., Želinský, T.: The dark side of gamification: financial platforms and the 'gamblification' of retail trading. New Media Soc. **23**(12), 3574–3592 (2021)
16. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
17. Werbach, K., Hunter, D.: For the Win: How Game Thinking can Revolutionize your Business (2012)
18. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, pp. 11121–11128 (2023)
19. Zhou, Y., Alon, U., Chen, X., Wang, X., Agarwal, R., Zhou, D.: Transformers can achieve length generalization but not robustly (2024)
20. Zou, J., et al.: Stock market prediction via deep learning techniques: a survey. CoRR (2023)

Cumulative True Range for Benchmarking of Trading Algorithms

Martin Drozda(✉) and Maros Cavojsky

Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Bratislava, Slovakia
{martin.drozda,maros.cavojsky}@stuba.sk

Abstract. This study evaluates trading algorithm efficiency by comparing simulated trading performance against theoretical metrics, specifically Cumulative True Range (CTR) and High-Low Pivot detection. Notably, despite its potential for benchmarking, CTR has surprisingly remained absent from existing literature. We examine these approaches across both fixed-position (without reinvestment) and compounding reinvestment frameworks using the Profit Capture Ratio (PCR), which measures the ratio of realized returns to theoretical potential. Our findings suggest that weekly CTR without reinvestment is a robust performance metric. It efficiently captures how an algorithm extracts profit amidst market volatility and unpredictability.

Keywords: Algorithmic Trading · Benchmarking · Cumulative True Range · High-low Pivot Algorithm · Williams Fractals · Renko Charts · Kagi Charts

1 Introduction

When comparing two trading algorithms that earned, for example, 100% and 10% in profits across distinct assets, absolute returns alone do not reveal how well each performed. To understand their true efficiency, one must estimate the maximum profit that could have been captured. This is illustrated in Fig. 1, where the price action of SPY (S&P500 ETF) and LQD (Investment Grade Corporate Bonds ETF) is shown.

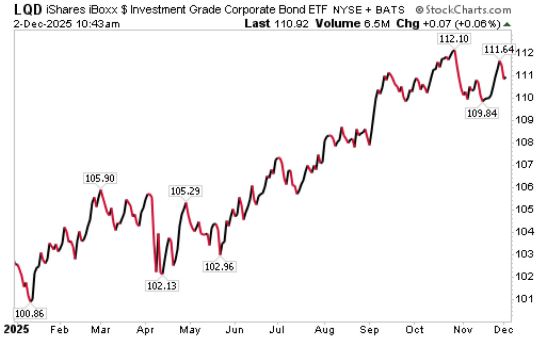
From the figure we can estimate that for SPY we could capture the price action from 575.45 to 683.39, or 18.76%, if we just buy and hold, or we could capture 99.65% in profits if we could correctly predict the price trajectory $575.45 \rightarrow 607.61 \rightarrow 546.63 \rightarrow \cdots \rightarrow 683.39$ (assuming that we can trade both long and short, all profits get reinvested, and minor price changes get ignored).

Similar statistics can be computed for LQD, where the buy-and-hold would yield a 10.69% profit and if we correctly predict the price trajectory $100.86 \rightarrow 105.90 \rightarrow 102.13 \rightarrow \cdots \rightarrow 111.64$ we could capture 29.63% in profits.

In both cases, it must be acknowledged that achieving these returns might be unrealistic in practice, as real-world trading does not allow us to identify local minima and maxima with perfect accuracy.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 599–610, 2027.
https://doi.org/10.1007/978-3-032-32104-6_50

(a) SPY



(b) LQD

Fig. 1. Comparison of SPY and LQD (both charts courtesy of StockCharts.com).

The open question that we study in this paper is what would be a *fair estimate of price action that can be captured by a trading algorithm*. We argue that Cumulative true range is a reasonable candidate, as it reflects the maximum price movement over a given period while discarding small intra-period fluctuations that would be difficult for an algorithm to exploit. Cumulative true range is derived from True range introduced by Wilder [10]. Usually, True range gets averaged over a certain number of time steps (the default is often set to 14) in order to compute Average true range [10], which is a standard volatility indicator.

Let us assume that for the time period i, the price action is represented as $OHLC_i$, where O_i = open price, H_i = highest price, L_i = lowest price, and C_i = close price. True range TR_i and Cumulative true range CTR are defined as follows:

$$TR_i = \max\{H_i - L_i, |H_i - C_{i-1}|, |L_i - C_{i-1}|\},$$

$$CTR_a^b = \sum_{i=a}^{b} TR_i,$$

where the aggregation of the time period i is, for example, 15-minutes, 30-minutes, 1-hour, 1-day, 1-week, or 1-month. The advantage of CTR is that it is straightforward to implement in many trading and testing environments, as it only requires summing a list of TR values.

To ensure comparability across different assets and price scales, we normalize the CTR by dividing it by the initial asset price. The normalized CTR is defined as:

$$CTR_{norm} = \frac{CTR_a^b}{P_a}, \tag{1}$$

where P_a is the initial price of the asset. This represents the price movement captured by a unit investment (1) over the specified period, while any profits captured are not reinvested.

Definition 1 (Profit Capture Ratio). *Let $\Pi_{\text{alg}} \geq 0$ denote the realized profit of the algorithm over a given time period, and let $\Pi_{\max} > 0$ denote the maximum theoretically attainable profit under the same price trajectory. We define the profit capture ratio (or profit capture efficiency) as*

$$PCR = \frac{\Pi_{\text{alg}}}{\Pi_{\max}} = \frac{\Pi_{\text{alg}}}{CTR_{norm}}. \tag{2}$$

The quantity $PCR \in [0, 1]$ measures the fraction of extractable profit that the algorithm successfully captures. We use CTR_{norm} as an estimate of Π_{max}.

In the above definition, we exclude cases where $\Pi_{\text{alg}} < 0$, as our focus is strictly on well-performing trading algorithms. Similarly, we disregard instances where $\Pi_{\max} = 0$, as this condition implies a stagnant price trajectory wherein the price neither increased nor decreased.

2 Related Work

Prior work in this area is limited. Several online charting platforms, such as stockcharts.com, employ a window-based method that identifies local extrema (pivots) by computing the maximum and minimum closing prices over a symmetric window [1]:

$$\max\{C_{j-L}, \ldots, C_j, \ldots, C_{j+L}\} \quad \text{and} \quad \min\{C_{j-L}, \ldots, C_j, \ldots, C_{j+L}\},$$

where L is a user-defined parameter. A key limitation of this approach is that confirming a local maximum or minimum requires L future observations, rendering it unsuitable for real-time applications. However, this look-ahead bias does not invalidate the method for historical performance evaluation, as the unconfirmed pivots at the end of the time series can simply be excluded from the total profit calculation. The pseudocode for this algorithm is shown in Algorithm 1. An instance of this algorithm is also used by stockcharts.com; see Fig. 1.

A related concept is Williams Fractals [11] that detect local extrema using a fixed window of size $L = 2$. A bullish fractal (local low) occurs at time i if the

```
Input  : Price series P[1..N], look-back L.
Output: Set of raw pivots: (j, P[j], HIGH/LOW).
pivots ← ∅;
for j ← L to N − L do
    window ← P[j − L ... j + L];
    mx ← max(window);
    mn ← min(window);
    if mx ≠ mn then
        if P[j] = mx then
            pivots ← pivots ∪{(j, P[j], HIGH)};
        else if P[j] = mn then
            pivots ← pivots ∪{(j, P[j], LOW)};
        end
    end
end
Sort pivots by index j;
return pivots;
```

Algorithm 1: High/Low Pivot detection Using Look-back Window

low at i is lower than the lows of the two preceding and two succeeding time steps. A bearish fractal (local high) occurs at time i if the high at i is higher than the highs of the two preceding and two succeeding time steps. Fractals are confirmed only after the future time steps are observed and are commonly used to highlight short-term turning points in price series.

Another alternative is the Zig–Zag algorithm [1] that identifies significant price turning points by filtering out movements below a predefined threshold. Let $\delta > 0$ be a reversal threshold. A pivot is confirmed when the price reverses by at least δ from the last confirmed pivot. During an uptrend, the algorithm tracks the maximum price and declares a HIGH pivot once the price declines by δ. During a downtrend, it tracks the minimum price and declares a LOW pivot once the price rises by δ. The resulting sequence of pivots alternates between highs and lows and captures only movements exceeding the specified threshold.

The threshold δ can also be defined dynamically as $\delta_t = k \cdot \mathrm{ATR}_t$, allowing the Zig–Zag to adapt to changing market volatility.

Yet another alternative are Japanese Kagi charts [8] that represent price movements using a sequence of vertical lines whose direction changes only when the price reverses by more than a predefined threshold. An upward line is drawn while prices continue to rise, and a downward line is drawn while prices continue to fall. A trend change is confirmed when the price reverses by a specified amount, commonly defined as a fixed value, a percentage, or an ATR-based threshold. By ignoring minor price fluctuations, Kagi charts emphasize major trend reversals and are therefore well suited for detecting changes in market trend.

Point-and-Figure (P&F) [4] charts represent price movements using columns of Xs and Os and ignore the passage of time entirely. A column of Xs denotes

rising prices, while a column of Os denotes falling prices. Price movements are discretized using a fixed box size, and a reversal occurs only when the price moves by a predefined number of boxes. Pivots can be computed by identifying column reversals, which correspond to significant changes in price direction exceeding the reversal threshold. Compared to Kagi charts, which draw continuous vertical lines and change direction based on a single reversal amount, Point-and-Figure charts use discrete columns and explicitly quantize price changes, resulting in a more structured representation of support, resistance, and pivot levels.

3 Problem Formulation and Analysis

3.1 Cumulative True Range

The basic property of the Cumulative True Range (CTR) is expressed by the following geometric inequality:

$$|C_i - C_{i-1}| \leq \max\{|H_i - C_{i-1}|, |L_i - C_{i-1}|\} \leq TR_i.$$

Since the closing price C_i is strictly bounded by the period's high H_i and low L_i, the net price change from the previous close C_{i-1} cannot exceed the maximum excursion relative to C_{i-1}. The True Range, by definition, captures this maximum possible deviation, accounting for both intraday volatility and inter-day gaps. Summing this relationship over a time horizon yields:

$$\sum_{i=a}^{b} |C_i - C_{i-1}| \leq \sum_{i=a}^{b} TR_i = CTR_a^b, \tag{3}$$

demonstrating that the CTR strictly dominates the sum of absolute close-to-close returns and thereby represents a more comprehensive measure of the total extractable price movement.

Inequality 3 is valuable for performance estimation of trading algorithms because it establishes a strict dominance relation between directional price displacement and total price volatility. Using CTR_a^b ensures that the algorithm is evaluated against complete price path rather than just price displacement.

3.2 Profit Sources of Trading Algorithms

When analyzing a trading algorithm, we can decompose its total profit into three distinct sources:

1. Volatility (V): The profit generated by capturing price variance along the asset's price path:

$$V = \kappa_0 \sum_{t=0}^{T} \epsilon_t,$$

where ϵ_i represents the price capture of individual volatility swings and κ_0 is the initial fixed number of shares, which does not change over time. If we

assume that asset prices follow a log-normal distribution, a standard assumption in continuous-time finance [3], the absolute magnitude of these swings (ϵ_i) scales proportionally with the price level.

2. Implicit Scaling (S): The portion of returns attributable to the directional price trend rather than active trading:

$$S = \kappa_0(P_T - P_0).$$

While V measures the profit extracted from price fluctuations, S represents the "pseudo-compounding" effect where absolute gains increase as the underlying asset price rises, even with a constant share count.

3. Physical Compounding: The active reinvestment of realized gains to increase the share count κ_t:

$$\kappa_t = \kappa_{t-1} + \frac{P_t - P_{t-1}}{P_t}.$$

Taking the above three effects into consideration, we can differentiate between two basic trading approaches:

1. Fixed Position (Non-Reinvesting): This setup utilizes only Volatility and Implicit Scaling. The profit Π_1 is driven by the asset's price trajectory while maintaining a fixed number of shares κ_0:

$$\Pi_1 = \kappa_0 \left(\Delta P + \sum \epsilon_t \right),$$

where $\Delta P = P_T - P_0$.

2. Variable Position (Reinvesting): This setup incorporates all three sources. By making N a dynamic variable, the strategy moves from linear share-growth to an exponential equity curve:

$$\Pi_2 = \sum_{t=1}^{T} \kappa_t(\Delta P_t + \epsilon_t),$$

where $\Delta P_t = P_t - P_{t-1}$.

3.3 Research Problem Formulation

Our research goal is to evaluate trading algorithm efficiency by comparing simulated trading performance against theoretical metrics, specifically Cumulative True Range (CTR) and High-Low Pivot detection. By analyzing these approaches under both fixed-position and reinvesting (compounding) frameworks, we aim to quantify the Profit Capture Ratio (PCR) as defined in Eq. 2. This comparison serves to validate the accuracy of heuristic profit potential metrics relative to realized trading outcomes.

4 Experimental Setup

Our methodology utilizes a trading algorithm based on Renko charts [5], a technique distinguished by its ability to decouple price movement from time. By requiring a predefined price change (the Renko height) to occur before a new unit is plotted, Renko charts effectively act as a low-pass filter, neutralizing market noise and isolating significant price trends [8]. This algorithm can be summarized as follows:

1. OHLC input data are converted to Renko charts.
2. OHLC(V) input data are used to compute various technical indicators [1] such as MACD, RSI, OBV etc. Subsequently, all computed technical indicators are normalized [5]. Several technical indicators, such as On-Balance Volume (OBV), also incorporate the volume component (V) as an input. Each indicator can be controlled by various input parameters; the applied parameters are shown in Table 1.
3. Calculated indicators are filtered based on Renko membership, with only the final value in each Renko being utilized. Note that each Renko contains at least a single OHLCV data point; thus, it is always possible to compute technical indicators within a Renko. This approach effectively filters out intermediate indicator values that do not coincide with the formation of a new Renko bar. The Renko height is determined by ATR, i.e., we adjust the height of each Renko with respect to observed volatility.
4. Renko charts allow for a specific way to label training/validation data. Let us assume that we are looking at the Renko chart shown in Fig. 2. It shows a specific way of computing the reward associated with this chart, where an upward Renko is valued +1 and a downward Renko is valued -1. In this case, the computed reward (label) would be 2.0.
   In general, future rewards can be discounted and defined in the following way:

$$reward = \sum_{t=0}^{\rho} \gamma^t r_t,$$

   where $0.0 < \gamma \leq 1.0$ is the discount rate, $r_t \in \{-1, +1\}$ is the reward derived from the i-th Renko orientation, and ρ is the time horizon for future reward computation ($\rho = 7$ in Fig. 2). This approach to computing discounted labels is motivated by reinforcement learning, where the discounted return needs to be computed in order to decide which action should be selected next [9].
5. A matrix $k \times ind$ is constructed, where k is the look-back time period and ind is the number of applied technical indicators. Each matrix is assigned a label. This constitutes the input to a machine learning approach. In this study, we set $k = 25$ and $ind = 23$.
6. The model architecture consists of an initial Batch Normalization layer applied along the indicator dimension, followed by two fully connected (Dense) layers of 256 and 128 units, respectively. Both hidden layers utilize *tanh* activation and a dropout rate of 5% for regularization. This MLP

(Multilayer perceptron) neural network concludes with a final Dense output layer designed for future reward regression.

7. The MLP architecture is trained using a Hybrid Mean Squared Error (MSE) loss function. It is designed to penalize absolute deviations while simultaneously rewarding directional alignment between the predicted return $\hat{y}$ and the ground truth y:

$$\mathcal{L}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 - \lambda\left[\frac{1}{n}\sum_{i=1}^{n}\operatorname{sgn}(y_i)\cdot\tanh(\hat{y}_i)\right], \tag{4}$$

where λ is a regularization hyperparameter that controls the influence of the directional agreement. In this study, we set $\lambda = 0.1$. We use the Adam optimizer [7] with the learning rate set to 0.002.

8. While the completion of a Renko is non-deterministic in live trading, our approach computes predictions for every OHLCV data point. This methodology is called candlestick iteration [5] and ensures continuous evaluation regardless of whether a new bar has fully formed.
9. The trading strategy is controlled by the future reward prediction. The algorithm assumes a directional position based on the sign of this forecasted return: the system enters a long position when a positive reward is predicted and a short position when a negative reward is predicted. To maximize capital efficiency, all profits and losses are computed with *full reinvestment*, resulting in a compounded equity curve. This ensures that the position size scales dynamically with the current portfolio value rather than remaining fixed.

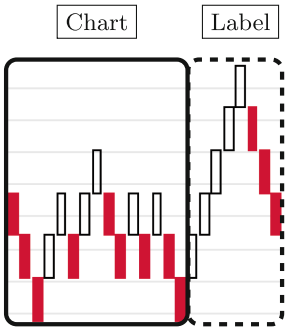


Fig. 2. Label for a chart is computed from future Renkos [5].

We use Alpaca as the source of historic OHLCV data [2]. Alpaca is a stock broker that focuses on algorithmic trading. To train the MLP architecture, we utilize 30-minute OHLCV data for S&P500 companies as well as several ETFs (Exchange traded funds). We utilize OHLCV data from 2016–2022 for training, 2023 for validation (including early-stop), and 2024–Oct. 2025 for final testing and performance evaluation.

Table 1. The list of applied indicators and their parameters.

Indicator	Parameters
RSI	window = 20
MACD	(12,26,9)
SMA	$\{20, 50, 100, 200\}$
EMA	$\{20, 50, 100, 200\}$
TTM squeeze	window = 20, ATR window=20, multiplier = 1.5
pivot pocket	length = 10, size = 1.3
ROC	window = 20
Bollinger bands	window = 20, std. dev. = 2.0
Keltner channels	window = 20, ATR window = 10, multiplier = 2.0
TSI	slow window = 25, fast window = 13
MFI	window = 14
VWAP	window = 14
ATR	window = 14
ADX	window = 14
OBV	no parameters
CMF	window = 14
AROON	window = 25

Since the applied technical indicators are normalized, the MLP models trained on 30-minute OHLCV data remain scale-invariant, allowing them to be applied directly to 1-day OHLCV data for daily return forecasting.

The Cumulative Total Return (CTR) is computed using both daily and weekly aggregations in order to capture volatility across different temporal scales. High/Low Pivot detection algorithm is only computed with daily aggregation.

Our code is implemented in Python 3.10. We use Keras 3.11.3 [6] to implement the neural network architecture. Keras is a high-level neural networks API that can be used with Python.

We use a PC with Ubuntu OS 22.04.4 LTS, Intel i9-13900KF CPU, 128GB DDR5 RAM, 4TB M.2 SSD and NVIDIA GeForce RTX 4090 GPU, which has 24GB GDDR6X VRAM and 16,384 stream processors.

5 Experimental Results

Our experimental results are shown in Table 2. The performance is computed for several companies such as AAPL (Apple), AMZN (Amazon), MU (Micron), TSLA (Tesla) and ETFs such as GLD (Gold), NUGT (Gold Miners Index Bull 2X), QQQ (Nasdaq Index), SPY (S&500 Index), TLT (20+ Year Treasury Bond) and XLU (Utilities Select Sector Index). The table presents the performance of the MLP architecture alongside the buy-and-hold (B&H) benchmark, the

Table 2. Theoretical Max. Performance (%) and Profit Capture Efficiency (PCR). **Algo** = Trading algorithm performance (%); **B&H** = Buy and hold (%); **Pivot** = High/Low Pivot detection (%).

Ticker	Algo	B&H	Compounding			Fixed position			PCR
			—CTR—		Pivot	—CTR—		Pivot	
			Daily	Weekly		Daily	Weekly		
AAPL	75.2	5.1	6,478.8	929.1	979.2	376.4	218.0	214.4	0.34
AMZN	84.0	7.3	11,820.5	1,123.1	1,220.2	505.5	267.0	280.8	0.31
GLD	54.8	54.1	751.3	278.5	304.5	266.9	160.2	166.3	0.34
MU	720.5	79.2	30,564.2	2,209.5	1,957.6	641.9	348.1	330.2	**2.07**
NUGT	**1,104.1**	**179.7**	118,987.9	4,748.2	3,366.6	**1,192.5**	**561.7**	475.5	1.97
QQQ	64.2	11.8	1,395.9	351.6	437.5	264.9	156.7	164.0	0.41
SPY	44.0	10.7	584.5	183.5	223.6	189.9	107.2	121.1	0.41
TLT	27.4	7.0	394.3	146.1	190.7	164.9	94.5	109.5	0.29
TSLA	356.9	24.4	**916,740.5**	**12,479.5**	**16,235.5**	882.6	506.1	**521.9**	0.70
XLU	61.8	21.5	1,026.3	225.7	274.5	264.0	128.5	140.9	0.48

CTR for both weekly and daily aggregations, the results of the High/Low Pivot detection algorithm, and the PCR (see Eq. 2).

The performance metrics are evaluated over a duration of approximately ten to eleven months, primarily from late 2024 through October 2025. The exact start and end dates depend on the Renko calculation for each ticker; because each Renko can aggregate multiple OHLCV data points, the effective trading period can shift by several days.

The strong performance of NUGT, a leveraged ETF, is primarily attributable to the sharp appreciation in gold prices during the observation period. Performance efficiency, measured with the Profit Capture Ratio (PCR) reached a peak of 2.07 for MU, whereas the lowest efficiency was recorded for TLT at 0.29.

We further observe that lower volatility, as estimated by the weekly CTR, correlates with diminished algorithmic performance. In these instances, the restricted fluctuations of the underlying price provide fewer opportunities for profit extraction. This can be observed in the case of TLT that has the Weekly CTR equal to 109.5 and PCR equal only to 0.29.

The results show that High/Low Pivot detection and Weekly CTR yield nearly identical outcomes for both fixed-position and compounding scenarios. Given its computational simplicity, CTR is the primary candidate for evaluating an algorithm's profit-capture efficiency.

A paired Wilcoxon signed-rank test confirms that the differences between Weekly CTR and High/Low Pivot detection are not statistically significant for either the compounding ($p = 0.32$) or fixed-position ($p = 0.49$) approaches, validating CTR as a robust proxy for theoretical maximum profit capture.

Additionally, our results indicate that the use of daily or weekly compounding produces unrealistically high performance estimates. Consequently, we conclude that non-compounding models provide a far more accurate representation of actual trading algorithm performance. It should be noted that this study does not focus on absolute returns; rather, our objective is to provide insight into the efficiency with which a trading algorithm captures available market profit.

6 Conclusion

This study evaluates the efficiency of a directional trading algorithm based on a MLP architecture across a selection of equities and ETFs. By defining the Profit Capture Ratio (PCR) as the ratio of realized algorithmic profit (Π_{alg}) and the weekly CTR (CTR_{norm}) without compounding, we established a metric for performance that is independent of absolute returns and instead focuses on capture efficiency.

The comparative analysis between compounded and fixed-position returns reveals that both daily and weekly compounding yield mathematically sound but practically unrealistic performance estimates (e.g., for TSLA $> 900,000\%$). These results should be interpreted as a theoretical upper bound, achievable only when we correctly predict each minor price change. Such a success rate is highly improbable in live markets; it represents a 'crystal ball' prediction scenario rather than a reproducible trading outcome.

