2026



28th International Conference

ECONOMIC COMPETITIVENESS AND SUSTAINABILITY 2026

PROCEEDINGS

March 26th–27th, 2026, Brno

Editors:
Petr David
Hana Vránová

MENDELU
Faculty
of Business
and Economics

Mendel University in Brno
Faculty of Business and Economics

28th ANNUAL INTERNATIONAL CONFERENCE

ECONOMIC COMPETITIVENESS AND SUSTAINABILITY 2026

PROCEEDINGS

Editors:
Petr David, Hana Vránová

March 26th–27th, 2026
Mendel University in Brno
Czech Republic

MENDELU
Faculty
of Business
and Economics

2026

Organizer:
International scientific conference Economic Competitiveness and Sustainability 2026 was organized by the Faculty of Business and Economics, Mendel University in Brno.

Scientific Board:
Ing. Peter Albrecht, Ph.D. (Mendel University in Brno, Czech Republic)
Ing. Irena Baláková, Ph.D. (Mendel University in Brno, Czech Republic)
doc. Ing. František Dařena, Ph.D. (Mendel University in Brno, Czech Republic)
prof. Ing. Petr David, Ph.D. (Mendel University in Brno, Czech Republic)
prof. Jan Fidrmuc, PhD (University of Lille, France)
prof. Dr. Jarko Fidrmuc (Zeppelin University, Germany)
prof. RNDr. Jan Hanousek, CSc. (Mendel University in Brno, Czech Republic)
Dr. Florian Horky (Zeppelin University, Germany)
prof. Ing. Lubor Lacina, Ph.D. (Mendel University in Brno, Czech Republic)
prof. Karsten Neuhoff, Ph.D. (German Institute for Economic Research, Germany)
doc. Ing. Svatopluk Kapounek, Ph.D. (Mendel University in Brno, Czech Republic)
Mgr. Martin Lakomý, Ph.D. (Mendel University in Brno, Czech Republic)
Ing. Radek Náplava, Ph.D. (Mendel University in Brno, Czech Republic)
Mgr. Bc. Ondřej Pavelek, Ph.D. (Mendel University in Brno, Czech Republic)
doc. Ing. Bc. Mojmír Sabolovič, Ph.D. (Prague University of Economics and Business, Czech Republic)
prof. Mgr. Ing. Magdalena Daria Vaverková, Ph.D. (Warsaw University of Life Sciences, Poland)
prof. Steven J. Yamarik, PhD (California State University Long Beach, USA)

Organization Board:
prof. Ing. Petr David, Ph.D.
Ing. Hana Vránová
Mgr. Silvie Trávníčková
Mgr. Jana Jirušková
Ing. Andrea Prudilová
Ing. Jana Smítalová

Editors:
Petr David (https://orcid.org/ 0000-0002-5055-0989)
Hana Vránová (https://orcid.org/ 0000-0001-8254-1422)

© Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ISBN 978-80-7701-102-0 (online ; pdf)
https://doi.org/10.11118/ecos2026

Publication *Economic Competitiveness and Sustainability 2026 Proceedings*
is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International licese.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

ABSTRACT

On March 26th and 27th, 2026, the Faculty of Business and Economics of Mendel University in Brno organized already the 28th international conference "Economic Competitiveness and Sustainability" with over 100 onsite participants from around the world. In the presented Proceedings you find 44 papers which were recommended by conference discussants and selected on the basis of a peer-review process. The presented research outputs contribute to and extend the current state of the art and will stimulate further