The results reveal a surprisingly large divergence between the compounding approaches and the trading algorithm performance. We argue that this is a structural consequence of the algorithm's discrete sampling intervals (1-day). By operating on fixed timeframes, the system incurs an unavoidable 'confirmation cost', sacrificing immediate execution to ensure trend validity. Quantifying this cost and optimizing the trade-off between signal latency and trend confirmation will be the subject of our future research.

Acknowledgments. This research was supported by the Slovak Research and Development Agency (APVV) under Grant No. APVV-24-0446.

References

1. Achelis, S.B.: Technical Analysis from A to Z, 2nd edn. McGraw-Hill, New York (2001)
2. Alpaca (2025). https://alpaca.markets/
3. Black, F., Scholes, M.: The pricing of options and corporate liabilities. J. Polit. Econ. **81**(3), 637–654 (1973)
4. Dorsey, T.J.: Point and Figure Charting: The Essential Application for Forecasting and Tracking Market Prices, 3rd edn. Wiley, Hoboken (2007)
5. Drozda, M., Cavojsky, M., Sebes, P.: On suitability of renko charts for algorithmic trading. In: International Scientific Conference on Distance Learning in Applied Informatics, pp. 69–81. Springer, Cham (2024)

6. Keras (2025). https://www.keras.io/
7. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Nison, S.: Beyond Candlesticks: New Japanese Charting Techniques Revealed. Wiley, Hoboken (1994)
9. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press (2018)
10. Wilder, J.W.: New concepts in technical trading systems. Greensboro, NC (1978)
11. Williams, B., Gregory-Williams, J.: Trading Chaos: Maximize Profits with Proven Technical Techniques. Wiley, Hoboken (2004)

Kagi Chart Representations as Structured Inputs for Deep Learning Models

Filip Hodon(✉), Maros Cavojsky, and Martin Drozda

Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Bratislava, Slovakia
{filip.hodon,maros.cavojsky,martin.drozda}@stuba.sk

Abstract. We investigate the efficacy of Kagi chart representations as structured inputs for deep learning models in algorithmic trading. By transforming noisy financial time series into trend-dependent Kagi segments, we capture significant market movements while suppressing stochastic fluctuations. We compare fixed percentage-based and ATR-based reversal thresholds, assessing their impact on data density, training stability, and predictive performance. Our empirical results show that ATR-based thresholds provide more robust performance across daily and 30-min data. The validation on QuantConnect confirms the practical applicability of the approach. Our main contribution is the demonstration that Kagi segment duplets offer a compact, information-dense encoding suitable for neural network models.

Keywords: Kagi chart · Algorithmic trading · Deep learning · Financial time series · Average true range · Input data encoding

1 Introduction

Financial time series are inherently noisy, non-stationary, and dominated by short-term fluctuations that often obscure the underlying market structure. Traditional time-based representations, such as OHLCV (open-high-low-close-volume) charts, preserve temporal ordering but provide limited mechanisms for separating meaningful price dynamics from stochastic noise. As a result, both discretionary traders and algorithmic trading systems typically rely on additional technical indicators, which are themselves derived transformations of already noisy data.

Price-based charting techniques offer an alternative perspective by discarding the uniform time axis and focusing exclusively on price movement. Among these methods, Kagi charts [9] construct market structure using discrete price changes and explicit trend confirmation rules rather than fixed time intervals. Originally developed in Japanese technical analysis, Kagi charts encode trend persistence and reversal logic based on prior price extremes, thereby embedding structural state transitions directly into the chart representation.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 611–623, 2027.
https://doi.org/10.1007/978-3-032-32104-6_51

In this work, we investigate the suitability of Kagi-based representations as structured inputs for deep learning models in algorithmic trading. We treat Kagi charts as a state-based encoding of market dynamics, where each segment represents a validated price movement rather than a raw time sample. This representation inherently suppresses stochastic fluctuations and emphasizes directional consistency. This paper presents a comparative evaluation of two classes of reversal mechanisms: fixed percentage-based thresholds and volatility-adaptive thresholds defined through the Average True Range (ATR), with the aim of assessing their impact on the statistical and economic performance of Kagi-based trading strategies.

By allowing the reversal size to scale with market volatility, ATR-based formulations provide an implicit form of normalization and noise control. We evaluate both approaches across multiple time resolutions, including daily and 30-min data, to assess their impact on data density, training stability, and predictive performance. The primary contribution of this study is a systematic empirical evaluation of Kagi chart representations as inputs to a Multi-Layer Perceptron neural network for trend classification and trading decision support.

We analyze their ability to reduce stochastic noise, stabilize structural patterns, and improve generalization in downstream trading models. Our results indicate that ATR-adaptive Kagi constructions offer a compact and information-dense encoding of price dynamics, making them particularly well suited for data-driven trading systems where robustness and scalability are critical.

2 Related Work

The application of non-temporal charting techniques in financial forecasting has gained popularity due to their ability to filter out market noise. Kagi charts, alongside Renko [6] and Point-and-Figure charts [5], represent a class of price-driven visualizations that prioritize price action over fixed time intervals.

Steve Nison [9] first popularized Kagi charts in Western finance, highlighting their utility in identifying trend reversals through structural shifts in line thickness and direction. A critical challenge in algorithmic trading is the non-stationarity and stochastic nature of financial time series.

The research by Tsay [12] emphasizes that traditional OHLCV data often contain significant noise that can degrade the performance of machine learning models. To address this, various denoising techniques have been proposed. The use of the Average True Range (ATR), introduced by Wilder [13], has become a standard for measuring volatility and is frequently used to create adaptive trading systems. For instance, Lau et al. [8] demonstrated that incorporating volatility-based indicators can significantly improve the robustness of automated decision-making. In recent years, the focus has shifted toward integrating these traditional techniques with deep learning.

Sezer and Ozbayoglu [11] proposed a CNN-based framework that treats technical indicators as images (CNN-Pred), similar to our approach of using structured Kagi segments as input features.

The concept of state-based representation, where the market is divided into discrete regimes, is further explored by Zhang et al. [14]. They argue that identifying validated trend states rather than raw price points leads to better generalization. Furthermore, the combination of technical analysis and deep learning has been validated by several studies. For example, Huang et al. [7] showed that neural networks can effectively learn from non-linear technical indicators to predict stock market movements.

Recent literature specifically addressing price-based charting in machine learning remains relatively sparse, but the notable work by Cohen [4] suggests that price-action representations like Renko and Kagi can provide a "compressed" view of the market, which is highly beneficial for training stability in neural networks.

Additionally, the importance of hyperparameter tuning in these hybrid systems, such as choosing the optimal reversal threshold, has been highlighted by Arévalo et al. [2] in their study on dynamic technical analysis. Finally, the works of Cavalcante et al. [3] provide a comprehensive survey on computational intelligence in the stock market, underlining the necessity of sophisticated feature engineering, such as our proposed Kagi segment duplets, to achieve consistent profitability.

3 Methodology

The proposed forecasting system employs a multi-stage approach designed to transform non-stationary financial time series into structurally significant state representations. This process integrates price-action-based charting, high-dimensional feature engineering via aggregated technical indicators, and a deep Multi-Layer Perceptron (MLP) architecture optimized for regression.

3.1 Data Transformation: Adaptive Kagi Charting

The dataset consists of *Open, High, Low, Close, and Volume* (OHLCV) indicators, which capture the underlying price dynamics and liquidity over the observed period. To emphasize the structural behavior of price movements and reduce market noise, the OHLCV data obtained from Alpaca [1] are transformed into Kagi charts. This transformation filters out insignificant fluctuations and highlights meaningful trend reversals. Subsequently, the Kagi representation is further processed to construct Kagi duplets, which serve as the final input structure for the proposed analysis. Unlike time-fixed representations, Kagi charts are purely price-driven, responding only to directional movements that exceed a predefined threshold, known as the reversal threshold. The Kagi chart is initialized at the first closing price $Close_1$, with its initial direction determined by comparing $Close_1$ to $Close_2$. The line is extended as long as the price continues in the current direction. Reversal occurs only when the price moves against the current trend by at least the reversal threshold. The chart starts as a Yang (thick) line. A transition from Yang (thick) to Yin (thin) occurs when the price

falls below the most recent prior swing low. Conversely, a transition from Yin (thin) to Yang (thick) occurs when the price exceeds the most recent prior swing high. Therefore, the line is thin (Yin) when the price is in a weaker state, i.e., after breaking a previous swing low, and remains thin until a new swing high is exceeded.

In this study, we evaluate two distinct approaches for defining the reversal threshold to ensure the model's adaptability to various market conditions:

- **Fixed Percentage-Based Reversal** The reversal threshold is defined as $\alpha * P$, where P denotes the current price and $\alpha \in (0, 1)$ represents the relative reversal sensitivity parameter.
- **Volatility-Adaptive Reversal (ATR)** Instead of using a fixed percentage threshold, the reversal magnitude is defined as a function of recent market volatility. Volatility is quantified using the True Range (TR) and its moving average, the Average True Range (ATR), which together provide a scale-sensitive measure of typical price fluctuations.

Let i denote the i-th observation time in the price series, with $i = 1, 2, \ldots, T$. The True Range at time i is computed as:

$$TR_i = \max\left(High_i - Low_i,\ |High_i - Close_{i-1}|,\ |Low_i - Close_{i-1}|\right), \tag{1}$$

where $High_i$, Low_i, and $Close_i$ denote the high, low, and closing prices at time i, respectively. The Average True Range is then computed as a rolling mean of the True Range over a time window. Its size is usually 14 periods, which we also apply herein.

Based on this volatility estimate, the adaptive reversal threshold of Kagi at time i is defined as:

$$\text{Threshold}_i = \text{ATR}_i \cdot \kappa, \tag{2}$$

where κ is a sensitivity parameter that controls how many units of average volatility are required to trigger a reversal. This volatility-adaptive reversal mechanism ensures that a change in Kagi line direction is triggered only when the price movement exceeds a multiple of its recent average variability, thereby filtering out minor fluctuations and improving the signal-to-noise ratio. The resulting structural transitions between Yang (green) and Yin (red) lines are illustrated in Fig. 1.

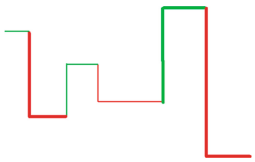


Fig. 1. Illustration of Kagi chart construction and structural transitions.

3.2 Kagi Duplet Construction and Feature Aggregation

To transform the dynamic structural changes of Kagi charts into a format suitable for supervised learning, we implement a **Duplet-based aggregation** strategy. A Kagi Duplet is defined as an atomic state representation derived from a sequence of two consecutive Kagi lines:

- **Sequence Capture:** By grouping two successive lines, the model captures local directional transitions (e.g., up-down, down-up, or trend continuation). This preserves the immediate price-action context while significantly reducing the dimensionality of the input space.
- **Max-Pooling Aggregation:** Let $\mathcal{I} = \{I^{(1)}, I^{(2)}, \ldots, I^{(M)}\}$ denote the set of technical indicators (Table 1). Furthermore, let $\mathcal{D} = \{D_1, D_2, \ldots, D_N\}$ denote the ordered sequence of Kagi duplets extracted from the time price series. Each duplet D_i spans a continuous time interval $[t_i^{\text{start}}, t_i^{\text{end}}]$ in the underlying OHLCV data, corresponding to the formation of two consecutive Kagi lines. For a given indicator $I^{(m)} \in \mathcal{I}$ and duplet D_i, we define the aggregated feature value using a max-pooling operator as:

$$I^{(m)}(D_i) \;=\; \max\big\{I^{(m)}(t) \;\big|\; t \in [t_i^{\text{start}},\, t_i^{\text{end}}]\big\}. \tag{3}$$

  This aggregation scheme captures the most extreme signal for each duplet and indicator during the Kagi structure formation. By focusing on these extremes, the resulting features highlight periods of peak momentum.

For each constructed duplet, we extract a comprehensive feature vector consisting of 23 technical indicators. As summarized in Table 1, these features are categorized into momentum, trend, volatility, and volume-based metrics to provide a holistic view of the market state. Certain technical indicators (MACD, Bollinger Bands, etc.) internally consist of multiple derived components, each of them is treated as a separate feature.

3.3 Sequential Input Preparation and Flattening

The final input for the Multi-Layer Perceptron is constructed as a temporal sequence of these structural states. We maintain a history of 15 consecutive duplets, which effectively grants the model a "structural memory" of the 16 most recent significant price movements. The resulting input tensor $\mathbf{X} \in \mathbb{R}^{15\times 23}$ is subsequently transformed via a flatten layer into a one-dimensional vector of size 345 ($15 \times 23 = 345$). This representation preserves the chronological order of market states while being entirely decoupled from uniform time, allowing the MLP to focus exclusively on the evolution of price-action structures.

3.4 Neural Network Architecture

The predictive engine is built upon a Multi-Layer Perceptron (MLP) architecture for high-dimensional regression tasks. The model processes the input through a sequential pipeline designed to extract features from a 15×23 temporal feature matrix:

Table 1. Feature Space: Technical indicators with specific window parameters.

Category	Technical Indicators (Parameters/Window)
Momentum	RSI (20), ROC (20), TSI (13, 25), Aroon, MACD(12, 26, 9)
Trend	SMA (20), EMA20, EMA50, ADX (14)
Volatility	Bollinger Bands (20, 2), Keltner Channels (20, 2), ATR (14)
Volume	MFI (14), VWAP (14), OBV, CMF, Pocket Pivot (10, 1.3)

- **Data Preprocessing & Flattening:** To stabilize training, the input matrix undergoes batch normalization. It is then flattened into a 345-dimensional feature vector, serving as the primary representation for subsequent layers.
- **Feature Projection:** A dense projection layer of 128 neurons with linear activation is employed to map the raw inputs into a latent feature space.
- **Non-linear Feature Extraction:** The core of the architecture consists of two hidden blocks:
  1. A dense layer with a size of 256 neurons.
  2. A dense layer with a size of 128 neurons.

  They both apply *tanh* as the activation function and incorporate a Dropout layer (rate = 0.05) to prevent overfitting and improve generalization.
- **Regression Output:** The final stage is a single-neuron output layer with linear activation.
- **Loss Function:** We use the Huber loss function.

4 Experiments

The experimental evaluation of Kagi chart-based strategies is conducted in a two-stage process. Initially, we utilize a custom-built simulation environment designed to facilitate rapid backtesting and hyperparameter optimization. A key feature of this simulation is its support for fractional share trading, allowing for precise capital allocation and realistic simulation. The strategy operates on a single asset at a time and follows an all-in allocation policy. Subsequently, the most promising models and prediction configurations are validated using *QuantConnect*[10], an industry-standard algorithmic trading platform, to ensure the robustness and reproducibility of the results in a production-grade environment.

4.1 Experimental Setup and Dataset

To ensure the generalizability of the results, the model is trained on historical data spanning from January 1, 2016, to December 31, 2022, and validated using the period from January 1, 2023, to December 31, 2023. The final performance was evaluated on a testing dataset covering September 1, 2024, to December 31, 2025 and evaluated on the following nine equities: *A, AAPL, MSFT, GOOGL,*

AMZN, TSLA, META, NVDA, MU. These nine equities represent a sample of current technology leaders.

The parameters introduced below apply exclusively to the trading logic used after the model produces its predictions. These parameters are varied only during the evaluation phase and do not affect the model itself or its learned representations. This separation ensures that the results are not influenced by overfitting or selection bias. The strategy performance was evaluated using a subset of Cartesian product of the following hyperparameter configurations:

- **Order Execution Levels ($[L, U]$):** The trading strategy executes orders based on the predicted price movement magnitude. When the model's predicted change exceeds the lower bound L, a buy or sell order is triggered in the direction of the prediction; movements below L are treated as noise and ignored. The upper bound U caps the maximum expected move, normalizing the signal strength within the interval $[L, U]$. Three configurations reflecting different entry sensitivities are evaluated: $[0.0, 10.0]$ (no minimum threshold), $[0.01, 10.0]$ (minimal noise filtering), and $[0.2, 10.0]$ (conservative, requiring stronger signals to execute).
- **Risk Management:** Four stop-loss thresholds (2%, 5%, 7%, and 10%) to assess capital protection.
- **Reversal Methodologies:**
  - **Fixed Percentage:** Ten thresholds ranging from 0.004 to 0.036: $(0.004, 0.008, 0.012, 0.014, 0.016, 0.02, 0.024, 0.028, 0.032, 0.036)$.
  - **Dynamic Volatility-Adjusted:** Using Average True Range (ATR) ranging from 0.6 to 3.8: $(0.6, 0.8, 1.2, 1.4, 1.6, 1.8, 2.2, 2.8, 3.2, 3.8)$.

For each equity, a total of 36 simulation configurations are evaluated, yielding 324 simulations in aggregate across all parameters. However, as the reversal threshold increases, some configurations are excluded due to insufficient data to support a statistically meaningful evaluation.

4.2 Results and Discussion: Daily Resolution

Initial experiments are conducted using daily (1-day) data to establish a baseline performance overview. Tables 2, 3 summarize the performance metrics of our simulation across various configurations. To ensure a clear interpretation of the results, the parameters and columns are defined as follows:

- **Reversal Threshold**: The threshold value utilized to identify trend shifts.
- **Sample Size ($d = 23$)**: Represents the size of the training matrix, defined as the base sample size multiplied by the factor $d = 23$.
- **Total Simulations**: The aggregate number of simulation runs performed per configuration.
- **Profit Ratio (%)**: The proportion of simulation outcomes yielding a positive result, expressed as a percentage.
- **Max Gain/Min Gain (Ratio)**: The maximum and minimum recorded returns, normalized as a ratio of the starting value.

Table 2. Daily simulation results for fixed percentage-based reversal

Reversal Threshold	Sample Size	Total Simulations	Profit Ratio (%)	Max Gain (Ratio)	Min Gain (Ratio)
0.004	5,565	324	78.1%	2.1904	0.6541
0.008	4,953	324	70.7%	3.1353	0.6163
0.012	4,102	324	64.8%	2.4673	0.5665
0.014	3,768	324	55.9%	2.6187	0.6389
0.016	3,492	320	69.7%	2.4031	0.7272
0.020	2,954	300	53.7%	2.3240	0.7065
0.024	2,509	296	61.8%	1.6320	0.6503

As shown in Table 2, lower fixed percentage thresholds (e.g., 0.004) yield the highest ratio of profitable simulations, but the number of available signals (sample size) decreases as the threshold increases. This trend is also evident in the dynamic ATR-based approach. The ATR-based results indicate that dynamic adjustment produces a more adaptive reversal mechanism. However, a critical observation during the daily data testing was the relationship between the reversal threshold (both fixed and dynamic) and the statistical significance of the results.

Table 3. Daily simulation results for ATR-based reversal.

Reversal Threshold	Sample Size	Total Simulations	Profit Ratio (%)	Max Gain (Ratio)	Min Gain (Ratio)
0.6 x ATR	3,455	324	71.6%	2.4792	0.8639
0.8 x ATR	2,774	292	45.6%	2.0219	0.7153

4.3 Transition to Intraday Resolution

A major limitation identified during the daily data experiments is the scarcity of data points at higher reversal levels. As the reversal threshold or ATR multiplier increases, the frequency of Kagi "shoulders" and "waists" (reversal points) significantly decreases. For daily data, testing values beyond the current range proved impractical, as the infrequent signals resulted in statistically insignificant performance metrics and failed to capture the nuances of the Kagi chart dynamics. To address this data sparsity and to investigate the strategy's behavior in a more high-frequency context, the next phase of our experiment shifts to 30-min intraday data. This higher resolution provides a significantly larger sample size, allowing for a more robust evaluation of the predictive power of Kagi charts across a wider spectrum of hyperparameters.

4.4 Results and Discussion: 30-Min Resolution

The application of the Kagi strategy to 30-min resolution data yielded a substantially larger dataset, which allowed us to test significantly higher reversal thresholds that are not feasible with daily bars.

Fixed Percentage-Based Analysis - As demonstrated in Table 4a, the shift to intraday data reveals a non-linear relationship between reversal threshold and profit ratio. While low thresholds (e.g., 0.004) produce a largest dataset of over 38,000 observations, the highest profitability ratio (77.08%) is actually achieved at a much higher threshold of 0.036. This indicates that in an intraday environment, the Kagi chart acts as a potent noise filter. Higher reversal thresholds suppress high-frequency noise inherent in 30-min bars and allow the model to focus on structurally significant price movements.

Dynamic ATR-Based Analysis - The dynamic volatility-adjusted reversal results (Table 4b) show an even more pronounced improvement. By scaling the reversal threshold to the current Average True Range, the strategy becomes highly adaptive to shifts in intraday volatility. The performance peaks at the 3.8×ATR multiplier, yielding a profitability ratio of 91.11%. This suggests that the strategy is most robust when the Kagi reversal is set significantly wider than the standard noise levels (ATR), confirming that the predictive power of Kagi patterns is most reliable for identifying major trend shifts rather than minor fluctuations.

Table 4. Overall simulation performance for ATR-based and percentage-based reversal thresholds. The best results are marked with an asterisk (*).

(a) Percentage-based reversal

Reversal Threshold	Sample Size	Profit Ratio	Total Sim.
0.004	38118	58.68	324
0.008	20206	59.72	324
0.012	12912	57.99	324
0.014	10857	63.54	324
0.016	9245	55.08	288
0.020	6973	55.42	252
0.024	5465	42.79	216
0.028	4445	62.78	180
0.032	3655	73.61	144
0.036	3090	**77.08***	144

(b) ATR-based reversal

Reversal Threshold	Sample Size	Profit Ratio	Total Sim.
0.6×ATR	40494	54.17	324
0.8×ATR	30970	58.33	324
1.2×ATR	20254	54.17	324
1.4×ATR	16883	63.54	324
1.6×ATR	14435	53.12	324
1.8×ATR	12551	55.56	324
2.2×ATR	9851	60.07	324
2.8×ATR	7161	71.55	288
3.2×ATR	5928	63.94	252
3.8×ATR	4664	**91.11***	180

4.5 Optimal Configurations and Stock Performance

Results across the tested instruments suggest that the ATR-based approach is more robust than the fixed percentage method for 30-min data. While the best percentage-based threshold (0.036%) yielded a respectable 77.08% profit ratio,

Table 5. Stock-level performance for the best percentage-based (0.036) and ATR-based (3.8×ATR) thresholds.

Stock	Percentage (0.036%)				ATR (3.8×ATR)			
	Sims	Max gain	Min gain	Avg gain	Sims	Max gain	Min gain	Avg gain
A	36	1.3736	0.8834	1.1171	36	1.2358	0.8977	1.1023
GOOGL	36	1.5651	1.0372	1.2696	36	1.7509	0.8398	1.3616
META	36	1.3810	0.8373	1.0887	36	1.5190	0.8627	1.0855
MSFT	-	-	-	-	36	1.4298	1.0663	1.2432
TSLA	36	1.8768	0.7609	1.1558	36	2.6756	0.6966	1.4875

it was consistently outperformed by the $3.8 \times$ ATR configuration, which reached a 91.11% success rate.

To ensure these results were not skewed by specific ticker behavior, we analyzed individual equity performance for these optimal settings. As shown in Table 5, the strategy remains consistently profitable across different stocks. Notably, high-beta stocks such as TSLA show significant "Max Gain" potential, particularly in the ATR-based configuration (2.6756), while maintaining an average gain well above 1.0.

4.6 QuantConnect Validation

The results obtained from the custom simulation environment indicate that high-threshold Kagi reversal strategies—specifically the 0.036 fixed reversal amount and the 3.8×ATR dynamic threshold—exhibit strong statistical performance. However, these experiments are conducted under idealized assumptions, including frictionless execution and the ability to trade fractional shares.

To assess how these strategies would behave under realistic trading conditions, both configurations are implemented in the QuantConnect platform using the LEAN engine. QuantConnect enforces market constraints such as discrete order sizing (whole-share trading only), bid–ask spreads, transaction fees, and execution latency, thereby providing a close approximation of real-world algorithmic trading.

By deploying the 0.036 reversal and the 3.8×ATR strategies in this environment, we aim to observe whether their favorable properties persist when subjected to practical limitations. The following tables report the resulting performance metrics and allow a direct comparison between the idealized custom backtests and the more realistic QuantConnect executions.

The columns reported in Tables 6, 7 summarize the performance of each trading strategy as follows. **Stock** denotes the traded financial instrument. **QC (Whole)** represents the final portfolio gain obtained from the QuantConnect backtesting environment, where trades are executed using whole shares only. **Custom (Frac.)** reports the final portfolio gain produced by the proposed custom simulator, which allows trading with fractional shares. **Fees ($)** indicate

the total transaction costs paid over the entire simulation period according to the Alpaca fee structure. **Win %** denotes the proportion of profitable trades relative to the total number of executed trades. Finally, **Avg. Win** represents the average return achieved on winning trades.

Table 6. Performance metrics for 3.6% reversal strategy: custom vs. QuantConnect

	End Gain				
Stock	QC (Whole)	Custom (Frac.)	Fees ($)	Win %	Avg. Win
TSLA	2.01	1.8768	27.05	61%	3.84%
META	1.32	1.3810	12.77	58%	2.43%
GOOGL	1.47	1.5651	13.61	49%	3.19%
A	1.13	1.3736	18.71	59%	1.27%

Table 7. Performance metrics for 3.8×ATR reversal strategy: custom vs. QuantConnect

	End Gain				
Stock	QC (Whole)	Custom (Frac.)	Fees ($)	Win %	Avg. Win
A	0.86	1.2358	14.43	47%	2.14%
GOOGL	1.55	1.7509	18.72	61%	2.55%
META	1.13	1.5190	19.60	51%	2.09%
MSFT	1.29	1.4298	22.25	52%	2.14%
TSLA	1.68	2.6756	13.45	70%	5.38%

Tables 6 and 7 present a side-by-side comparison of the two best-performing Kagi reversal configurations evaluated in this study: the fixed 0.036 reversal amount and the dynamic 3.8×ATR threshold. For each instrument, the final performance obtained in the custom simulator (allowing fractional shares) is contrasted with the corresponding outcome achieved in the QuantConnect environment, which enforces whole-share trading and realistic market frictions.

For the 0.036 reversal strategy (Table 6), the QuantConnect results remain broadly consistent with the custom backtests. While small deviations in end gain are observed across assets, the strategy preserves its positive edge in all cases. Importantly, the win rates and average profit per winning trade remain stable, indicating that the underlying signal quality is robust to real-world trading conditions.

The results of the 3.8×ATR strategy (Table 7) arguably suggest a more efficient generalization of the underlying market patterns. In several cases—most

notably TSLA and GOOGL—the QuantConnect implementation achieves competitive or superior performance relative to the idealized custom simulator. This suggests that the volatility-adjusted reversal threshold adapts well to realistic market microstructure effects. The high win rates, particularly the 70% observed for TSLA, further confirm that the strategy maintains strong directional accuracy under production-grade execution constraints.

5 Conclusion

In this study, we investigated the efficacy of Kagi chart representations as structured inputs for deep learning models in algorithmic trading. By transforming noisy, time-based financial series into trend-dependent Kagi segments, we demonstrated a robust method for capturing significant market movements while suppressing stochastic fluctuations.

Our empirical results highlight two critical factors for the success of such models. First, the transition from fixed percentage-based reversals to volatility-adaptive thresholds using the Average True Range (ATR) significantly enhanced the model's ability to adapt to varying market conditions. Second, our comparative analysis across different timeframes revealed that daily data often lacks the density required for training complex neural architectures like the MLP used in our framework.

The move to 30-min interval data proved essential, providing the necessary volume of Kagi structures to achieve stable training and consistent profitability across all simulations. The primary contribution of this work lies in the validation of Kagi segment duplets as a compact, information-dense encoding for neural networks. This representation not only simplifies the learning task by focusing on validated trend states but also demonstrates remarkable resilience in backtesting environments. Future research will aim to conduct a more extensive evaluation of the proposed strategies by applying them to a broader set of financial instruments, such as S&P 500 constituents, across multiple timeframes. Additionally, we intend to investigate the robustness of these strategies under varying market conditions and explore potential enhancements through alternative parameterizations and risk management techniques.

Acknowledgments. This research was supported by the Slovak Research and Development Agency (APVV) under Grant No. APVV-24-0446.

References

1. Alpaca Securities LLC: Alpaca API Documentation (2026). https://alpaca.markets/
2. Arévalo, R., García, J., Guijarro, F., Peris, R.I.: A dynamic trading strategy based on a genetic algorithm. Appl. Soft Comput. **58**, 336–344 (2017)

3. Cavalcante, R.C., Brasileiro, R.C., Souza, V.L., Nobrega, J.P., Evangelista, A.L.: Computational intelligence and financial markets: a survey and future directions. Expert Syst. Appl. **55**, 194–211 (2016)
4. Cohen, D.: Trading with Price Action: Strategies for Renko and Kagi Charts. Financial Times Press (2020)
5. Dorsey, T.J.: Point and Figure Charting: The Essential Application for Forecasting and Strategic Market Timing, 4th edn. Wiley, Hoboken (2013)
6. Drozda, M., Cavojsky, M., Sebes, P.: On suitability of renko charts for algorithmic trading. In: International Scientific Conference on Distance Learning in Applied Informatics, pp. 69–81. Springer, Cham (2024)
7. Huang, W., Nakamori, Y., Wang, S.Y.: Forecasting stock market movement direction with support vector machines. Comput. Oper. Res. **32**(10), 2513–2522 (2005)
8. Lau, A., et al.: Adaptive trading strategies using volatility-based indicators. J. Comput. Financ. **12**(2), 55–82 (2008)
9. Nison, S.: Beyond Candles: New Japanese Charting Techniques Revealed. Wiley, Hoboken (1994)
10. QuantConnect Corporation: LEAN Algorithmic Trading Engine (2026). https://www.quantconnect.com/lean/. Open-source algorithmic trading engine
11. Sezer, O.B., Ozbayoglu, M.: Algorithmic financial trading with deep convolutional neural networks: time series to image conversion approach. Appl. Soft Comput. **70**, 525–538 (2018)
12. Tsay, R.S.: Analysis of Financial Time Series. Wiley, Hoboken (2005)
13. Wilder, J.W.: New Concepts in Technical Trading Systems. Trend Research (1978)
14. Zhang, Z., Zohren, S., Roberts, S.: Automated stock trading via equilibrium-based reinforcement learning. arXiv preprint arXiv:1907.03665 (2019)

The Impact of BinaryTarget Engineering on Sector-Wise Class Differentiation in Next-Day Return Prediction

Lívia Kelebercová(✉), Michal Munk, and Jakub Marták

Faculty of Natural Sciences and Informatics, Constantine the Philosopher University, Nitra, Slovakia

{lkelebercova,mmunk,jakub.martak}@ukf.sk

Abstract. Financial markets are complex, non-linear systems where short-term price prediction remains a fundamental challenge. While machine learning is frequently applied to directional classification, the structural impact of target construction specifically the choice of return threshold is often overlooked in favor of model architecture. This study investigates how different binary target formulations influence the statistical separability of baseline OHLCV (Open, High, Low, Close, Volume) predictors across various GICS sectors. Using a multi-asset equity dataset, we defined six binary targets ranging from simple direction-based signs to progressively stricter return thresholds. We developed a sector-wise meta-analytic framework employing hypothesis testing and introduced two novel metrics: Relative Differentiation (*RD*) and Cumulative Differentiation (*CD*) to quantify class-conditional distribution shifts. Empirical findings reveal that simple sign-based targets yield minimal statistical differentiation for price-level variables across most sectors. However, as return thresholds increase, predictor separability intensifies significantly. Notably, trading volume emerged as the most robust differentiator, achieving its highest *RD* values under the strictest movement criteria. Target labeling is a critical structural design choice. Stricter thresholds filter market noise and enhance the signal-to-noise ratio of baseline predictors, suggesting that sectoral heterogeneity must be integrated into financial labeling strategies to improve model stability.

Keywords: Financial Machine Learning · Target Engineering · Sector-Wise Analysis · Next-Day Return Prediction · Statistical Class Separability · GICS Classification

1 Introduction

Financial markets represent complex, adaptive, and highly non-linear systems characterized by stochastic dynamics, structural breaks, and heterogeneous behavioral interactions. Despite decades of research in econometrics and quantitative finance, predicting short-term asset price movements remains a challenging task. The difficulty arises not only from market efficiency considerations but also from noise dominance, regime shifts, volatility clustering, and cross-sector structural differences.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 624–634, 2027.
https://doi.org/10.1007/978-3-032-32104-6_52

In empirical asset pricing and financial machine learning, prediction problems are commonly formulated either as regression tasks (forecasting future returns or prices) or as classification tasks (predicting the direction of future movement). Directional prediction is often preferred in practice because many trading strategies depend primarily on correctly identifying the sign of future returns rather than their precise magnitude.

A fundamental but frequently underemphasized aspect of classification-based financial modeling is target construction. Transforming continuous returns into discrete labels is not merely a preprocessing step but a modeling decision that shapes class balance, signal-to-noise ratio, and statistical separability of predictors. Simple direction-based labeling (positive vs. negative return) may capture minimal price movements that are economically insignificant, whereas threshold-based definitions introduce stricter criteria for defining meaningful movements. As the return threshold increases, the resulting classification task becomes more selective, potentially enhancing class differentiation but also altering sample composition.

Another key characteristic of financial markets is sectoral heterogeneity. Firms belonging to different economic sectors are exposed to distinct macroeconomic drivers, demand structures, regulatory environments, and liquidity conditions. Consequently, the statistical relationship between baseline market variables such as open, high, low, close prices and trading volume and future returns may vary systematically across sectors. Ignoring this heterogeneity may obscure sector-specific structural patterns.

This study investigates how alternative binary target formulations influence the statistical separability of baseline OHLCV predictors across GICS sectors. Instead of focusing on predictive model performance, we examine whether the conditional distributions of predictors differ significantly between target classes and how this differentiation changes as return thresholds become progressively stricter. By introducing measures of Relative Differentiation (RD) and Cumulative Differentiation (CD), we provide a structured sector-wise meta-analytic perspective on the interaction between target construction and predictor behavior.

2 Related Works

2.1 Direction-Based Labeling Approaches

Binary classification of next-period return direction has become a standard framework in financial machine learning. Early influential work by Kim applied support vector machines to stock direction prediction, demonstrating improved performance over traditional statistical models [1]. Huang et al. further compared SVM with backpropagation neural networks for forecasting stock index direction [2].

Ballings et al. evaluated multiple classifiers, including Random Forest and logistic regression, for stock price direction prediction and showed that ensemble methods often outperform single classifiers [3]. Patel et al. proposed a fusion-based machine learning approach for stock index prediction using technical indicators [4]. Krauss et al. extended this line of research by applying deep neural networks and gradient-boosted trees in a statistical arbitrage setting, converting daily probability forecasts into long–short trading strategies [5].

Sezer et al. provided a comprehensive review of deep learning methods for financial time series forecasting, highlighting the dominance of recurrent neural networks in directional prediction tasks [6].

Although these studies confirm the viability of direction-based classification, they generally assume a fixed binary labeling scheme and do not systematically analyze how alternative threshold definitions affect statistical properties of predictors.

2.2 Trend-Based and Alternative Labeling Approaches

Recent literature emphasizes that label definition significantly influences predictive performance and economic interpretation.

Zhao et al. proposed a formally redefined trend labeling framework inspired by Dow Theory and retracement principles, integrating it into a time-weighted LSTM model [7]. Their results demonstrate that trend-aware labeling improves classification accuracy.

Wu et al. introduced a "continuous trend labeling" method designed to capture sustained market movements rather than short-term noise [8]. They argue that traditional next-day comparison labeling fails to reflect underlying trend continuity.

Kovačević et al. developed an optimal trend labeling framework and introduced the concept of labeling robustness, showing that different labeling algorithms may lead to distinct financial outcomes despite similar accuracy levels [9].

López de Prado provided a systematic treatment of financial labeling in *Advances in Financial Machine Learning* [10]. He introduced the triple-barrier method, which defines labels based on profit-taking, stop-loss, and time horizons, and proposed meta-labeling to separate signal direction from trade selection. This work formalized labeling as a structural design choice in financial modeling.

Tsinaslanidis and Kugiumtzis proposed a structural prediction scheme based on perceptually important points (PIPs) and dynamic time warping (DTW), focusing on subsequence similarity rather than direct binary labeling [11].

2.3 Return Predictability and Sectoral Heterogeneity

Return predictability has been studied extensively from both cross-sectional and sectoral perspectives. Rapach et al. investigated international stock return predictability and the role of industry structure [12]. Bhojraj and Swaminathan examined macromomentum effects in industry portfolios [13], while Chordia and Swaminathan analyzed cross-autocorrelations in stock returns [14].

Gervais et al. studied the high-volume return premium, demonstrating that trading activity is associated with future returns [15]. These studies highlight that return dynamics and predictor relevance may vary across industries and liquidity regimes.

Despite this extensive literature on return predictability and industry effects, relatively little research has examined how alternative target constructions interact with sectoral heterogeneity in shaping predictor distributional separability. Most prior work evaluates forecasting accuracy rather than analyzing the statistical structure induced by different labeling schemes.

3 Materials and Methods

3.1 Dataset and Market Variables

The empirical analysis was conducted on a multi-asset equity dataset containing daily OHLCV information across a broad set of financial instruments. Each observation corresponds to one daily bar for a given ticker, including the standard price-level variables open, high, low, and close, as well as trading volume measured by volume_base. Each instrument was additionally assigned to a Global Industry Classification Standard (GICS) sector, enabling sector-level stratification of statistical effects.

The dataset further contains engineered technical indicators derived from OHLCV data; however, the present study focuses specifically on the baseline OHLCV predictors in order to quantify their sector-dependent statistical relevance under alternative target formulations.

3.2 Target Variables

Binary target variables were defined to represent next-day price movement under both direction-based and threshold-based formulations. Let $close_t$ denote the closing price at trading day t, and $close_{t+1}$ the closing price at the subsequent trading day. The one-day return was computed as

$$r_{t+1} = \frac{close_{t+1}}{close_t} - 1.$$

Based on this return definition, six binary classification targets were constructed:

- **BIN_T1**: indicator of whether $close_{t+1} > close_t$
- **BIN_T2**: indicator of whether $r_{t+1} > 0$
- **BIN_T3_05**: indicator of whether $r_{t+1} > 0.0005$
- **BIN_T3_10**: indicator of whether $r_{t+1} > 0.0010$
- **BIN_T3_20**: indicator of whether $r_{t+1} > 0.0020$
- **BIN_T3_50**: indicator of whether $r_{t+1} > 0.0050$

These targets represent progressively stricter definitions of positive next-day movement, allowing the evaluation of how target formulation affects the statistical separability of predictors.

3.3 Sector-Wise Meta-Analytic Framework

To account for heterogeneity across market segments, all statistical analyses were performed separately within each GICS sector. For each sector S, each target definition Y, and each predictor variable $X \in \{open, high, low, close, volume_base\}$, a hypothesis test was conducted to evaluate whether the predictor differs between the two target classes.

Formally, for each sector-target-predictor combination, the null hypothesis was defined as

$$H_0 : E[X_s \mid (Y_s = 0)] = E[X_s \mid (Y_s = 1)],$$

where $E[X_s \mid (Y_s = k)]$ denotes the expected value of predictor X given class k. The alternative hypothesis assumes a statistically significant difference between the class-conditional expected values.

The resulting p-values were used to assess statistical significance at conventional thresholds ($p < 0.05$, $p < 0.01$, $p < 0.001$). This procedure yielded a sector-wise map of statistically significant differentiation across predictors and target definitions.

In addition to statistical significance, the magnitude of class separation was quantified using Relative Differentiation (*RD*). *RD* provides a standardized measure of the strength of differentiation between the two target classes for a given predictor within a sector.

To summarize predictor behavior across sectors, Cumulative Differentiation (*CD*) was computed as an aggregated measure reflecting the accumulation of sectoral differentiation contributions for each predictor–target combination. This enabled comparison of cross-sector robustness across alternative target formulations.

4 Results

A sector-wise meta-analysis was conducted to evaluate the statistical differentiation of OHLCV predictors (open, high, low, close, and volume_base) across binary return targets. The targets included two direction-based formulations (BIN_T1 and BIN_T2) and four threshold-based formulations (BIN_T3_05, BIN_T3_10, BIN_T3_20, BIN_T3_50). For each predictor–target combination, hypothesis tests were performed within GICS sectors to determine whether predictor distributions differed significantly between target classes. Statistical significance was assessed using p-values, and effect magnitude was quantified using Relative Differentiation (*RD*). For each target, the sum of *RD* across all sector groupings equals 1.

4.1 Direction-Based Targets

For the price-level predictors (open, high, low, close), an identical sectoral structure was observed under both BIN_T1 and BIN_T2.

Across all four price predictors, statistically significant differentiation was observed exclusively in the Energy sector at the 1% significance level ($p < 0.01$), with $RD = 0.08$. All remaining sectors (Information Technology, Utilities, Financials, Health Care, Industrials, Real Estate, Communication Services, ETF/Index, Materials, Consumer Discretionary, Consumer Staples, and Unknown Sector) were jointly non-significant ($p > 0.05$) and collectively accounted for $RD = 0.92$. This pattern was consistent across open, high, low, and close, and identical for BIN_T1 and BIN_T2.

For the trading activity predictor volume_base, the sectoral structure differed from the price-level variables. Under both BIN_T1 and BIN_T2, the Industrials sector reached the highest level of statistical significance ($p < 0.001$) with $RD = 0.08$. A second statistically significant grouping comprised Financials, Health Care, Real Estate, and ETF/Index, which reached $p < 0.01$ and accounted for $RD = 0.31$ ($CD = 0.38$). The remaining sectors (Utilities, Materials, Consumer Discretionary, Communication Services, Consumer Staples, Information Technology, Energy, and Unknown Sector) were non-significant ($p > 0.05$) and collectively contributed $RD = 0.62$.

Thus, under direction-based targets, price-level differentiation was concentrated exclusively in Energy, whereas volume-based differentiation exhibited a two-tier significant structure.

4.2 Threshold-Based Targets

Under threshold-based targets, sectoral differentiation became more distributed across sectors for all predictors.

BIN_T3_05
For the open predictor, the Unknown Sector reached $p < 0.001$ ($RD = 0.08$), and Information Technology reached $p < 0.01$ ($RD = 0.08$). The sectors Utilities, Real Estate, and Energy reached $p < 0.05$ and jointly accounted for $RD = 0.23$ ($CD = 0.38$). The remaining sectors (Health Care, Industrials, Consumer Discretionary, Materials, Financials, Communication Services, Consumer Staples, and ETF/Index) were non-significant and accounted for $RD = 0.62$.

For high and low, a similar structure was observed: Unknown Sector and Information Technology reached $p < 0.01$ ($RD = 0.15$), while Utilities, Real Estate, and Energy reached $p < 0.05$ ($RD = 0.23$). The remaining sectors accounted for $RD = 0.62$ and were non-significant.

For close, Information Technology and Unknown Sector reached $p < 0.01$ ($RD = 0.15$). The sectors Utilities, Real Estate, Health Care, and Energy reached $p < 0.05$ and jointly accounted for $RD = 0.31$ ($CD = 0.46$). The remaining sectors accounted for $RD = 0.54$.

For volume_base, the sectors Unknown Sector, ETF/Index, and Industrials reached $p < 0.001$ and jointly accounted for $RD = 0.23$. Health Care reached $p < 0.01$ ($RD = 0.08$). The remaining sectors were non-significant and accounted for $RD = 0.69$.

BIN_T3_10
For open, high, low, and close, an identical structure was observed. The Unknown Sector reached $p < 0.001$ ($RD = 0.08$). Information Technology reached $p < 0.01$ ($RD = 0.08$). Utilities reached $p < 0.05$ ($RD = 0.08$). All remaining sectors collectively accounted for $RD = 0.77$ and were non-significant.

For volume_base, ETF/Index and Unknown Sector reached $p < 0.001$ ($RD = 0.15$). The sectors Health Care, Information Technology, and Industrials reached $p < 0.05$ and jointly accounted for $RD = 0.23$. The remaining sectors contributed $RD = 0.62$ and were non-significant.

BIN_T3_20
For open and high, the sectors Unknown Sector and Materials reached $p < 0.001$ ($RD = 0.15$). Information Technology reached $p < 0.01$ ($RD = 0.08$). The sectors Real Estate, Financials, Utilities, and ETF/Index reached $p < 0.05$ and jointly accounted for $RD = 0.31$ ($CD = 0.54$). The remaining sectors (Industrials, Energy, Consumer Discretionary, Health Care, Communication Services, and Consumer Staples) were non-significant ($RD = 0.46$).

For low and close, Unknown Sector and Materials reached $p < 0.001$ ($RD = 0.15$), and Information Technology reached $p < 0.01$ ($RD = 0.08$). The sectors Real Estate, Financials, and Utilities reached $p < 0.05$ ($RD = 0.23$; $CD = 0.46$). The remaining sectors accounted for $RD = 0.54$.

For volume_base, ETF/Index, Unknown Sector, and Information Technology reached $p < 0.001$ and jointly accounted for $RD = 0.23$. The sectors Materials, Utilities, and Consumer Discretionary reached $p < 0.05$ and jointly accounted for $RD = 0.23$ ($CD = 0.46$). The remaining sectors contributed $RD = 0.54$.

BIN_T3_50

For open and high, the sectors ETF/Index, Materials, Unknown Sector, and Financials reached $p < 0.001$ and jointly accounted for $RD = 0.31$. Information Technology reached $p < 0.01$ ($RD = 0.08$). The sectors Utilities and Health Care reached $p < 0.05$ ($RD = 0.15$; $CD = 0.54$). The remaining sectors accounted for $RD = 0.46$.

For low, the sectors Materials, Unknown Sector, ETF/Index, and Financials reached $p < 0.001$ ($RD = 0.31$). Information Technology and Utilities reached $p < 0.01$ ($RD = 0.15$; $CD = 0.46$). Health Care reached $p < 0.05$ ($RD = 0.08$; $CD = 0.54$). The remaining sectors contributed $RD = 0.46$.

For close, the sectors ETF/Index, Materials, Unknown Sector, and Financials reached $p < 0.001$ ($RD = 0.31$). Information Technology reached $p < 0.01$ ($RD = 0.08$). Utilities and Health Care reached $p < 0.05$ ($RD = 0.15$; $CD = 0.54$). The remaining sectors accounted for $RD = 0.46$.

For volume_base under BIN_T3_50, a markedly different structure was observed. A large sector grouping comprising Consumer Discretionary, ETF/Index, Financials, Information Technology, Materials, Real Estate, Unknown Sector, Utilities, Energy, and Communication Services reached $p < 0.001$ and accounted for $RD = 0.77$. Consumer Staples reached $p < 0.05$ ($RD = 0.08$; $CD = 0.85$). The remaining sectors, Health Care and Industrials, were non-significant and jointly accounted for $RD = 0.15$.

5 Discussion

The empirical findings reveal several structural regularities in how target construction influences sector-wise distributional separability of baseline OHLCV predictors. The results indicate that the statistical differentiation of predictors is not only sector-dependent but also strongly sensitive to the definition of the return target.

5.1 Direction-Based vs. Threshold-Based Targets

For the direction-based targets (BIN_T1 and BIN_T2), statistically significant differentiation of price-level predictors (open, high, low, close) was observed exclusively in the Energy sector. In all other sectors, the distribution of predictor values between classes did not differ significantly. Moreover, the Relative Differentiation (RD) associated with the significant Energy effect was modest ($RD = 0.08$), while the non-significant aggregated sector group accounted for $RD = 0.92$. This pattern was identical across both direction-based formulations.

Simple sign-based next-day return labeling shows weak cross-sector separation for OHLC variables, supporting claims that it mainly captures short-term noise rather than meaningful trends (Wu et al. [8]; López de Prado [10]). Consistent with this, sign-based targets failed to differentiate 12 of 13 sectors, suggesting daily price changes are dominated by noise rather than signal. Increasing return thresholds improves sectoral differentiation, indicating that stricter labeling filters out noise and better reveals underlying structural relationships (Wu et al. [8]).

In contrast, threshold-based targets (BIN_T3_05 to BIN_T3_50) produced substantially broader sectoral differentiation. As return thresholds increased, statistically significant sectors expanded across multiple industries. Under BIN_T3_20 and BIN_T3_50, several sectors reached $p < 0.001$ for price-level predictors, including combinations of Unknown Sector, Materials, Financials, and ETF/Index, depending on the predictor. *RD* values under higher thresholds reached 0.31 at the sector-group level for price variables, indicating stronger concentration of class separation within statistically significant sector clusters.

This structural shift supports the theoretical argument that filtering out minor price fluctuations enhances the statistical distinction between return classes. Zhao et al. [7] demonstrate that redefining trend structure improves predictive modeling performance. The present findings complement that perspective by showing that stricter threshold definitions enhance statistical separability even prior to model estimation. Thus, improved predictive performance reported in trend-based studies may partly reflect improved intrinsic class structure.

5.2 The Role of Volume

The behavior of volume_base differs structurally from that of price-level predictors. Even under direction-based targets, volume exhibited statistically significant differentiation in more sectors than price variables. Specifically, Industrials reached $p < 0.001$, and a second grouping (Financials, Health Care, Real Estate, ETF/Index) reached $p < 0.01$. This two-tier significant structure contrasts with the single-sector significance observed for price variables.

Under threshold-based targets, the differentiating power of volume increased further. For BIN_T3_05 and BIN_T3_10, significant differentiation appeared across multiple sectors, including ETF/Index, Unknown Sector, and Industrials. Under BIN_T3_50, a large aggregated sector group reached $p < 0.001$ and accounted for $RD = 0.77$, representing the strongest observed class separation across all predictors and targets. This indicates a high concentration of distributional separation in volume under extreme return thresholds.

These results align with the literature linking trading activity to return dynamics. Gervais et al. [15] document that unusually high trading volume predicts subsequent return performance, while Chordia and Swaminathan [14] show that volume contributes to return cross-autocorrelations. The present findings extend these insights by demonstrating that the sector-wise discriminative strength of volume intensifies as return thresholds become stricter. Higher thresholds likely isolate events associated with stronger market participation, thereby amplifying volume-related differentiation.

5.3 Sectoral Heterogeneity

The results reveal pronounced sectoral heterogeneity. Under direction-based targets, only Energy displayed statistically significant differentiation for price-level predictors. However, under threshold-based targets, sectors such as Information Technology, Materials, Financials, Utilities, ETF/Index, and Unknown Sector emerged as statistically significant depending on the threshold level.

This sectoral variation suggests that the statistical relationship between baseline OHLCV predictors and future returns is not uniform across industries. Rapach et al. [12] demonstrate that industry structure plays a role in return predictability across markets, while Bhojraj and Swaminathan [13] show that momentum effects differ across industry portfolios. The present findings provide complementary distributional evidence: sector-specific structures influence the degree to which predictors differentiate return classes under alternative labeling schemes.

Importantly, these patterns indicate that a uniform target construction strategy applied across all sectors may obscure structural differences in predictor behavior.

5.4 Labeling as Structural Design

Recent research emphasizes the importance of labeling robustness and economic interpretability [6, 10]. However, much of the literature evaluates labeling schemes primarily through predictive accuracy or trading performance. The present study approaches the problem from a structural perspective by analyzing distributional separability prior to model estimation.

As return thresholds increase, Cumulative Differentiation (*CD*) patterns indicate stronger concentration of statistically significant sector groups. Higher thresholds redistribute *RD* toward sectors exhibiting stronger class separation, thereby reshaping the statistical landscape of the dataset. This suggests that labeling is not merely a technical preprocessing step but a structural design choice that alters feature relevance and cross-sector differentiation patterns.

Consequently, labeling decisions may influence downstream model stability, feature importance rankings, and generalization capacity, even before algorithmic modeling choices are made.

6 Conclusion

This study examined how alternative binary return target formulations affect sector-wise statistical separability of baseline OHLCV predictors. By comparing direction-based and progressively stricter threshold-based targets across GICS sectors, we demonstrate that labeling design materially alters the distributional structure of explanatory variables.

Direction-based targets produced limited differentiation for price-level variables, with statistical significance confined to a single sector. In contrast, threshold-based targets generated broader and stronger sectoral separation, particularly under higher thresholds. Trading volume exhibited consistently stronger and more widespread differentiation compared to price-level predictors, especially under extreme threshold definitions.

Overall, the findings indicate that target construction plays a fundamental role in shaping the statistical environment in which predictive models operate.

7 Contributions

The main contributions of this study are threefold. First, we demonstrate empirically that return threshold selection directly influences predictor distributional separability. Second, we introduce Relative Differentiation (*RD*) and Cumulative Differentiation (*CD*) as sector-wise metrics for evaluating labeling-induced structural effects. Third, we provide evidence of sector-dependent heterogeneity in baseline OHLCV relevance under alternative target formulations. By shifting focus from predictive accuracy to statistical structure, this work complements existing literature on financial labeling methodologies.

8 Limitations

Several limitations should be acknowledged. First, the analysis is restricted to daily OHLCV data. Intraday data may exhibit different dynamics and stronger volume-related effects. Second, the study focuses exclusively on baseline price-level and volume predictors. Incorporating technical indicators, volatility measures, or macroeconomic variables could alter differentiation patterns.

Third, hypothesis testing is conducted independently within sectors without modeling cross-sector spillovers or temporal regime shifts.

Fourth, the study evaluates distributional separability but does not directly assess predictive model performance under each labeling scheme. Although stronger separation suggests potential performance improvements, this relationship is not explicitly tested here.

9 Future Research Directions

Future research may extend this work in several directions, beginning with the integration of predictive modeling to directly evaluate whether stronger Relative Differentiation (*RD*) and Cumulative Differentiation (*CD*) values translate into improved out-of-sample performance. Furthermore, investigating regime-dependent labeling effects during crisis versus non-crisis periods could reveal how market volatility impacts structural separability. The analysis could also be extended to multi-class labeling frameworks, including trend-continuity definitions as described by Wu et al. [8]. Ultimately, a deeper understanding of how target construction interacts with market structure may contribute to more robust and economically meaningful financial machine learning systems.

Acknowledgments. This research was supported by the Slovak Research and Development Agency (APVV) under Grant No. APVV-24-0446.

References

1. Kim, K.: Financial time series forecasting using support vector machines. Neurocomputing. **55**(1–2), 307–319 (2003)
2. Huang, W., Nakamori, Y., Wang, S.: Forecasting stock market movement direction with SVM. Comput. Oper. Res. **32**(10), 2513–2522 (2005)
3. Ballings, M., Van den Poel, D., Hespeels, N., Gryp, R.: Evaluating multiple classifiers for stock price direction prediction. Expert Syst. Appl. **42**(20), 7046–7056 (2015)
4. Patel, J., Shah, S., Thakkar, P., Kotecha, K.: Predicting stock market index using fusion of machine learning techniques. Expert Syst. Appl. **42**(4), 2162–2172 (2015)
5. Krauss, C., Do, X.A., Huck, N.: Deep neural networks, gradient-boosted trees, random forests: statistical arbitrage on the S&P 500. Eur. J. Oper. Res. **259**(2), 689–702 (2017)
6. Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M.: Financial time series forecasting with deep learning: a systematic review. Appl. Soft Comput. **90** (2020)
7. Zhao, Z., Rao, R., Tu, S., Shi, J.: Time-weighted LSTM model with redefined labeling. In: ICTAI (2017)
8. Wu, D., Wang, X., Su, J., Tang, B., Wu, S.: A labeling method for financial time series prediction based on trends. Entropy. **22** (2020)
9. Kovačević, M., et al.: Optimal trend labeling in financial time series. IEEE Access. **11**, 83822–83832 (2023)
10. López de Prado, M.: Advances in Financial Machine Learning. Wiley (2018)
11. Tsinaslanidis, P.E., Kugiumtzis, D.: PIPs and DTW prediction scheme. Expert Syst. Appl. **41** (2014)
12. Rapach, D.E., Strauss, J.K., Zhou, G.: International stock return predictability. J. Financ. Econ. **111**(1), 1–22 (2014)
13. Bhojraj, S., Swaminathan, B.: Macromomentum. J. Finance. **61**(1), 429–451 (2006)
14. Chordia, T., Swaminathan, B.: Trading volume and cross-autocorrelations. J. Finance. **55**(2), 913–935 (2000)
15. Gervais, S., Kaniel, R., Mingelgrin, D.H.: The high-volume return premium. J. Finance. **56**(3), 877–919 (2001)

Geometric and Topological Perspectives on Financial Market Dynamics

Jakub Marták(✉), Michal Munk, Daša Munková, and Lívia Kelebercová

Constantine the Philosopher University in Nitra, 949 01 Nitra, Slovakia
{jakub.martak,mmunk,dmunkova,lkelebercova}@ukf.sk

Abstract. This study investigates the application of Topological Data Analysis (TDA) as a quantitative framework for detecting structural regime shifts in financial markets. Using intraday stock market data from 2016 to 2025, grouped by economic sectors, we transform price series into logarithmic returns and construct rolling daily time windows. For each window, persistent homology is computed on the corresponding point clouds, and temporal structural changes are quantified using the Wasserstein distance between consecutive persistence diagrams. The resulting time series of Wasserstein distances is compared with sectoral price dynamics to evaluate the correspondence between topological transitions and major market events. The analysis reveals that significant financial disruptions – most notably the 2020 global market crash associated with the COVID-19 pandemic – are consistently characterized by sharp increases in Wasserstein distance across all sectors. Additional sector-specific structural changes are detected during periods of macroeconomic instability, including the 2015–2016 selloff, the 2018 cryptocurrency crash, the 2022 geopolitical crisis, and subsequent regional market downturns. The findings suggest that persistent homology captures multiscale structural reorganizations in financial return distributions and provides a robust, model-independent indicator of market regime shifts. By focusing on geometric and topological invariants rather than solely on statistical moments or volatility measures, TDA offers a complementary perspective for financial risk assessment and systemic instability detection. Future research may further integrate topological descriptors with predictive modeling frameworks to enhance early-warning capabilities in financial markets.

Keywords: Topological Data Analysis · Persistent Homology · Market Stock

1 Introduction

Financial markets are characterized by high volatility and complex nonlinear dynamics, where substantial price movements may occur within very short time intervals. Traditional approaches to technical analysis rely on visual inspection of charts and heuristic pattern recognition. Although such methods may provide insights, they are often computationally intensive, subjective, and difficult to scale to high-dimensional datasets.

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 635–648, 2027.
https://doi.org/10.1007/978-3-032-32104-6_53

Topological Data Analysis (TDA) comprises a collection of methodologies grounded in algebraic topology for extracting multiscale structural information from complex datasets. In many real-world applications, including financial markets, data are high-dimensional, noisy, incomplete, and often concentrated near lower-dimensional manifolds. The fundamental premise of TDA is that meaningful information is encoded in the intrinsic shape of the data, rather than solely in pairwise distances or local statistical summaries.

Topology [1] is a mathematical discipline concerned with studying the properties of objects that are preserved under continuous transformations, particularly their connectivity and neighborhood relations. Together with algebra, topology forms one of the foundations of modern mathematics. In topology, a square and a circle are considered equivalent because one can be continuously transformed into the other without cutting or gluing.

The general workflow of TDA consists of:

1. data collection;
2. construction of a geometric representation (e.g., a graph) from the data;
3. extraction of topological properties (e.g., number of vertices, edges, connected components);
4. analysis based on these properties (ranging from basic data mining to advanced machine learning).

One of the central components in TDA is the Mapper algorithm, which provides a compressed representation of the topological structure of high-dimensional data. The Mapper algorithm transforms data into an interactive graphical representation that (a) facilitates exploratory data analysis and (b) identifies structural patterns within the data.

The general steps of the Mapper algorithm are as follows:

1. begin with data represented in an n-dimensional coordinate system, where n is the number of variables;
2. reduce the dimensionality using a suitable technique (e.g., principal component analysis or factor analysis);
3. partition the reduced data into overlapping subsets;
4. return to the original high-dimensional data and apply a clustering method (e.g., k-means, k-nearest neighbors) within each subset;
5. construct a graph in which nodes represent clusters and edges represent overlaps between clusters [2, 3].

The resulting graph provides insight into the global structure of the data, including branching behavior, loops, and connected components.

Persistent homology constitutes one of the principal computational tools of TDA. It enables the identification of topological features across multiple spatial scales and provides a principled mechanism for distinguishing structural signal from noise. Its goal is to identify fundamental topological properties of the data that are robust to noise and extreme values.

A triangle [4] is a fundamental geometric object because any two-dimensional geometric shape can be decomposed into triangles. The generalization of this concept to n-dimensional space is called an *n-simplex*. A collection of simplices forms a *simplicial*

complex. For example, both a triangle and a torus contain a hole. In topology, shapes that can be continuously deformed into one another are considered equivalent; this equivalence is called a homeomorphism. The concept of holes generalizes to higher dimensions and forms the basis of homology groups. When discussing homology, we are essentially analyzing holes in various dimensions.

The general steps of Persistent Homology analysis are as follows:

1. convert the data into a simplicial complex by constructing n-dimensional spheres around data points;
2. if two spheres intersect, create 1-simplices (line segments) connecting the corresponding points;
3. gradually increase the radius of the spheres, allowing higher-dimensional simplices to form;
4. continue this process until the structure becomes fully connected.

The evolution of topological features is captured and quantified by tracking connected components, loops, and higher-dimensional holes across scales. The significance of these features is determined by their persistence: features that exist across a wide range of scales are considered more important than those that disappear quickly [5].

In mathematics, the Wasserstein distance is a distance function defined between probability distributions on a given metric space M. If each distribution is viewed as a unit amount piled on M, the metric is the minimum "cost" of turning one amount into the other. While other topological summaries (e.g., Betti numbers, Euler characteristics, barcodes, persistence landscapes) may be employed in our research as well, the Wasserstein distance provides a continuous and metrically stable measure of structural change between consecutive time segments [2, 5, 6]

$$l_1(u, v) = \int_{-\propto}^{\propto} |x - y| d\pi(x, y), \tag{1}$$

where $\Gamma(u, v)$ is the set of (probability) distributions on $R \times R$ whose margins are u and v on the first and second factors respectively. For a given value x, $x(u)$ gives the probability of u at position x, and the same for $v(x)$ [6].

The objective of this study is to investigate TDA as a mathematically rigorous framework for detecting structural patterns in financial data and for assessing the likelihood of future market movements. To investigate TDA's advantages and disadvantages for researching market developments over a long period of time.

Rather than relying solely on metric properties of the data, TDA emphasizes qualitative geometric and topological structures that remain invariant under continuous transformations.

This study is structured as follows. The Related Work section reviews existing research applying Topological Data Analysis in various scientific domains. The Materials and Methods section describe the acquisition and preprocessing of financial data. The Results section presents the resulting graphs illustrating price movements across different economic sectors and compares them with model predictions. Finally, the Discussion section interprets and evaluates the findings.

2 Related Work

The application of TDA has expanded rapidly across finance, dynamical systems, biomedicine, and molecular sciences. In this section, we review representative contributions that illustrate the methodological breadth and empirical relevance of TDA.

Tsung-Wen Yen et al. [5] applied TDA and persistent homology to the analysis of stock markets in Singapore and Taiwan. Using data from the Taiwan Capitalization Weighted Stock Index (TAIEX), they computed persistence barcodes, persistence diagrams, persistent entropy, bottleneck distances, Betti numbers, and Euler characteristics. Their results demonstrated that during periods of financial instability and market crashes, homological features exhibited reduced persistence. In both markets, they identified consistent topological signatures associated with crash events, reflected in the behavior of Betti numbers, Euler characteristics, and persistent entropy. These findings suggest that topological invariants may serve as structural indicators of systemic financial stress.

De Silva et al. [7] developed a topological framework for the analysis of recurrent dynamical systems. Their approach formalizes periodic, quasiperiodic, and recurrent behaviors as topological features detectable via persistent cohomology. By reconstructing phase spaces and analyzing their multiscale topological structure, they demonstrated that persistent (co)homological invariants provide a robust representation of recurrent dynamics, offering an alternative to purely spectral or metric-based analyses.

Bukkuri et al. [4] investigated applications of TDA in oncology, with particular emphasis on persistent homology. They reported several successful applications, including prediction of treatment response and prognosis, tumor segmentation, computer-aided diagnosis, disease classification, and characterization of cellular architecture. Their work highlights the capacity of TDA to extract clinically relevant structural information from high-dimensional biomedical data and outlines future research directions for integrating topological techniques into cancer analytics.

Rabadán et al. [8] employed TDA to identify functionally relevant genetic alterations in cancer. While certain mutations occur frequently across tumors, many appear only in small subsets of samples. By applying topological methods to integrated genomic datasets, the authors identified 38 cancer-associated genes, including previously unreported alterations. Their results demonstrate that topological integration of heterogeneous biological data enhances the detection of rare but structurally significant genomic events.

Nielson et al. [9] applied TDA to preclinical datasets involving traumatic brain injury and spinal cord injury. Their data-driven topological analysis identified patterns across a complex syndromic network, revealing interactions between spinal cord and traumatic brain injuries, as well as adverse drug effects. Notably, their results indicated that perioperative hypertension was a stronger predictor of long-term recovery than any of the pharmacological interventions tested. This study illustrates the utility of TDA in uncovering multivariate dependencies in high-dimensional biomedical data.

Shultz [10] provided a conceptual overview of TDA in the context of economic research. He compared topological approaches with traditional econometric methods, discussed core TDA methodologies (including Mapper and persistent homology), and

highlighted potential avenues for future applications in economics. His work emphasizes the methodological complementarity between topological invariants and classical statistical inference.

Finally, Wee and Jiang [11] presented a comprehensive survey of the development, methodology, and applications of TDA in molecular sciences. They traced the evolution of TDA from qualitative geometric descriptors to quantitative and predictive frameworks incorporating persistent homology, persistent Laplacians, and topological machine learning. Their review documents the transformative impact of TDA in biomolecular stability analysis, protein–ligand interactions, drug discovery, materials science, topological sequence analysis, and viral evolution. Furthermore, they discussed recent advances in integrating TDA with machine learning and artificial intelligence, enabling predictive modeling in protein engineering, solubility and toxicity estimation, and novel material discovery. They have also discussed the limitations of TDA.

Collectively, these studies demonstrate that TDA provides a mathematically principled and domain-agnostic framework for extracting multiscale structural information from complex, high-dimensional data. The diversity of applications – from financial markets to genomics and neuroscience – underscores its potential relevance for modeling and interpreting financial time-series data.

3 Materials and Methods

This section describes the data source, dataset construction, and preprocessing procedures employed prior to the application of TDA.

3.1 Dataset

The dataset (Table 1) was obtained from Yahoo!® Finance historical records for selected equity tickers covering the period 2016–2025. Data acquisition was performed using the *yfinance* Python library, which provides programmatic access to publicly available financial time-series data. For each ticker, intraday time-series data were retrieved with the following attributes: timestamp (open_date), open price, high price, low price, close price, and traded volume (volume_base).

Table 1. Example of file *DIS.csv* (header and first five rows).

open_date	open	high	low	close	volume_base
2016-01-04 09:30:00	96.33	96.39	95.23	95.87	2621116
2016-01-04 10:00:00	95.85	95.86	95.3	95.74	946225
2016-01-04 10:30:00	95.73	95.8	95.25	95.3	758515
2016-01-04 11:00:00	95.29	95.51	95.07	95.32	934236
2016-01-04 11:30:00	95.32	95.32	94.8	94.97	1009872
…	…	…	…	…	…

3.2 Data Preprocessing

All preprocessing procedures were implemented in Python using standard data analysis libraries (e.g., *NumPy* and *pandas*). The preprocessing pipeline consisted of data cleaning, sectoral classification, filtering for temporal completeness, and variable selection.

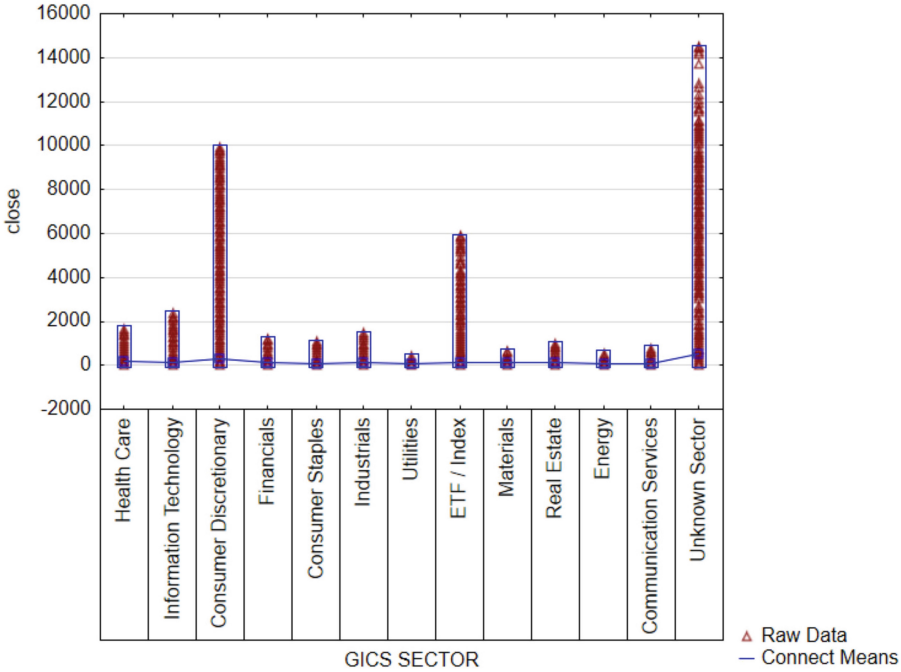


Fig. 1. Variability plot for close value of acquired data.

The initial dataset contained over 500 files. These were categorized according to the Global Industry Classification Standard (GICS) sectors in which the respective corporations operate: Communication Services, Consumer Discretionary, Consumer Staples, Energy, ETF Index, Financials, Health Care, Industrials, Information Technology, Materials, Real Estate, Utilities, and Unknown Sector (Fig. 1).

Tickers were excluded from further analysis if:

1. they contained substantial missing values;
2. they did not span the entire predetermined observation window (2016–2025);
3. the corresponding company was established after the beginning of the study period.

The "Unknown Sector" category was also excluded due to excessive heterogeneity and inconsistent temporal coverage. This sector consisted only of a few unrelated tickers, that did not belong to any other category. After filtering, approximately 27% of the original time-series data (in terms of total observations) were retained for TDA (Fig. 2).

However, approximately 93% of tickers remained usable after removing incomplete or structurally inconsistent records, exactly 500 tickers.

Examples of ticker names from various sectors include: AAPL (Apple), AMZN (Amazon.com, Inc), BAC (Bank of America Corporation), CMS (CMS Energy Corp), DVA (DaVita Inc.), EA (Electronic Arts), KO (The Coca-Cola Company), NEM (Nemetschek), PSX (Phillips 66), UDR (United Dominion Realty Trust), VGT (Vanguard Information Technology ETF), WAB (Westinghouse Air Brake Technologies Corporation) (Table 2).

Table 2. Distribution of tickers across individual sectors.

Communication Services	19
Consumer Discretionary	47
Consumer Staples	36
Energy	21
ETF Index	31
Financials	72
Health Care	56
Industrials	69
Information Technology	66
Materials	26
Real Estate	29
Utilities	28

Each dataset contained five primary numerical variables: open, high, low, close, and volume_base. For sector-level aggregation, these columns were retained and aligned across tickers.

For subsequent topological analysis and comparative testing, a univariate representation was required at each stage. In particular, the close price was selected as the principal variable of interest, as it represents the final transaction price within a given interval and is commonly used in financial time-series modeling. The price variables (open, high, low, close) exhibit strong convergence within sectors due to their shared dependence on underlying price dynamics. In contrast, volume_base, representing the total number of traded shares or contracts within a time interval, displays substantially different statistical behavior and variance structure (Fig. 3).

The preprocessing stage thus transforms raw financial log files into structured, sector-specific time-series datasets suitable for geometric and topological modeling.

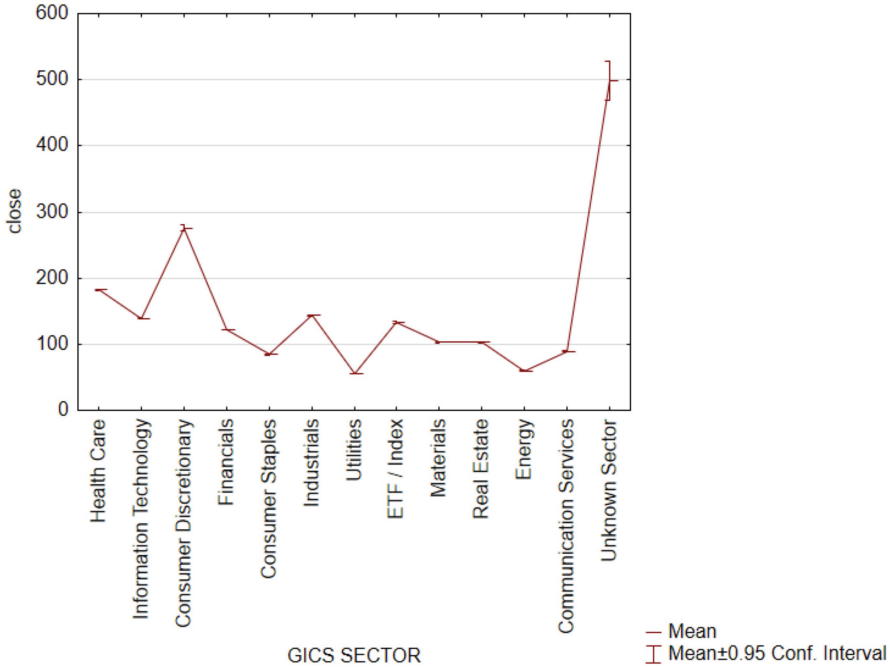


Fig. 2. Means plot for close value of acquired data.

4 Results

The first step of the analysis consisted of transforming raw price levels into logarithmic returns. It provides relative price changes; it transforms stock prices into a sequence representing the trend in the stock's price [2].

To compute persistent homology, the data were partitioned into rolling time windows. In this study, each window corresponds to one trading day, comprising 13 observations. Persistent homology was computed on each window independently.

Function *Rips()* from *ripser* library for *python* initializes the persistence diagrams. We have used input parameter "*maxdim* = *2*" which means, that only 2-dimensional persistence diagrams are allowed, this was due to limit of computation power of our hardware. For each time window, we compute two persistence diagrams with the function *rips.fit_transform* that uses list of into logarithmic returns for the single time windows as an input parameter.

To quantify temporal changes in topological structure, we computed the Wasserstein distance between persistence diagrams derived from consecutive time windows. The Wasserstein distance measures the minimal "transport cost" required to match topological features between two diagrams. It measures the minimal effort to match features from diagram 1 to diagram 2 (minimal moving cost). We computed Wasserstein distance via *persim.wasserstein* function, that uses two diagrams and a Boolean, whether or not should the function return matching information and cross-similarity matrix.

For each sector, we plotted:

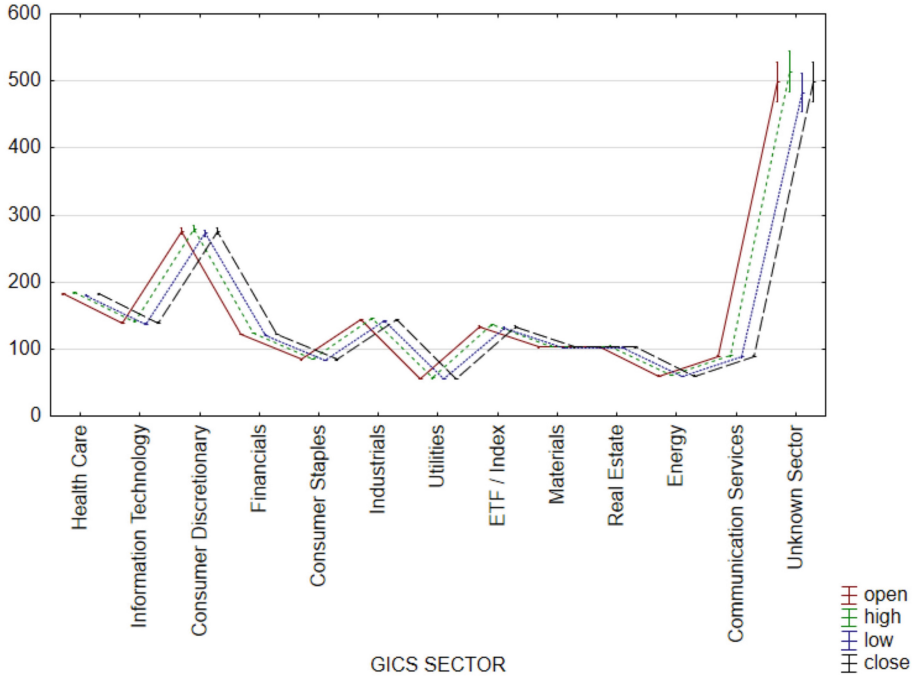


Fig. 3. Means plot for multiple variables of acquired data.

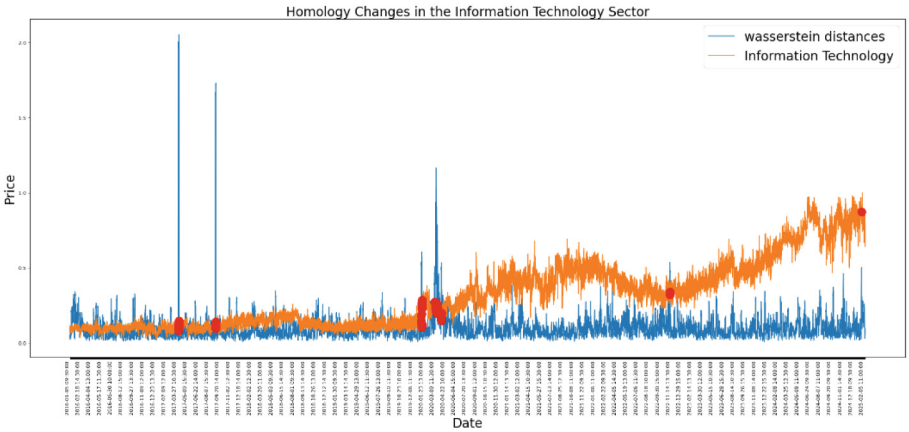
1. the temporal evolution of Wasserstein distances from 2016 to 2025;
2. the weighted average of close prices within the sector, normalized by the maximum observed value during the period;
3. markers indicating the largest price jumps between neighboring time windows.

The red markers represent statistically significant price discontinuities (i.e., extreme log-returns) exceeding a predefined threshold. These serve as reference points for comparing metric price volatility with topological variability.

The Information Technology sector (Fig. 4) exhibit moderate long-term growth with several pronounced homological transitions. Significant increases in Wasserstein distance occur during 2016–2017, early 2020 (marked with red points on the graphs). The late 2022 structural change exceeds the 0.5 threshold and is followed by accelerated growth, indicating a regime shift detectable through persistent homology.

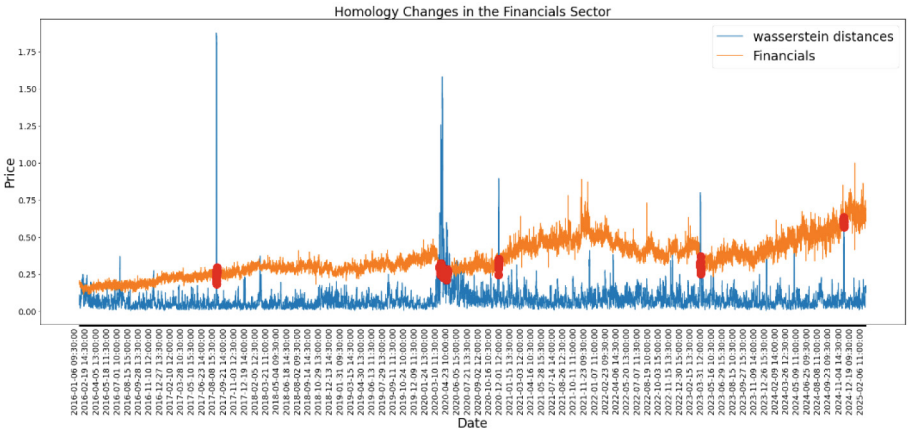
The Financials sector (Fig. 5) display modest growth trajectories. The Financials sector shows weaker overall upward momentum. It experiences two major homological disruptions in 2020, as well as additional significant changes in: 2016–2017 and late 2024 (marked with red points on the graphs). The concurrence of these topological shifts with macroeconomic stress periods suggests that persistent homology captures systemic instability across the sector.

The Real Estate sector (Fig. 6) shows comparable long-term behavior, with substantial structural changes occurring primarily in: 2016 and 2020. The second half of the study period is characterized by several large fluctuations in Wasserstein distance,



Note: marked points - the largest leaps in the prices

Fig. 4. Homology Changes in the Information Technology Sector.



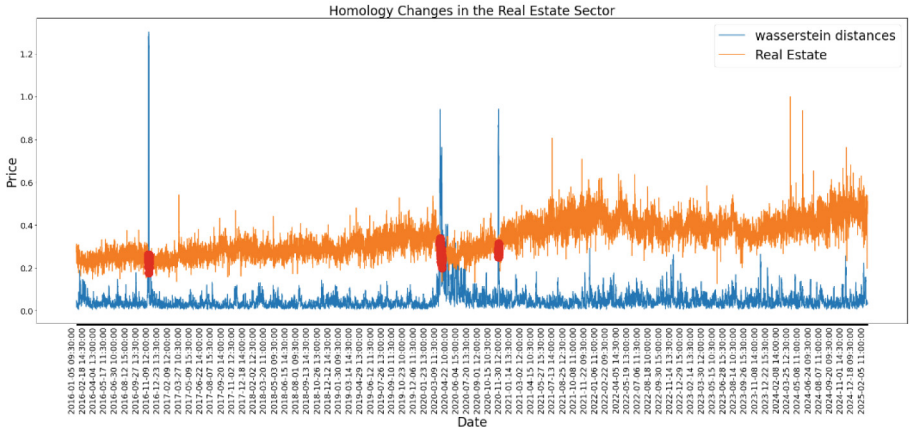
Note: marked points - the largest leaps in the prices

Fig. 5. Homology Changes in the Financial Sector.

indicating episodic structural reorganizations in return distributions (marked with red points on the graphs).

5 Discussion

Implications of TDA include robustness to noise and outliers. A core strength of persistent homology is the ability to, distinguishing significant structures from unsignificant. TDA can also analyze data across varying scales simultaneously, allowing it



Note: marked points - the largest leaps in the prices

Fig. 6. Homology Changes in the Real Estate Sector.

to capture local and global data behaviors. TDA provides intuitive visual representations of complex, high-dimensional data, making them more interpretable even to non-mathematicians, as was the case in this paper. Graphical representation of the homology changes clearly showed the largest leaps in the prices.

Limitations of Persistent homology include high computational cost, as was the case in this paper. Its calculations can be exponentially complex relative to the number of data points, that's why we were only able to compute persistent diagrams with 2 dimensions. The results of the Persistent Homology depend heavily on choices like the filter function, proximity metrics, and resolution parameters. TDA can often identify the existence of a topological feature, but it can be challenging to determine its exact location, that's why we have used the scatterplot in our graphs, for greater emphasis. Related to that, this methodology is based on abstract mathematical concepts, which requires expertise in algebraic topology. In the end, TDA may not improve predictive power over simpler, well-known models in scenarios where the structure is not inherently topological [13].

Wasserstein distance is generally considered a robust indicator of systemic transitions. It can be both lagging and leading indicator, based on the circumstances. It is generally classified neither as lagging nor as leading indicator in traditional sense. It can be used as lagging indicator for measuring the changes in historical data, but it can also be used as a leading indicator in anomaly detection (1).

A comparative analysis of sectoral graphs reveals a common structural feature: a pronounced increase in the Wasserstein distance followed by a substantial decline in prices in February 2020 (Figs. 4, 5, and 6). This phenomenon coincides with the global financial shock triggered by the COVID-19 pandemic [14]. Across all sectors, persistent homology detected abrupt reconfigurations in the topology of log-return point clouds, reflected in sharp peaks of Wasserstein distances between consecutive time windows. These peaks indicate rapid structural transitions in market dynamics, consistent with extreme volatility and systemic uncertainty.

Earlier significant shifts are observable in 2016, corresponding to the 2015–2016 stock market selloff (Figs. 4, 5, and 6). This period was characterized by multiple macroeconomic stressors, including slowing GDP growth in China, a sharp decline in petroleum prices (falling below USD 27 per barrel), the Greek sovereign debt crisis, and the political and economic uncertainty surrounding Brexit in United Kingdom. The corresponding elevations in Wasserstein distance suggest that TDA captured structural fragmentation in return distributions during this multi-factor instability [12].

Another observable structural disturbance appears in 2018, associated with the Cryptocurrency crash, also known as the Bitcoin crash, following the speculative boom of 2017. Although primarily centered on digital asset markets, the spillover effects appear detectable in several traditional sectors (Fig. 4), indicating that topological signatures may reflect cross-market contagion [15].

The year 2022 exhibited renewed market turbulence following the January peak. This period coincides with the Russian invasion of Ukraine [16, 17], triggered the 2022 Russian stock market crash and broader geopolitical and economic instability (Fig. 4).

In 2023, a sector-specific decline is visible in the financial sector (Fig. 5), despite 2023 being widely regarded as a strong year for global equity markets. This divergence highlights the sectoral sensitivity of topological measures and suggests that persistent homology may detect localized structural stress not immediately evident in aggregate indices.

Further moderate declines are observable in 2024 within the Financial sector (Fig. 6). These fluctuations may be partially attributable to financial instability in Asian markets, including downturns in China and Japan. Contributing factors include the sluggish post-pandemic recovery of the Chinese real estate market and the appreciation of the Japanese yen, which led to the unwinding of yen carry trades and subsequent global selloffs [18, 19]. The associated increases in Wasserstein distance indicate detectable topological restructuring during these periods.

Finally, the 2025 market downturn (Fig. 4) aligns with trade policy tensions initiated by tariff announcements under Donald Trump, followed by retaliatory tariffs imposed by China [20]. The resulting structural shifts were again captured by peaks in Wasserstein distance, reinforcing the sensitivity of persistent homology to macroeconomic and geopolitical shocks.

These findings suggest that persistent homology captures regime transitions characterized by rapid structural reorganization in financial return distributions.

6 Conclusion

This study investigated the applicability of Topological Data Analysis as a quantitative framework for detecting market crashes and significant structural shifts in financial markets. Using sectoral financial data from 2016 to 2025, we demonstrated that persistent homology, combined with Wasserstein distances between consecutive persistence diagrams, provides a robust indicator of systemic transitions.

The empirical results show that major global financial disruptions – ranging from localized market downturns to the global crash induced by the COVID-19 pandemic –are consistently reflected as sharp topological transitions. Less severe, sector-specific

disturbances are also detectable, indicating that TDA captures both global and localized structural changes.

While financial markets remain inherently complex and partially unpredictable, the results indicate that topological descriptors offer complementary insights beyond traditional volatility or correlation-based measures. By focusing on the intrinsic geometric structure of return distributions, TDA provides a multiscale perspective on market dynamics.

Future research should continue to explore the possibilities to improve the ways, how to used TDA methods to explore market conditions to further optimize the financial stability.

Acknowledgement. This research was supported by the Slovak Research and Development Agency (APVV) under Grant No. APVV-24-0446.

References

1. Wee, J., Jiang, J.: A review of topological data analysis and topological deep learning in molecular sciences. J. Chem. Inf. Model. **65**(23), 12691–12706 (2025)
2. Bukkuri, A., Andor, N., Darcy, I.: Applications of topological data analysis in oncology. Front. Phys. **4** (2021). https://doi.org/10.3389/frai.2021.659037
3. Almgren, K., Kim, M., Lee, J.: Mining social media data using topological data analysis. In: 2017 IEEE International Conference on Information Reuse and Integration (IRI), pp. 144–153. IEEE (2017)
4. medium.com: https://medium.com/@kennywang2003/topological-data-analysis-mapper-f0c44ecbbb3d. Last accessed 7 Feb 2026
5. Wasserman, L.: Topological data analysis. Annu. Rev. Stat. Appl. **5**, 501–532 (2018)
6. docs.scipy.org: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wasserstein_distance.html. Last accessed 29 Mar 2026
7. Skraba, P., de Silva, V., Vejdemo-Johansson, M.: Topological analysis of recurrent systems, pp. 1–5 (2012). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-107210
8. Rabadán, R., Mohamedi, Y., Rubin, U., et al.: Identification of relevant genetic alterations in cancer using topological data analysis. Nat. Commun. **11**, 3808. https://doi.org/10.1038/s41467-020-17659-7
9. Nielson, J.L., et al.: Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. Nat. Commun. **6**(1), 8581 (2015)
10. Shultz, C.: Applications of topological data analysis in economics (2023). Available at SSRN 4378151.
11. Yen, P., Cheong, S.: Using Topological Data Analysis (TDA) and persistent homology to analyze the stock markets in Singapore and Taiwan. Front. Phys. **9** (2021). https://doi.org/10.3389/fphy.2021.572216
12. Shan L., Andrea C., Paresh D.: Stock market suffers worst one-day drop since 2008. Los Angeles Times (2015). Last updated 15 Aug 2015
13. Yara, S., Reinhard, L.: Topological data analysis in biomedicine: a review. J. Biomed. Inform. **130**, 104082 (2022). https://doi.org/10.1016/j.jbi.2022.104082
14. Coronavirus creates a true 'bear market' for Wall Street. Daily Breeze. Last updated 13 Mar 2013
15. Bradsher, K.: China's economy, by the numbers, is worse than it looks. The New York Times (2019). Last updated 19 Jan 2019

16. Thompson M., Liakos C.: Russian stocks crash 33% and ruble plunges to record low. CNN (2022). Last updated 25 Aug 2022
17. Gura, D.: Stocks sink, sending the S&P 500 to a bear market. National Public Radio (2022). Last updated 14 June 2022
18. Economists: China's stock market nightmare is nowhere near over the situation ought to worry Xi Jinping. Economists (2024). Last updated 7 Feb 2024
19. He L., Stewart M., Thompson M.: Global stock rout intensifies with Japan's Nikkei and other Asian markets plunging. CNN (2024)
20. Grantham-Philips W. Trump's tariffs have launched global trade wars. Here's a timeline of how we got here. AP News (2025). Last updated 10 Apr 2025

Enhancing Time Series Forecasting Using Persistent Homology

Patrik Sebes(✉), Martin Drozda, and Maros Cavojsky

Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Bratislava, Slovakia
{patrik.sebes,martin.drozda,maros.cavojsky}@stuba.sk

Abstract. Accurate forecasting of financial time series remains a challenging problem due to noise, nonstationarity, and complex market dynamics. In this paper, we investigate whether features derived from Persistent Homology, a method from Topological Data Analysis, can improve deep learning stock price forecasting. We extract topological features from sliding windows of historical price data using time-delay embedding and combine them with standard market features and technical indicators. We evaluate our approach on daily stock data from the U.S. market using several state-of-the-art forecasting models, including transformer-based, convolutional, linear, and tree-based architectures. Our experimental results demonstrate that technical indicators provide the most consistent and statistically significant performance improvements across assets and models. In contrast, Persistent Homology features generally failed to outperform the baseline and, when combined with technical indicators, often diluted their predictive power. Overall, our findings suggest that standard topological feature engineering may act as high dimensional noise in this specific forecasting setup, highlighting the need for more specialized strategies or alternative applications such as pattern recognition.

Keywords: Algorithmic Trading · Deep Learning · Feature Extraction · Topological Data Analysis · Persistent Homology

1 Introduction

Time series forecasting aims to predict future values of a temporally ordered signal using historical observations. Although deep learning models have demonstrated strong empirical performance, accurate forecasting remains challenging in settings characterized by nonstationarity, irregular sampling, noisy measurements, or latent regime changes. In such cases, methods relying solely on local temporal patterns or frequency-based features may fail to capture the global structural properties of the underlying dynamics.

Persistent homology (PH) is a method in Topological Data Analysis (TDA) that provides a principled way to quantify the underlying shape of data. Rather

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 649–661, 2027.
https://doi.org/10.1007/978-3-032-32104-6_54

than focusing on individual observations or local statistics, PH analyzes how topological structures, such as connected components and loops, emerge and persist across multiple scales. These structures are summarized using representations such as persistence diagrams and barcodes [1,16].

Importantly, these topological summaries can be featurized into numerical representations, which can then be used as additional input features for standard machine learning and deep learning models.

2 Related Work

The application of Topological Data Analysis (TDA) to time series forecasting has gained significant traction, evolving from simple feature extraction to complex integration within deep learning architectures. Existing literature can be broadly categorized into methods that modify neural network architectures to process topological information and methods that utilize topological descriptors as supplementary input features.

2.1 Deep Learning Integration and Architectural Innovations

Several studies have proposed integrating persistent homology (PH) directly into the learning mechanism of neural networks. Zeng et al. [15] introduced *topological attention*, a mechanism that computes PH features on short, overlapping windows and uses an attention layer to weigh their contribution. They reported state-of-the-art performance on the M4 benchmark by using a specialized architectural design.

Building on this direction, Lin et al. [6] proposed the Topological Information Supervised (TIS) framework. They utilized a conditional Generative Adversarial Network (GAN) to synthesize topological features and introduced a topological consistency loss to guide the training process. Their results showed consistent gains across various backbones, suggesting that enforcing topological constraints can improve model robustness. Beyond supervised forecasting, Kim et al. [5] introduced *TopoCL*, which treats temporal and topological summaries as distinct modalities for contrastive representation learning. This approach aims to learn invariant representations useful for downstream tasks, including forecasting.

In the spatio-temporal domain, Nguyen et al. [9] constructed multi-scale graphs from PH filtrations and employed an attention-routed ensemble of graph neural networks, demonstrating the utility of topology in capturing complex spatial dependencies in traffic data.

2.2 Topological Feature Engineering in Finance

While architectural modifications offer theoretical novelty, feature engineering approaches focus on extracting robust descriptors to enhance standard models. This direction is particularly relevant to our work. In the financial domain, de

Jesus Jr. et al. [3] demonstrated that augmenting neural baselines with persistence diagram statistics (such as feature counts and entropy) improves forecasting accuracy across multiple cryptocurrency and traditional instrument datasets. Similarly, Heo and Jung [2] proposed a method where domain knowledge modulates PH computations via an influence vector, allowing for more targeted feature extraction.

2.3 Differentiation of Our Work

Although previous works like Zeng et al. [15] and Lin et al. [6] show the potential of TDA, they often rely on complex, custom architectures that may be difficult to deploy or interpret. Conversely, while de Jesus Jr. et al. [3] explored feature engineering, there is a lack of comprehensive evaluation regarding how vectorized topological features interact with the most recent state-of-the-art Transformer-based models (e.g., PatchTST, iTransformer) specifically in the context of the US stock market. Our work bridges this gap by systematically evaluating a pure feature engineering approach. Instead of modifying the model architecture, we extract a rich set of vectorized topological signatures and test whether they provide complementary information to modern deep learning forecasters and standard technical indicators.

3 Methodology

3.1 Process Workflow

The simplified workflow is summarized in Fig. 1, which shows the main steps and how they flow. Each section is described in more detail later in the article.

3.2 Feature Extraction Process

After we collected the data, we created four feature sets, which are described in more detail in Sect. 5.5. For the Baseline feature set, we work with the features we already obtained. For the Technical Indicators feature set, the mentioned indicators are calculated in the standard way using the TA library [11].

The PH feature extraction process is based on Persistent Homology (PH). First, the time series is transformed into a higher-dimensional space using Takens' Embedding Theorem. This allows us to treat the sequence of prices as a geometric object. Specifically, the univariate time series is mapped into a d-dimensional Euclidean space $\mathbb{R}^d$, where $\mathbb{R}$ denotes the set of real numbers and d represents the *embedding dimension*.

The resulting point cloud is then analyzed using a filtration process. By growing hyperspheres of radius ϵ around each point, we track the birth and death of topological features. The parameter ϵ represents a distance threshold in the embedded space. We focus on features such as connected components (H_0) and loops (H_1). Features that persist over a wide range of ϵ values are

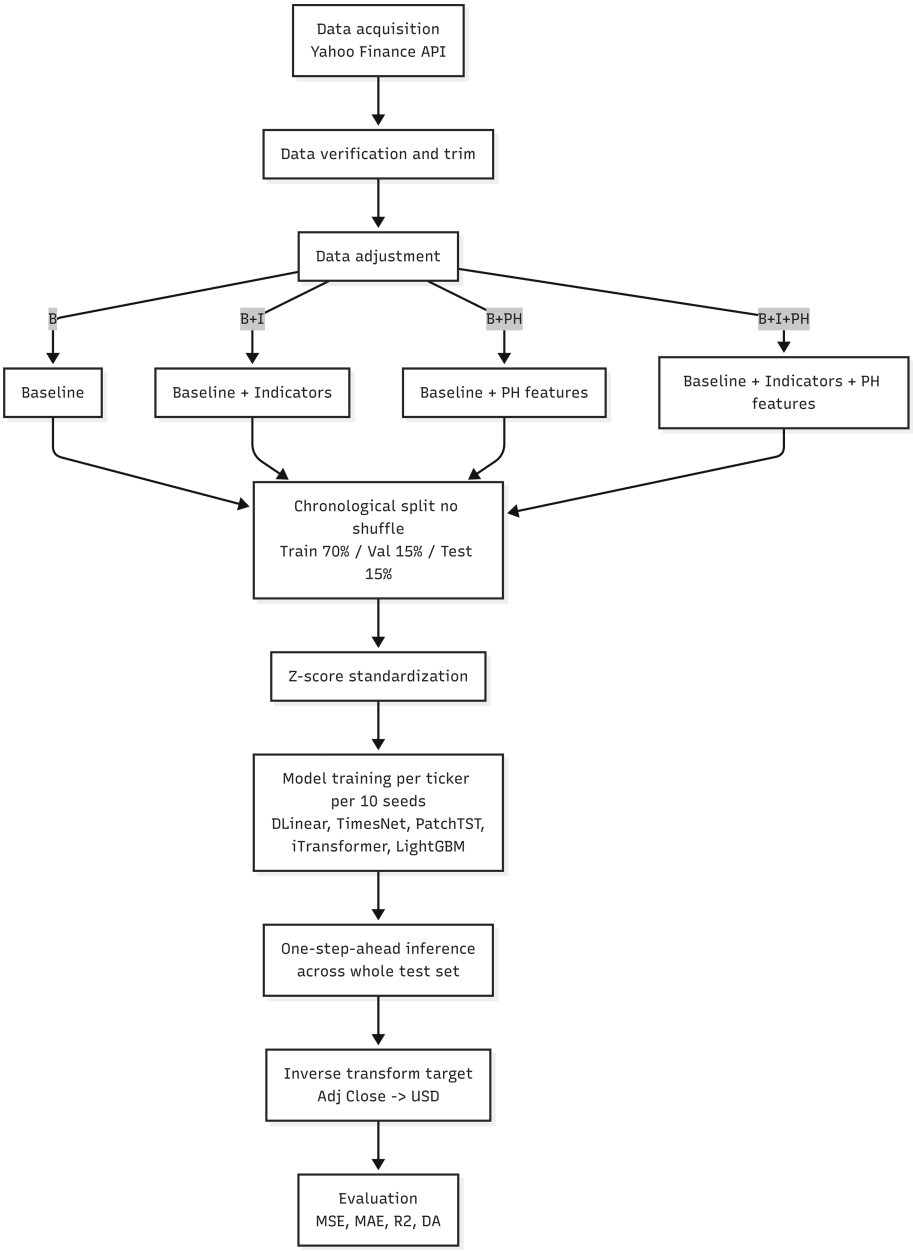


Fig. 1. Simplified Process Workflow Diagram.

located far from the diagonal in the Persistence Diagram and are considered robust topological signatures.

Our methodology follows a standard TDA workflow. In **Phase 1**, we segment the time series into sliding windows of length w (where w represents the number of consecutive trading days, e.g., $w = 60$) and apply time-delay embedding to generate a point cloud in $\mathbb{R}^d$. **Phase 2** utilizes persistent homology to track the evolution of topological components (H_n) across changing scales of ϵ, capturing the structural shape of market dynamics. Finally, in **Phase 3**, we vectorize these descriptors into a fixed-length numerical feature vector $V \in \mathbb{R}^K$ (where K is the total number of extracted topological statistics), ensuring compatibility with the downstream predictive model.

We extracted a total of 24 numerical features (12 for each homology dimension H_0 and H_1) to vectorize the topological structure. Let $\mathcal{D}_k = \{(b_i, d_i)\}_{i=1}^N$ be the persistence diagram for dimension k, consisting of N birth-death pairs. We define the *persistence* (lifetime) of the i-th feature as $l_i = d_i - b_i$. The extracted features are:

1. **Count:** The total number of topological features N in the diagram (cardinality of $\mathcal{D}_k$), representing the overall topological activity.
2. **Persistence Statistics (Sum, Mean, Std, Max):** Summary statistics derived from the set of lifetimes $\{l_i\}$. *Sum* represents the total persistence; *Mean* and *Std* capture the average duration and variability; *Max* identifies the most dominant structure (e.g., the most significant loop).
3. **Persistent Entropy:** A measure of the diversity of feature lifespans, calculated as $E = -\sum p_i \log(p_i)$, where $p_i = l_i / \sum l_j$. High entropy indicates many features of similar importance (noise), while low entropy suggests a few dominant signals.
4. **Timeline Statistics (Birth, Death, Midlife):** The average values of the birth times (b_i), death times (d_i), and midlife points defined as $m_i = (b_i + d_i)/2$. These metrics locate *when* topological features typically emerge and vanish during the filtration.
5. **Betti Curve Statistics (Mean, Std, Max):** The Betti curve $\beta_k(\epsilon)$ records the number of active features at filtration step ϵ. We discretized this function and computed its *Mean* (average connectivity), *Std* (fluctuation in connectivity), and *Max* (peak topological complexity) over the filtration range.

3.3 Warmup Period and Window Size

First, a *warmup period size* denoted by *wps* is defined. This parameter represents the maximum number of past time steps required by any technical indicator or persistent homology feature to be computed correctly. The warmup period ensures that, for every time point t, a sufficiently long historical record is available. Without this constraint, feature values at early time points would be undefined or unreliable due to missing past observations. All points $t < t_{wps}$ are therefore excluded from further analysis, as valid feature values cannot be guaranteed in this interval. After the warmup period, features are computed using a sliding window with a step size of one time point. For each time point t, a window of length w is defined over the interval $[t - w, t - 1]$.

Persistent homology features are computed from the data contained in this window and assigned to time point t. These features capture the topological structure of events that occurred prior to t and do not use any future information. This construction preserves temporal causality and ensures consistent feature definitions across the entire time series.

3.4 Training and Testing

After feature extraction, data are split **chronologically** (no shuffling) into train/val/test with a **70/15/15** ratio to preserve temporal structure and prevent look-ahead leakage. Features are z-score standardized. Metrics are reported after inverse transforming the target to the original USD scale. We evaluate one-step-ahead forecasting ($h = 1$). At time t, the model ingests the last L observations and predicts $t+1$. Evaluation proceeds sequentially with a 1-step sliding window: inputs always use ground-truth history and no prediction feedback is used. This is standardly used evaluation process.

4 Problem Formulation

For this paper, we consider the following research questions. We aim to answer these questions using experimental results:

- **RQ1 (Performance Impact)**
  Does the inclusion of Persistent Homology (PH) features provide a statistically significant improvement in forecasting accuracy compared to Baseline or Baseline with Standard Technical Indicators?
- **RQ2 (Architectural Sensitivity and Data Robustness)**
  Is the performance gain from topological features consistent across state-of-the-art architectures, or is it dependent on specific model types? How will it behave with different architectures and data?

5 Experimental Setup

5.1 Data Selection

We focus on the United States (US) stock market using **daily data granularity**, chosen for its high liquidity, maturity, and status as a standard benchmark in financial forecasting. To ensure a representative and high-quality dataset, we selected only large-cap assets with high trading volumes and extensive historical records, thereby avoiding the noise and manipulation associated with low-liquidity stocks. We followed the Global Industry Classification Standard (GICS)[8] to guarantee sector diversity, selecting one representative stock from each of the 11 economic sectors to minimize correlation. Additionally, the S&P 500 index is included to capture systematic market trends.

The selected assets are: XOM (Exxon Mobil) for Energy, APD (Air Products and Chemicals) for Materials, BA (Boeing) for Industrials, AMZN (Amazon) for Consumer Discretionary, PG (Procter & Gamble) for Consumer Staples, JNJ (Johnson & Johnson) for Health Care, JPM (JPMorgan Chase) for Financials, AAPL (Apple) for Information Technology, GOOGL (Alphabet) for Communication Services, NEE (NextEra Energy) for Utilities, and AMT (American Tower) for Real Estate. In addition, we include SPY (SPDR S&P 500 ETF) as an all sectors benchmark to capture broad market dynamics.

5.2 Data Preparation

We compiled a dataset of daily historical data sourced from the Yahoo Finance API [13], spanning from **January 1, 2016, to December 31, 2025**. We utilized the *Adjusted Closing Price* to account for corporate actions such as dividends and stock splits, ensuring a consistent economic signal. Comprehensive validation checks were performed to detect missing values, structural errors, and logical inconsistencies (e.g. High $<$ Low). The dataset successfully passed all quality controls, requiring no action other than trimming to the specified date range.

5.3 Model Selection

To ensure the universality and robustness of the proposed feature engineering approach, we select a diverse set of baseline models representing distinct architectural families, including Transformers, CNNs, and Linear models. Our evaluation prioritizes state-of-the-art architectures published in top-tier conferences (e.g., NeurIPS, ICLR) within the last five years, avoiding older benchmarks like ARIMA. We specifically included both *Channel-Independent* and *Channel-Mixing* approaches to assess how different multivariate modeling mechanisms utilize the added topological features. Additionally, simple linear baselines are employed to verify that complex models provide genuine signal extraction beyond trend extrapolation. The complete list of evaluated models is detailed in Table 1.

Table 1. Selected models

Acronym	Model Name	Architecture
DLinear [14]	Decomposition Linear	Linear/MLP
LightGBM [4]	Light Gradient Boosting Machine	Gradient Boosting
TimesNet [12]	Temporal 2D-Variation Network	CNN
PatchTST [10]	Patch Time Series Transformer	Transformer
iTransformer [7]	Inverted Transformer	Transformer

5.4 Model Settings

In our experiments, we used a fixed set of parameters across all experiments to ensure a fair and consistent comparison. The key parameters used throughout the experiments are summarized in Table 2.

The hyperparameters were selected to provide a consistent and computationally acceptable experimental setup across all feature sets and forecasting architectures. The PH embedding parameters were set to an embedding dimension of 3 and a time delay of 2 as a compromise between representational capacity and computational efficiency, allowing the delay embedding to capture non-trivial local geometric structure without introducing excessive dimensionality or instability in the topological descriptors.

The lookback window was fixed to 60 days to provide sufficient recent historical context for both the PH-based feature extraction and the forecasting models. The warmup period was set to 100 points because the longest technical indicators used in our study, namely SMA100 and EMA100, require 100 past data points for reliable computation. This ensures that all derived features are well-defined before model training and evaluation begin.

We considered a one-step-ahead forecasting horizon, as this is the standard prediction setting. Finally, 100 epochs for training with early stopping patience of 10 was chosen to allow convergence while limiting overfitting and unnecessary computation. The adjusted close price was used as the prediction target because it provides a financially meaningful series corrected for corporate actions such as dividends and stock splits.

Table 2. Key Hyperparameters used in the Experiments.

Hyperparameter	Value
Embedding dimension	3
Time delay	2
Window size	Adaptive
Warmup period size	100
Sliding window step	1
Lookback window	60
Forecast horizon	1
Epochs	100
Early stopping patience	10
Prediction target	Adjusted Close

5.5 Feature Sets

To evaluate the contribution of topological descriptors, we evaluated three distinct feature groups: **Baseline (B)**, comprising raw OHLCV data; **Technical**

Indicators (I), representing standard momentum and trend signals; and **PH features (PH)**, which capture non-linear structural market dynamics via persistent homology. We assess these sets both individually and in combination (e.g. B+I versus B+I+PH) to determine whether topological features provide complementary predictive power. A detailed breakdown of all features is provided in Table 3.

Table 3. List of Features in each Feature set.

Feature set	Group	Features
B (5)	OHLCV (5)	Open, High, Low, Close, Volume
I (11)	Technical Indicators (11)	SMA20, SMA50, SMA100, EMA20, EMA50, EMA100, MACD12, MACD26, RSI14, ATR14, ROC14
PH (24)	H_0 (9)	Count, Entropy, Persistence (Sum, Mean, Std, Max), Birth mean, Death mean, Midlife mean
	H_1 (9)	Count, Entropy, Persistence (Sum, Mean, Std, Max), Birth mean, Death mean, Midlife mean
	Betti (6)	H_0 Curve (Mean, Std, Max), H_1 Curve (Mean, Std, Max)
Total		**40 Features**

6 Experimental Results

6.1 Metric Evaluation

Tables 4, 5, 6 and 7 report the forecasting results for all evaluated models, feature sets, and assets using metric Mean Squared Error (MSE). Results are reported as mean ± standard deviation over 10 runs. The best results for each metric are highlighted with bold, second best results are highlighted with underline. All values are rounded to two decimal places using standard rounding rules. In addition to MSE, we evaluated Mean Absolute Error (MAE) and Directional Accuracy (DA). We omit detailed MAE results as they strongly correlate with MSE, yielding identical model rankings. DA is also excluded, as values fluctuated near 0.50 across all experiments. Welch's t-tests confirmed no statistically significant improvement ($p > 0.05$) in directional prediction over the baseline, even in cases where MSE significantly decreased.

LightGBM's weak performance was caused by shift in price levels between the training and test sets. Since tree based models cannot extrapolate beyond the range of training data, the model failed to predict the higher prices in the test set. This led to systematic underestimation and high error metrics.

6.2 Statistical Significance Analysis

To evaluate the impact of different feature sets, we conducted Welch's t-test (unequal variances) comparing the Baseline against the augmented sets. This test was chosen due to the substantial heteroscedasticity observed across the $n = 10$ random seed runs. The introduction of Technical Indicators often reduced prediction variance compared to the baseline. We tested the null hypothesis (H_0) of equal means at a significance level of $\alpha = 0.05$.

Table 6 shows that adding Technical Indicators (B+I) yields statistically significant improvements across the majority of models. For state-of-the-art architectures like PatchTST and iTransformer, MSE values decreased substantially (e.g., for AAPL, PatchTST improved from 33.71±1.88 to 19.25±0.29, $p < 0.001$). Furthermore, the lower standard deviations in the B+I experiments indicate that technical indicators contribute to the stability and robustness of the training process.

Table 4. Results Metric MSE Features B (mean ± std).

Ticker	DLINEAR	TIMESNET	PATCHTST	ITRANSFORMER	LIGHTGBM
AAPL	**31.61 ± 1.21**	43.67 ± 4.65	33.71 ± 1.88	38.01 ± 2.37	4117.74 ± 110.83
AMT	26.73 ± 1.91	30.72 ± 7.22	27.88 ± 8.45	21.80 ± 2.76	**15.93 ± 1.17**
AMZN	32.29 ± 0.67	39.80 ± 8.99	33.34 ± 3.48	**30.38 ± 1.18**	1448.82 ± 16.89
APD	**40.78 ± 1.36**	50.86 ± 8.23	42.58 ± 1.75	42.10 ± 2.11	202.64 ± 9.34
BA	27.87 ± 2.07	37.89 ± 8.86	29.57 ± 2.87	33.26 ± 4.48	**20.57 ± 0.28**
GOOGL	48.64 ± 5.10	53.58 ± 5.81	**29.96 ± 2.76**	34.47 ± 6.03	4968.67 ± 26.26
JNJ	15.09 ± 3.80	9.37 ± 2.10	**7.58 ± 0.39**	9.15 ± 1.85	260.47 ± 1.14
JPM	**24.73 ± 0.88**	35.57 ± 5.02	24.76 ± 3.30	28.55 ± 1.72	12890.55 ± 16.08
NEE	3.65 ± 0.52	3.85 ± 0.44	**3.15 ± 0.16**	3.28 ± 0.23	5.58 ± 0.08
PG	**4.86 ± 0.17**	6.56 ± 0.61	5.28 ± 0.41	6.32 ± 0.61	262.56 ± 0.87
SPY	**56.88 ± 4.05**	71.17 ± 2.70	65.08 ± 7.84	67.77 ± 1.09	26729.28 ± 190.36
XOM	3.83 ± 0.13	4.77 ± 0.30	3.84 ± 0.36	**3.79 ± 0.26**	123.16 ± 1.62

Table 5. Results Metric MSE Features B+PH (mean ± std).

Ticker	DLINEAR	TIMESNET	PATCHTST	ITRANSFORMER	LIGHTGBM
AAPL	47.81 ± 2.30	77.59 ± 15.37	**35.22 ± 2.50**	53.13 ± 4.41	4195.17 ± 28.59
AMT	26.82 ± 2.05	117.92 ± 47.72	**19.47 ± 2.20**	24.76 ± 2.01	20.97 ± 0.94
AMZN	93.37 ± 12.98	219.46 ± 10.47	**36.50 ± 2.31**	46.36 ± 0.83	1709.60 ± 25.84
APD	73.82 ± 11.06	214.49 ± 93.04	**38.84 ± 1.90**	56.90 ± 1.89	223.64 ± 6.20
BA	53.19 ± 5.05	269.71 ± 31.74	39.47 ± 8.04	48.96 ± 1.04	**20.79 ± 0.62**
GOOGL	55.55 ± 5.57	511.76 ± 49.04	**37.35 ± 1.26**	44.10 ± 2.96	5288.62 ± 25.06
JNJ	20.86 ± 2.33	47.81 ± 26.27	**6.25 ± 0.69**	8.58 ± 1.32	269.13 ± 4.78
JPM	82.79 ± 11.83	129.05 ± 36.49	**27.39 ± 1.59**	37.95 ± 0.62	13720.72 ± 38.10
NEE	5.30 ± 0.42	6.32 ± 1.00	**2.42 ± 0.11**	3.63 ± 0.22	5.85 ± 0.21
PG	18.45 ± 4.65	23.88 ± 2.26	**5.65 ± 0.44**	7.04 ± 0.06	262.23 ± 3.18
SPY	239.59 ± 55.23	138.66 ± 22.63	**86.08 ± 8.63**	103.58 ± 3.26	28902.91 ± 121.98
XOM	13.92 ± 2.30	16.33 ± 2.69	**5.05 ± 0.66**	6.46 ± 0.50	101.64 ± 3.95

Persistent Homology features (Table 5) rarely provided a statistically significant benefit over the baseline. In many instances, MSE increased, suggesting that PH features introduce noise or complexity that the models fail to leverage effectively. Consequently, the null hypothesis (H_0) could not be rejected for the majority of tickers.

Finally, the combination of all features (Table 7) generally underperformed the B+I configuration. While often better than the Baseline or B+PH, the inclusion of PH appears to dampen the positive effects of Technical Indicators. This suggests that Technical Indicators are the primary driver of predictive accuracy, while the topological features, in this specific setup, do not possess sufficient predictive power.

Table 6. Results Metric MSE Features B+I (mean ± std).

Ticker	DLINEAR	TIMESNET	PATCHTST	ITRANSFORMER	LIGHTGBM
AAPL	27.04 ± 2.28	52.68 ± 25.82	**19.25 ± 0.29**	22.67 ± 1.36	4001.54 ± 46.76
AMT	**11.89 ± 0.61**	28.99 ± 5.03	12.73 ± 0.89	13.55 ± 1.02	14.63 ± 0.36
AMZN	22.24 ± 1.48	35.37 ± 6.02	**20.38 ± 0.27**	22.89 ± 0.46	1681.16 ± 7.48
APD	34.46 ± 3.99	64.47 ± 12.38	**25.92 ± 1.31**	29.12 ± 1.15	191.83 ± 5.50
BA	21.76 ± 1.68	32.44 ± 3.35	**19.15 ± 0.85**	21.37 ± 0.67	22.15 ± 0.78
GOOGL	23.51 ± 3.81	55.11 ± 20.55	**19.27 ± 2.42**	26.17 ± 0.78	5133.74 ± 11.98
JNJ	10.61 ± 3.75	11.23 ± 2.59	5.15 ± 1.97	**4.67 ± 0.33**	265.04 ± 1.50
JPM	20.32 ± 3.82	37.50 ± 6.48	**19.24 ± 1.03**	22.18 ± 1.19	13519.71 ± 17.02
NEE	1.86 ± 0.26	4.74 ± 0.94	**1.82 ± 0.07**	1.90 ± 0.05	5.15 ± 0.20
PG	**3.53 ± 0.37**	7.01 ± 1.43	3.55 ± 0.20	3.85 ± 0.23	263.18 ± 1.81
SPY	56.56 ± 19.16	115.63 ± 55.95	**44.70 ± 2.13**	49.62 ± 0.56	27600.87 ± 112.78
XOM	**2.70 ± 0.10**	4.45 ± 0.73	2.76 ± 0.02	3.02 ± 0.21	88.57 ± 0.69

Table 7. Results Metric MSE Features B+I+PH (mean ± std).

Ticker	DLINEAR	TIMESNET	PATCHTST	ITRANSFORMER	LIGHTGBM
AAPL	46.04 ± 8.02	93.92 ± 8.99	**26.48 ± 1.74**	40.25 ± 0.81	4212.71 ± 15.84
AMT	22.21 ± 1.39	74.13 ± 12.62	**13.38 ± 0.91**	19.65 ± 0.27	21.55 ± 1.54
AMZN	91.91 ± 16.39	78.73 ± 8.29	**25.71 ± 1.15**	37.63 ± 2.35	1762.25 ± 33.23
APD	58.34 ± 7.28	108.36 ± 33.48	**31.65 ± 1.58**	45.63 ± 3.64	204.86 ± 3.17
BA	53.39 ± 5.14	87.04 ± 13.40	26.64 ± 2.91	41.04 ± 3.47	**20.76 ± 0.15**
GOOGL	47.92 ± 3.07	199.10 ± 73.00	**27.72 ± 4.28**	36.57 ± 3.44	5413.34 ± 19.13
JNJ	20.28 ± 4.26	23.62 ± 5.78	**5.27 ± 0.44**	6.32 ± 0.09	269.96 ± 1.60
JPM	72.75 ± 16.48	85.07 ± 37.47	**24.39 ± 1.32**	31.71 ± 0.10	13945.35 ± 69.79
NEE	4.09 ± 0.39	5.47 ± 0.74	**1.90 ± 0.03**	2.92 ± 0.17	6.14 ± 0.38
PG	16.48 ± 6.34	12.46 ± 2.33	**4.50 ± 0.39**	5.79 ± 0.20	265.46 ± 1.12
SPY	227.24 ± 42.00	164.27 ± 36.65	**60.23 ± 7.25**	83.79 ± 1.91	29135.06 ± 87.17
XOM	11.93 ± 4.17	10.94 ± 0.88	**3.75 ± 0.19**	5.47 ± 0.29	113.97 ± 3.02

7 Conclusion

In this study, we evaluated the effectiveness of integrating features derived from Topological Data Analysis (specifically Persistent Homology) into deep learning models for financial time series forecasting. We compared these features against standard Technical Indicators across a diverse set of U.S. stocks.

Our experimental results show that Technical Indicators (B+I) provided the most significant and consistent improvement in forecasting accuracy (MSE), whereas Persistent Homology features (B+PH) often failed to improve, or even degraded, model performance.

The main advantages of this approach are its independence from forecasting architecture and its ability to capture global structural properties of time series. In theory, persistent topological descriptors may also offer greater robustness to noise. Its main disadvantages are the large number of methodological choices, hyperparameters and the computational complexity of the PH pipeline. The final performance may depend on several methodological choices, such as point cloud construction, embedding parameters, filtration, and vectorization, for which no universally optimal configuration is known.

We attribute this disparity to the fundamental difference in how these features interact with market dynamics. The superior performance of Technical Indicators can be largely explained by their widespread adoption. These metrics are monitored by a vast number of human traders and algorithmic trading systems. In this sense, Technical Indicators can act as a self-fulfilling prophecy. They may not just describe the market, but actively shape it.

In contrast, Persistent Homology extracts abstract geometric structures from the data embedding. While mathematically rigorous, these features lack a direct link to market behavior. Since market participants do not make trading decisions based on topological features, these features do not drive the markets. Consequently, when added to the models, they likely acted as high dimensional noise rather than actionable signal, obscuring the clear patterns provided by price history and Technical Indicators.

These findings do not imply that Persistent Homology is unsuitable for financial time series prediction. Our experiments confirm that providing topological features to deep learning models is not a straightforward path to superior performance. However, our study was limited to specific embedding strategies and fixed hyperparameters. We did not explore alternative point cloud construction methods, filtration or vectorization techniques. Furthermore, the utility of Persistent Homology extends beyond direct feature extraction. Future research should investigate its potential in other areas, such as regime change detection, time series clustering, or pattern recognition, where its ability to capture global structural properties may prove more advantageous.

Acknowledgments. This research was supported by the Slovak Research and Development Agency (APVV) under Grant No. APVV-24-0446.

References

1. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discrete Comput. Geom. **37**(1), 103–120 (2007)
2. Heo, E., Jung, J.H.: Persistent homology of featured time series data and its applications. AIMS Math. (2024)
3. de Jesus Jr., L.C., Fernández-Navarro, F., Carbonero-Ruz, M.: Enhancing financial time series forecasting through topological data analysis. Neural Comput. Appl. (2025)
4. Ke, G., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems (2017)
5. Kim, N., Baik, H., Yoon, Y.: TopoCL: topological contrastive learning for time series. arXiv preprint arXiv:2502.02924 (2025)
6. Lin, Z., Zulkepli, N.F.S., Kasihmuddin, M.S.M., Gobithaasan, R.U.: Time-series forecasting via topological information supervised framework with efficient topological feature learning. Expert Syst. Appl. (2026)
7. Liu, Y., et al.: iTransformer: inverted transformers are effective for time series forecasting. In: International Conference on Learning Representations (2024)
8. MSCI Inc.: The global industry classification standard. https://www.msci.com/indexes/index-resources/gics
9. Nguyen, V.T., Pham, D.A., Le, A.T., Peter, J., Gust, G.: Persistent homology-induced graph ensembles for time series regressions. arXiv preprint arXiv:2503.14240 (2025)
10. Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: long-term forecasting with transformers. In: International Conference on Learning Representations (2023)
11. TA-Lib contributors: Ta-lib: Technical analysis library for python. https://github.com/TA-Lib/ta-lib-python
12. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: TimesNet: temporal 2D-variation modeling for general time series analysis. In: International Conference on Learning Representations (2023)
13. Yahoo Inc.: Yahoo finance. https://finance.yahoo.com/
14. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: AAAI Conference on Artificial Intelligence (2023)
15. Zeng, S., Graf, F., Hofer, C., Kwitt, R.: Topological attention for time series forecasting. In: Advances in Neural Information Processing Systems, vol. 34, pp. 24871–24882 (2021)
16. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete Comput. Geom. **33**(2), 249–274 (2005)

Portfolio Optimization with TD3 Algorithm Using Technical Indicators

Teodora Šimon(✉), Peter Likavec, Maros Cavojsky, and Martin Drozda

Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Bratislava, Slovakia

teodora.simon@stuba.sk, xlikavec@stuba.sk, maros.cavojsky@stuba.sk, martin.drozda@stuba.sk

Abstract. We explore the application of the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm to portfolio optimization. Financial markets present high-dimensional and non-stationary environments where traditional reinforcement learning (RL) methods often suffer from value overestimation and instability. Unlike other approaches that use OHLCV (Open, High, Low, Close, Volume) data, we evaluate our approach using only a diverse set of normalized technical indicators. Such indicators are derived from OHLCV data; however, they are widely used by investors who rely on technical analysis. Our experimental results demonstrate that the agent achieves stable convergence. We further analyze the policy's behavior using the HerfindahlHirschman Index (HHI) and portfolio turnover, revealing a shift toward more concentrated capital allocation during the inference phase.

Keywords: Algorithmic Trading · Deep Reinforcement Learning · Portfolio Optimization · TD3

1 Introduction

Algorithmic trading and portfolio optimization have undergone significant transformation with the advent of Deep Reinforcement Learning (DRL) [29]. Traditional portfolio management often relies on mean-variance optimization, which typically requires accurate estimation of expected returns and covariance matrices, tasks that are notoriously difficult in volatile markets. DRL offers a paradigm shift by enabling agents to learn optimal trading policies directly from high-dimensional market data through interaction and feedback.

However, applying DRL to finance is not without challenges. Actor-critic methods, such as Deep Deterministic Policy Gradient (DDPG) [15], are prone to overestimation bias and high variance, which can lead to suboptimal policies and unstable training. To address these issues, we employ the TD3 algorithm [12], which introduces clipped double Q-learning, delayed policy updates, and target policy smoothing. This research focuses on constructing a robust DRL framework

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 662–675, 2027.
https://doi.org/10.1007/978-3-032-32104-6_55

for portfolio selection that specifically accounts for transaction costs and investor risk aversion. Unlike other approaches that use raw OHLCV (Open, High, Low, Close, Volume) data, we evaluate our approach using a diverse set of normalized technical indicators. Such indicators are derived from OHLCV data and are widely used by investors who rely on technical analysis.

2 Related Work

Value-based reinforcement learning methods first estimate a value function and then derive a policy from it. Thus, the quality of the policy depends entirely on the accuracy of the estimated action-value function $Q(s, a)$. In high dimensional state or action spaces, tabular representations of Q-values become infeasible, so the function approximation with neural networks is used. This introduces approximation error stemming from network architecture and training dynamics. Such error appears as noise in the Q-value estimates. When actions are selected via a maximization step, $\max_a Q(s, a)$, this noise induces a positive bias, causing systematic overestimation. Thrun and Schwartz [30] first identified this effect, and Hasselt [11] later formalized it in the context of Double Q-learning (Double QL). As a result, Q-value overestimation in value-based methods can lead to suboptimal policies and unstable or even divergent training, especially in high-dimensional environments.

Prior work has examined bias-variance challenges in reinforcement learning. Pendrith [22] showed that RL algorithms lack adaptive mechanisms to control estimator variance, leading to instability in noisy or non-Markov environments. They introduced *ccBeta*, an adaptive per-parameter learning-rate method based on online error autocorrelation, validated through theory, simulation, and real-robot experiments.

Hasselt [11] analyzed how uncertainty in estimated Markov decision process (MDP) transitions and rewards propagates through the Bellman equation. Using second-order approximations, they derived closed-form bias and variance expressions for value estimates and showed that variance typically dominates, enabling practical confidence intervals and statistically sound policy evaluation.

Fujimoto et al. [9] studied function-approximation error in actorcritic methods, proposing an improved Deep Deterministic Policy Gradient (DDPG) [15]. While Double DQN [31] and Double Q-learning [11] mitigate overestimation bias, Double Q-learning can introduce high variance that still harms policy performance. To address this, they proposed Clipped Double Q-learning, which uses the minimum of two independently learned Q-value estimates to prevent optimistic errors from propagating.

Additional work explored variance-reduction strategies in Q-learning, including averaging past Q-networks [3] and soft-update mechanisms to avoid overfitting noisy early estimates [8]. Temporal-difference methods also suffer from compounding errors due to bootstrapping on imperfect estimates, with mitigation typically achieved through multi-step returns or hybrid Monte-Carlo/TD updates.

Portfolio construction can be optimized using Machine Learning (ML), RL, and Deep RL (DRL). ML supports forecasting, asset pre-selection, and decision-making. Rubesam [26] applies ML-based portfolio construction and compares it with traditional methods; Mavruk [16] uses ML to detect investor herding; and Fereydooni and Mahootchi [7] employ ML for algorithmic trading in currency markets.

RL is well-suited to dynamic financial markets due to its sequential decision-making framework. RL methods are grounded in the MDP [1,21], with algorithms such as QL, Deep QL (DQL), DDPG, and PPO representing common extensions. DDPG is particularly important for continuous-action portfolio optimization [15]. High-dimensional portfolio DQL models have been explored by Pigorsch and Schäfer [24], while multi-agent DQL approaches were developed by Park et al. [20] and Shavandia and Khedmati [27].

DRL has become a strong approach for portfolio optimization, combining deep learning's representation power with RL's sequential decision-making. This enables strategies that adapt to dynamic, stochastic, and high-dimensional markets. Bühler et al. [5] introduced Deep Hedging, a DRL method incorporating risk constraints and transaction costs, while Li [14] proposed a DRL framework for online portfolio selection that explicitly models transaction costs. Yang [33] proposed an ensemble DRL method for automated trading suited to high-dimensional markets. Zhao [35] introduced an asset-correlation-based DRL approach that improves portfolio selection by leveraging cross-asset relationships. Cui [6] incorporated CVaR-based risk and investor risk aversion into a DRL portfolio-construction framework. Betancourt and Chen [4] developed a DRL model capable of handling markets with a dynamic number of tradable assets, a key requirement for transaction-cost-aware trading.

Earlier DRL portfolio methods often ignored transaction costs or lacked explicit risk-aversion modeling. These limitations were addressed by integrating meanvariance constraints into DRL [10,32], developing cost-sensitive DRL portfolio models [34], and incorporating asset variability and correlations through repeated DRL decision processes [18].

3 Methodology

This section outlines the core principles of the *TD3 algorithm* applied in *Deep reinforcement learning for portfolio selection* [12].

In *value-based* RL, function approximation errors often lead to overestimated value functions and suboptimal policies. These issues persist in an *actor-critic* setting. Fujimoto et al. [9] address this problem with TD3, which builds on the Deterministic Policy Gradient (DPG) algorithm [28], an actor-critic approach that trains a deterministic policy using learned value estimates. When a policy is updated via gradient ascent on an approximate value function, small update steps locally maximize the estimated value, but not necessarily the true value. If the critic is systematically optimistic, the policy update will favor actions with positive estimation error, causing the policy's true performance to be overestimated. This explains why overestimation bias can arise in actor-critic methods

even without explicit action maximization. To reduce overestimation bias in actor-critic methods, the TD3 authors introduced a clipped variant of Double QL. In Double QL, action selection (the greedy step) and action evaluation (target computation) are separated across two value estimators to prevent shared error accumulation. The Double DQN idea of separating selection (online network) from evaluation (target network) is specific to value-based DQN, while the clipped variant adapts this separation to actor-critic settings.

Target update using the clipped variant of Double Q-learning is defined as:

$$y_1 = r + \gamma \min_{i=1,2} Q_{\theta'_i}\big(s', \pi_{\phi_1}(s')\big), \tag{1}$$

where y_1 denotes the target value used to update critic $Q_{\theta'_1}$, r is the reward observed after taking action a from state s, γ with $0 \leq \gamma < 1$ is the discount factor, $\min_{i=1,2}(\cdot)$ is the clipped Double Q operator applied over the two target critics, $Q_{\theta'_i}(\cdot)$ are the two target critic networks with parameters θ'_i, and $\pi_{\phi_1}(\cdot)$ is the actor network with parameters ϕ_1, used to select the next action at state s' when computing the target for critic 1.

Another key issue is variance. High-variance value estimates produce noisy policy-gradient updates, destabilizing learning; variance not only contributes to overestimation but also harms overall training stability. TD learning amplifies this problem because each update bootstraps on an approximate next-state value, leaving a residual TD error that accumulates over time. To improve stability, target networks are commonly used. Since deep function approximators require many gradient steps to converge, slowly updated target networks provide a more stable learning objective and make better use of experience data. The weights of the target network are slowly updated by τ at each time step using $\theta' \leftarrow \tau\theta + (1-\tau)\theta'$, where τ is the soft-update rate that determines how quickly the target network's weights move toward the online network's weights.

A limitation of deterministic policies is their tendency to overfit to sharp peaks in the value function. Because the critic's targets depend on a single deterministic action, they become highly sensitive to function-approximation errors, increasing variance. To mitigate this, the authors introduce target policy smoothing, a regularization strategy inspired by SARSA [29]. Instead of using one deterministic action to compute the critic's target, small clipped noise is added to the target action, effectively averaging value estimates over nearby actions. This enforces the idea that similar actions should yield similar values, producing smoother value functions and more stable policy updates.

The target update with target policy smoothing is defined as follows:

$$y = r + \gamma\, Q_{\theta'}\big(s',\, \pi_{\phi'}(s') + \varepsilon\big), \qquad \varepsilon \sim \mathrm{clip}\big(\mathcal{N}(0, \tilde{\sigma}),\, -c,\, c\big), \tag{2}$$

where y denotes the smoothed target used to update the critic; r is the reward obtained after taking action a from state s; γ, with $0 \leq \gamma < 1$, is the discount factor; $Q_{\theta'}(\cdot)$ is the target critic network parameterized by θ'; $\pi_{\phi'}(s')$ is the target actor generating the next-state action, parameterized by ϕ'; ε is the target policy smoothing noise drawn from a Gaussian distribution $\mathcal{N}(0, \tilde{\sigma})$ and clipped

elementwise to the interval $[-c, c]$; c controls the maximum magnitude of the smoothing perturbation; and $\tilde{\sigma}$ denotes the standard deviation of the smoothing noise prior to clipping. Note that a single critic $Q_{\theta'}$ is used here for clarity of exposition; in the full TD3 update, the clipped Double Q operator $\min_{i=1,2} Q_{\theta'_i}(\cdot)$ is retained alongside the smoothing noise, yielding the combined target defined in Eq. 1.

The TD3 algorithm is depicted in Fig. 1, where $\tilde{a}$ denotes the noisy target action used for target policy smoothing; ε denotes the target-smoothing noise; $\tilde{\sigma}$ is the standard deviation of this noise; d is the policy-delay parameter; τ is the target-update rate; and N denotes the mini-batch size.

Algorithm 1. Twin Delayed Deep Deterministic Policy Gradient (TD3)

Initialize critic networks $Q_{\theta_1}, Q_{\theta_2}$, and actor network π_ϕ with random parameters θ_1, θ_2, ϕ
Initialize target networks $\theta'_1 \leftarrow \theta_1$, $\theta'_2 \leftarrow \theta_2$, $\phi' \leftarrow \phi$
Initialize replay buffer $\mathcal{B}$
for $t = 1$ to T **do**
  Select action with exploration noise $a \sim \pi_\phi(s) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma)$
  Observe reward r and next state s'
  Store transition tuple (s, a, r, s') in $\mathcal{B}$
  Sample mini-batch of N transitions (s, a, r, s') from $\mathcal{B}$
  $\tilde{a} \leftarrow \pi_{\phi'}(s') + \varepsilon, \quad \varepsilon \sim \mathrm{clip}(\mathcal{N}(0, \tilde{\sigma}), -c, c)$
  $y \leftarrow r + \gamma \min_{i=1,2} Q_{\theta'_i}(s', \tilde{a})$
  Update critics:

$$\theta_i \leftarrow \arg\min_{\theta_i} \frac{1}{N} \sum (y - Q_{\theta_i}(s, a))^2$$

  if $t \bmod d = 0$ **then**
    Update ϕ using deterministic policy gradient:

$$\nabla_\phi J(\phi) = \frac{1}{N} \sum \nabla_a Q_{\theta_1}(s, a)\big|_{a=\pi_\phi(s)} \nabla_\phi \pi_\phi(s)$$

    Update target networks:

$$\theta'_i \leftarrow \tau\theta_i + (1-\tau)\theta'_i, \qquad \phi' \leftarrow \tau\phi + (1-\tau)\phi'$$

  end if
end for

In our work, we apply a model-free DRL framework introduced by Jiang [12], within which TD3 is applied as the underlying DRL algorithm. The portfolio value at time t is given by:

$$P_t = P_{t-1} + p_t^\intercal k_t, \tag{3}$$

where P_{t-1} denotes the portfolio value from the previous period, $p_t = [p_{1,t}, \ldots, p_{N,t}]$ represents the vector of closing prices of all N assets at time t, and $k_t =$

$[k_{1,t}, \ldots, k_{N,t}]$ is the vector of trading actions at time t, where each $k_{n,t}$ specifies whether asset n is bought, held, or sold.

In the portfolio selection problem, transaction costs must be incorporated into every trading operation. Following [33], the transaction cost at time t is defined as:

$$c_t^{\text{tran}} = \xi \times p_t^{\intercal} k_t, \tag{4}$$

where ξ is a positive constant representing the transaction cost rate associated with executing a new trade. The term $p_t^{\intercal}$ denotes the transpose of the asset price vector p_t, and $k_t = [k_{1,t}, ..., k_{N,t}]$ is the portfolio decision vector - also referred to as the trading action.

Beyond transaction costs, the objective function also incorporates underlying portfolio risk, captured by the portfolio variance σ_t^2, together with the investor's risk aversion coefficient β. The risk penalty is introduced through the cost term:

$$c_t^{\text{risk}} = \beta \sigma_t^2. \tag{5}$$

The portfolio trading decision is formulated as a MDP, represented by the tuple $(\mathcal{S},\mathcal{A},\mathcal{P},r)$, where $\mathcal{S}$ denotes the state space, defined by a vector of state variables $s_t = [p_t, m_{t-1}]$, $\mathcal{A}$ is the action space containing all feasible actions $a_t \in \{buy, hold, sell\}$, $\mathcal{P}$ denotes the state transition probability, r the reward function and m_t is the vector of asset holdings (number of shares) at time t.

The trading strategy is characterized by a portfolio policy $\pi(a_t, s_t)$, and the reward function $r_t(s_t, a_t, s_{t+1})$ is obtained by applying this policy once the next state s_{t+1} is realized. The objective reward function incorporates both risk and transaction costs:

$$r_t = p_t^{\intercal} m_t - \beta \sigma^2 t - \xi p_t^{\intercal} |k_t|, \tag{6}$$

where $p_t^{\intercal} m_t$ represents the total market value of the portfolio. The term $\beta \sigma^2 t$ denotes the volatility penalty reflecting the investor's risk aversion, while $\xi p t^{\intercal} |k_t|$ accounts for the transaction costs associated with executing trades.

The agent aims to learn an optimal trading policy $\pi(s_t, a_t)$ that selects actions maximizing the reward. Policy performance is evaluated through the action-value function $Q^{\pi}(s_t, a_t)$, or Q-function, which can be expressed via the Bellman equation as described by Wiering and Otterlo [19]:

$$Q^{\pi}(s_t, a_t) = E_{s_{t+1}} \left[r_t(s_t, a_t, s_{t+1}) + \gamma \, E_{a_{t+1}} \left[Q^{\pi}(s_{t+1}, a_{t+1}) \right] \right]. \tag{7}$$

Q-learning is a widely used RL method that updates the action-value function $Q^{\pi}(s_t, a_t)$ recursively to maximize cumulative rewards and derive the optimal trading policy. The temporal-difference (TD) error quantifies the gap between predicted and actual Q-values as the agent transitions between states, and the Q-value is updated by scaling this error with the learning rate [19]:

$$Q_\pi^{\text{New}}(s_t, a_t) \leftarrow Q_\pi^{\text{Old}}(s_t, a_t) + \alpha\Big(r_t(s_t, a_t, s_{t+1}) + \gamma \max_{a_{t+1} \in \mathcal{A}} Q_\pi(s_{t+1}, a_{t+1}) - Q_\pi^{\text{Old}}(s_t, a_t)\Big). \tag{8}$$

4 Experimental Setup

We utilize historical data sourced from the Alpaca API Platform's Market Data API [2]. Our analysis focuses on the period from 2016 to 2023, using 30-minute interval data. Data from 2016 to 2019 are used for training, and data from 2019 to 2023 for testing.

Our experiments were conducted on the following ten tickers: AAPL, MSFT, AMZN, GOOGL, META, TSLA, NVDA, JPM, JNJ, and XOM. While our experiments successfully demonstrate the TD3 algorithm's ability to learn robust trading policies, we restrict the agent's action space exclusively to long positions and cash allocations

We acknowledge that the exclusion of short selling represents a notable limitation, as it restricts the strategy's capacity to generate alpha during persistent market downturns or individual asset depreciation. With a long-only restriction, the agent can defend against anticipated price drops only by exiting its positions, as it cannot capitalize on negative market movements. Consequently, because the portfolio can only profit from rising prices, it might be vulnerable to broad market crashes. This lack of built-in protection may lead to steeper financial losses compared to strategies that use short selling to hedge against market risk. Furthermore, truncating the action weights to the $[0, 1]$ interval limits the RL agent's expressiveness, preventing it from fully capitalizing on high-conviction bearish signals identified within the technical indicator state space.

Nevertheless, this long-only restriction is a deliberate design choice intended to manage environmental complexity. Incorporating short selling into a simulated market environment requires modeling complex real-world frictions such as dynamic borrowing rates, margin maintenance requirements, and the theoretical infinite risk associated with unbounded price increases. By isolating the action space to long positions, we eliminate these confounding variables, allowing us to establish a controlled baseline that cleanly evaluates the core efficacy of the TD3 algorithm and the predictive utility of the selected technical indicators.

4.1 Indicators

Unlike Jiang [12], who relies primarily on raw OHLC data, our model utilizes normalized technical indicators as inputs. To ensure comparability between heterogeneous features and to mitigate the influence of absolute value magnitudes, all indicators were normalized prior to training. Indicators with a natural central reference point were centered around their neutral value. For example, the RSI,

defined on the interval $[0, 100]$, was shifted by 50, yielding a transformed range of $[-50, 50]$. The ADX was treated analogously but centered at its conventional neutrality threshold of 25.0. Furthermore, indicators derived from moving averages (such as EMA and SMA) were scaled relative to the current closing price. For any moving-average-based indicator I_t, the normalized value was computed as:

$$I_t^{(\text{norm})} = \frac{I_t}{\text{Close}_t}.$$

This transformation expresses the relative deviation of the indicator from the prevailing price and is invariant to the absolute price level of the asset, thereby significantly improving the cross-asset comparability of the input features.

In total, our model incorporates 17 technical indicators to enrich the state space and enhance decision-making. These include the RSI and Rate of Change (ROC) with a 20-day window; MACD with parameters $(12, 26, 9)$; Simple Moving Average (SMA) and Exponential Moving Average (EMA) with periods of 20, 50, 100, and 200; Bollinger Bands formed by a 20-day SMA and a standard deviation multiplier of 2.0; Keltner Channels with a 20-day window, a 10-day ATR window, and a multiplier of 2.0; and the TTM Squeeze utilizing a 20-day window, a 20-day ATR window, and a multiplier of 1.5. Additional indicators comprise the TSI (slow 25, fast 13), MFI, VWAP, ATR, ADX, and CMF, each computed with a 14-day window, along with AROON using a 25-day window. Finally, the asset's closing price is also included as a baseline input feature.

4.2 Training and Testing Setup

The differences between the training and testing modes lie primarily in the settings that influence the agent's behavior. The core processing logic remains the same in both cases, with the differences summarized in Table 1.

Table 1. Differences Between Training and Testing Modes

Property	Training Mode	Testing Mode
Exploration (noise)	Exploration noise is used when selecting actions.	Noise is not used so it does not influence action selection
Agent learning	The TD3 parameters are updated.	Learning is disabled; parameters remain unchanged
Replay buffer	Transitions are stored and sampled for mini-batch learning.	Replay buffer is not used for learning, so storing samples is unnecessary
Agent initialization	The agent is initialized without prior experience.	The trained model is loaded

The number of training episodes is set to 50, the learning rate is 0.0003 and the batch size is 128. Both the actor and critics use the Adam optimizer [13], and

critics are trained using Mean Squared Error (MSE) loss. ReLU [17] is applied as an activation function in hidden layers, and the Sigmoid activation function is used in the output layer.

In order to provide a robust performance estimate, we use 5 independent training and testing runs (herein we refer to them as run0–run4).

4.3 Implementation Details

Our implementation uses Python 3.12 and PyTorch as a machine learning framework. Pandas and NumPy handle data processing and numerical computations. The experiments are run on two devices: MacBook Pro (M3 Pro, 36 GB RAM), and a PC with Ubuntu 22 LTS OS and GPU (CUDA): NVIDIA GeForce RTX 3090.

Both TD3 actor and critic network is a three-layer MLP that flattens the input, processes it through two ReLU-activated hidden layers, and outputs continuous actions in [0, 1] via a Sigmoid layer. The actor network uses entmax [23] to normalize portfolio output weights. The algorithm uses an initial noise set to 0.3, the final noise is 0.05, and the noise annealing period is 500, defining how long it takes to gradually reduce the exploration noise from its initial value to its final value. The replay buffer size is set to $100,000$ and the initial cash is $10,000$. For simplicity, we treat the transaction-cost constant as identical for all assets, but the model can be adapted to reflect heterogeneous transaction costs and liquidity conditions.

5 Experimental Results and Performance Analysis

The key challenge in designing portfolio optimization approaches using RL is their overall convergence. The results related to convergence are presented in Fig. 1. This figure shows how the cumulative reward evolved over 500 episodes, the upward trend indicates that the agent is successful in learning portfolio weights to achieve higher returns.

Figure 2 displays the weight distribution in the final training episode. Instead of allocating capital evenly, the agent adjusts weights based on market conditions, aiming to maximize the reward, while responding dynamically to market changes. We can see that there are time steps when the portfolio is more weighted towards cash and other time steps when the portfolio is weighted towards stock allocation.

As shown in Fig. 2, the training phase exhibits some stochastic noise. The explorationexploitation trade-off effects are reflected in frequent portfolio rebalancing. In contrast, Fig. 3 reveals a more stable policy. The “vertical striping” observed during training is replaced by coherent, contiguous blocks of capital allocation. This shift indicates that the agent has transitioned from a stochastic exploratory phase to a more deterministic exploitation of learned market patterns.

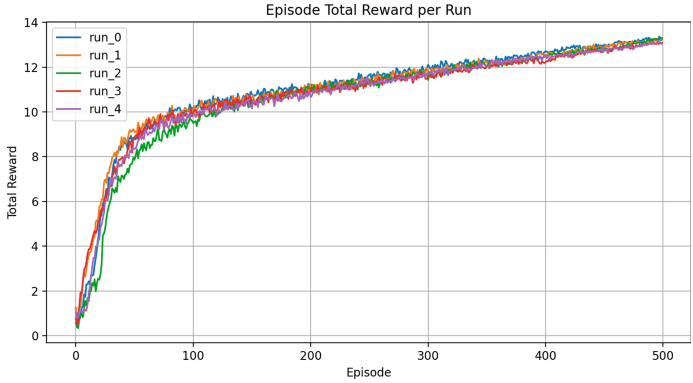


Fig. 1. The cumulative reward during training.

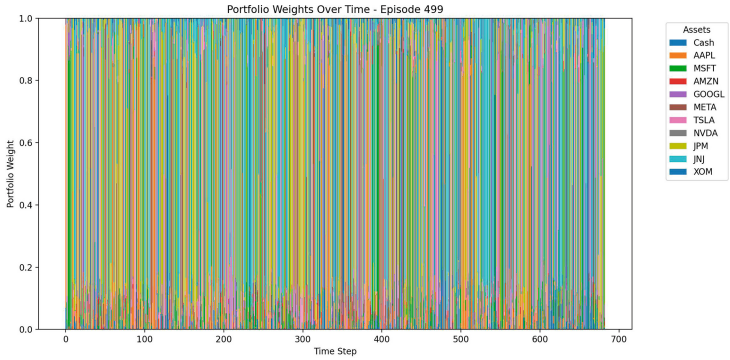


Fig. 2. Portfolio weights during training (run_0).

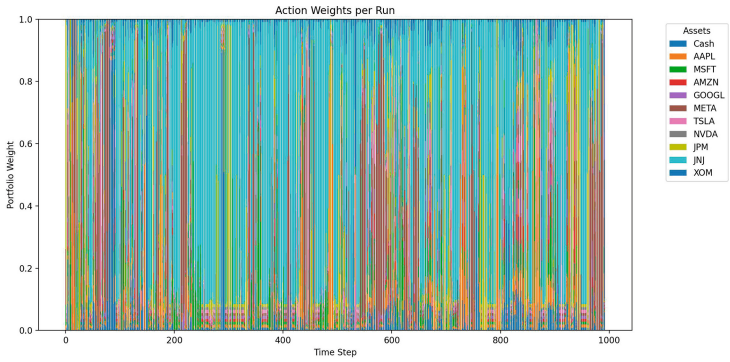


Fig. 3. Portfolio weights during testing (run_0).

To quantitatively assess the behavior of the TD3 agent, we apply a formal statistical comparison between the training phase and the deterministic test phase. We focus on two critical metrics: the Herfindahl-Hirschman Index (HHI) to measure asset concentration, and the Portfolio Turnover to measure temporal policy stability.

The concentration of weights is measured using the HHI, defined as:

$$HHI_t = \sum_{i=1}^{N} (w_{i,t})^2, \tag{9}$$

where $w_{i,t}$ represents the weight of asset i at time t.

The Portfolio Turnover, which quantifies the rebalancing frequency and the magnitude of change in the agent's actions, is defined as:

$$\mathcal{T} = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{i=1}^{N} |w_{i,t+1} - w_{i,t}|, \tag{10}$$

where T is the total number of time steps, N is the number of assets, and $w_{i,t}$ represents the weight of asset i at time t.

The test episode exhibited a mean HHI of 0.764 ($\sigma = 0.162$), representing a 15.58% increase over the training mean. This suggests that during inference, the agent converges on a"winner-take-all" strategy, significantly reducing portfolio diversification.

The results indicate a substantial degradation in stability during the test phase. The mean turnover rose from 1.040 during training to 1.381 in Fig. 2 (a 32.79% increase). This high turnover confirms that the deterministic policy is highly reactive to state-space noise, leading to excessive trading activity.

6 Conclusion

We have presented an application of the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm for the problem of continuous portfolio optimization. By utilizing normalized technical indicators as state inputs and incorporating transaction costs and risk aversion into the reward function, we developed an agent capable of navigating high-dimensional asset spaces.

Our experimental results demonstrate that the TD3 agent successfully achieves convergence, as evidenced by the upward trend in cumulative rewards during the training phase. A critical observation from our analysis is the profound qualitative shift in policy behavior between the training and testing phases. Once exploration noise is removed during inference, the agent defaults to a purely deterministic, greedy policy. This yields an unintended "winner-take-all" strategy, evidenced by a 15.58% increase in the Herfindahl-Hirschman Index (HHI), where the algorithm aggressively allocates maximum capital to whichever asset exhibits a marginally higher expected Q-value, completely abandoning the

risk-mitigating principles of portfolio diversification. Compounding this concentrated idiosyncratic risk is the strategy's severe temporal instability. Despite acting deterministically, portfolio turnover surged by 32.79% during testing. This erratic rebalancing reveals that the agent is hypersensitive to the inherent noise within technical indicator data; rather than capturing stable macroeconomic trends, the model continuously liquidates its concentrated positions in response to microscopic fluctuations in the state space.

Consequently, while the TD3 architecture proves capable of mathematical convergence, our future research will address how this excessive portfolio concentration can be mitigated, for instance, through action-space regularization or advanced smoothing techniques. We also intend to validate our enhanced models using QuantConnect [25], an event-based simulator, for robust testing of these trading strategies.

Acknowledgments. This research was supported by the Slovak Research and Development Agency (APVV) under Grant No. APVV-24-0446.

References

1. Almahdi, S., Yang, S.Y.: An adaptive portfolio trading system: a risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. Expert Syst. Appl. **87**, 267–279 (2017)
2. Alpaca Markets: Alpaca market data API documentation. https://docs.alpaca.markets/docs/about-market-data-api. Accessed 12 Feb 2026
3. Anschel, O., Baram, N., Shimkin, N.: Averaged-DQN: variance reduction and stabilization for deep reinforcement learning. In: International conference on machine learning, pp. 176–185. PMLR (2017)
4. Betancourt, C., Chen, W.H.: Deep reinforcement learning for portfolio management of markets with a dynamic number of assets. Expert Syst. Appl. **164**, 114002 (2021)
5. Buehler, H., Gonon, L., Teichmann, J., Wood, B.: Deep hedging. Quantit. Financ. **19**(8), 1271–1291 (2019)
6. Cui, T., Ding, S., Jin, H., Zhang, Y.: Portfolio constructions in cryptocurrency market: a CVaR-based deep reinforcement learning approach. Econ. Model. **119**, 106078 (2023)
7. Fereydooni, A., Mahootchi, M.: An algorithmic trading system based on a stacked generalization model and hidden Markov model in the foreign exchange market. Glob. Financ. J. **56**, 100825 (2023)
8. Fox, R., Pakman, A., Tishby, N.: Taming the noise in reinforcement learning via soft updates. arXiv preprint arXiv:1512.08562 (2015)
9. Fujimoto, S., Hoof, H., Meger, D.: Addressing function approximation error in actor-critic methods. In: International Conference on Machine Learning, pp. 1587–1596. PMLR (2018)
10. García-Galicia, M., Carsteanu, A.A., Clempner, J.B.: Continuous-time reinforcement learning approach for portfolio management with time penalization. Expert Syst. Appl. **129**, 27–36 (2019)
11. Hasselt, H.: Double Q-learning. Adv. Neural. Inf. Process. Syst. **23** (2010)

12. Jiang, Y., Olmo, J., Atwi, M.: Deep reinforcement learning for portfolio selection. Glob. Financ. J. **62**, 101016 (2024)
13. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Li, B., Wang, J., Huang, D., Hoi, S.C.: Transaction cost optimization for online portfolio selection. Quantit. Financ. **18**(8), 1411–1424 (2018)
15. Lillicrap, T.P.: Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971 (2015)
16. Mavruk, T.: Analysis of herding behavior in individual investor portfolios using machine learning algorithms. Res. Int. Bus. Financ. **62**, 101740 (2022)
17. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814 (2010)
18. Olschewski, S., Diao, L., Rieskamp, J.: Reinforcement learning about asset variability and correlation in repeated portfolio decisions. J. Behav. Exp. Financ. **32**, 100559 (2021)
19. Otterlo, M.V., Wiering, M.: Reinforcement learning: state-of-the-art. Adapt. Learn. Optim. **12** (2012)
20. Park, H., Sim, M.K., Choi, D.G.: An intelligent financial portfolio trading strategy using deep Q-learning. Expert Syst. Appl. **158**, 113573 (2020)
21. Peck, J., Yang, H.: Investment cycles, strategic delay, and self-reversing cascades. Int. Econ. Rev. **52**(1), 259–280 (2011)
22. Pendrith, M.D., Ryan, M.R., Sammut, C.: Estimator variance in reinforcement learning: Theoretical problems and practical solutions. University of New South Wales, School of Computer Science and Engineering (1997)
23. Peters, B., Niculae, V., Martins, A.F.T.: Sparse sequence-to-sequence models. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 1504–1519. Association for Computational Linguistics, Florence (2019). https://aclanthology.org/P19-1146
24. Pigorsch, U., Schäfer, S.: High-dimensional stock portfolio trading with deep reinforcement learning. In: 2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr), pp. 1–8. IEEE (2022)
25. QuantConnect (2025). https://www.quantconnect.com/
26. Rubesam, A.: Machine learning portfolios with equal risk contributions: evidence from the Brazilian market. Emerg. Mark. Rev. **51**, 100891 (2022)
27. Shavandi, A., Khedmati, M.: A multi-agent deep reinforcement learning framework for algorithmic trading in financial markets. Expert Syst. Appl. **208**, 118124 (2022)
28. Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., Riedmiller, M.: Deterministic policy gradient algorithms. In: International Conference on Machine Learning, pp. 387–395. PMLR (2014)
29. Sutton, R.S., Barto, A.G., et al.: Reinforcement Learning: An Introduction, vol. 1. MIT Press, Cambridge (1998)
30. Thrun, S., Schwartz, A.: Issues in using function approximation for reinforcement learning. In: Proceedings of the 1993 Connectionist Models Summer School, pp. 255–263. Psychology Press (2014)
31. Van Hasselt, H., Guez, A., Silver, D.: Deep reinforcement learning with double Q-learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30 (2016)
32. Wang, H., Zhou, X.Y.: Continuous-time mean-variance portfolio selection: a reinforcement learning framework. Math. Financ. **30**(4), 1273–1308 (2020)

33. Yang, H., Liu, X.Y., Zhong, S., Walid, A.: Deep reinforcement learning for automated stock trading: an ensemble strategy. In: Proceedings of the First ACM International Conference on AI in Finance, pp. 1–8 (2020)
34. Zhang, Y., Zhao, P., Wu, Q., Li, B., Huang, J., Tan, M.: Cost-sensitive portfolio selection via deep reinforcement learning. IEEE Trans. Knowl. Data Eng. **34**(1), 236–248 (2020)
35. Zhao, T., Ma, X., Li, X., Zhang, C.: Asset correlation based deep reinforcement learning for the portfolio selection. Expert Syst. Appl. **221**, 119707 (2023)

Author Index

A
Aberšek, Boris 27
Aberšek, Metka Kordigel 27
Ádám, Norbert 495
Adámek, Miroslav 362
Afonso, Luís 248
Agarski, Martina 129
Akimjaková, Beáta 211
Akimjakova, Beata 75

B
Bachmann, Pavel 285
Balogh, Michal 222
Balogh, Zoltán 495, 508, 523, 538, 571
Balogh, Zoltán 548
Belko, Jozef 187
Benko, Ľubomír 39
Benková, Lucia 39
Biernatova, Katerina 52
Bilek, Martin 63
Bilous, Iryna 75
Boatca, Maria Elena 300
Burgerová, Jana 85

C
Čech, Pavel 482
Cavojsky, Maros 587, 599, 611, 649, 662
Černá, Miloslava 153, 312
Coufal, Petr 98

D
Deusdado, Leonel 248
Diaconescu, Andra 300
Draghici, Anca 300
Drozda, Martin 587, 599, 611, 649, 662
Ďuriač, Denis 508

F
Főző, Ladislav 571
Flogie, Andrej 163
Fodor, Kristián 523
Fortuzi, Shkëlqim 495
Francisti, Jan 538

G
Gangur, Mikuláš 362
Gergely, Sebestyén 457
Gogh, Elod 110, 119
Gomotiuk, Oksana 322
Gomotiyk, Oksana 75, 469
Gorghiu, Gabriel 199

H
Hlavačka, Marián 538
Hlavsova, Hana 63
Hodon, Filip 587, 611
Homotiuk, Viktoriia 75
Hrabčák, Marek 548
Hubálovská, Marie 420

J
Jakab, Imrich 129, 337
Jakabová, Silvia 211
Jaroš, Miroslav 15
Jedina, Maria Anna 141

K
Kalous, Ladislav 153
Karmaz, Mykola 322
Kelebercová, Lívia 624, 635
Kerneža, Maja 163
Kerneža, Maja 27
Kilianova, Kristyna 175
Kockova, Petra 175
Kolesnyk, Oleksandr 469
Kollárová, Dana 129
Kopják, József 385, 457
Korba, Peter 571
Korenova, Lilla 52, 141
Kostolanyova, Katerina 175
Kováč, Ernest 187
Kováč, Jaroslav 187
Kovari, Attila 110, 119
Krabonja, Maja Vičič 163

© European Alliance for Innovation 2027
Z. Balogh (Ed.): DIVAI 2026, EAI/SICC, pp. 677–678, 2027.
https://doi.org/10.1007/978-3-032-32104-6

L
Lamanauskas, Vincentas 199
Lengyelfalusy, Tomáš 322, 469
Likavec, Peter 662

M
Machkova, Veronika 63
Magdin, Martin 495
Makáň, Filip 187
Makan, Filip 211, 322, 469
Marochok, Stanislav 222
Marták, Jakub 624, 635
Martinčík, David 235
Mikulecký, Peter 482
Milková, Eva 98
Mls, Karel 482
Molnár, György 272, 561
Moreiras, Mariana 248
Morze, Nataliia 349
Mucka, Oto 75
Munk, Michal 372, 624, 635
Munková, Daša 635

N
Németh, István 272
Nagy, Enikő 561
Nemec, Radek 260

O
Okechukwu, Chinyere Chidera 285
Otčenášková, Tereza 482

P
Palade, Marta 300
Pecuchová, Janka 39
Pekarčík, Ľubomír 187
Pekarcik, Lubomír 469
Piskura, Vladimír 85
Pitner, Tomáš 15
Ponce, Daniela 482
Poulová, Petra 312
Poulová, Petra 285
Pribeanu, Costin 199
Procházka, Václav 235
Proft, Radoslav 432
Pucherová, Zuzana 129

R
Rohlíková, Lucie 445
Rudenskyi, Rostyslav 211

S
Savitska, Viktoriia 322
Schmid, Angelika 141
Sebes, Patrik 649
Sebestyén, Gergely 272, 385
Shyshak, Andriana 211
Sik, Dávid 272
Šimon, Teodora 662
Skokanová, Katarína 337
Smyrnova-Trybulska, Eugenia 349
Sojková, Olga Martinčíková 235
Sojková, Olga Martinčíková 362
Stopjaková, Viera 3
Sudorová, Júlia 372
Szűcs, István 385

T
Tóbel, Imre 395
Tartaraj, Azeta 75, 538
Tlučhoř, Jan 235
Trenz, Oldřich 15
Tučník, Petr 482
Tvrdoň, Miroslav 187, 211

V
Vejvodová, Jana 445
Voštinár, Patrik 432
Voborník, Petr 405
Vopršal, Jan 420
Vozár, Martin 571

W
Weiss, Ervín 322
Wendler, Mark 457

Y
Yaroshchuk, Iryna 469

Z
Záhorec, Ján 372
Zilkova, Katarina 141
Žváčková, Andrea 482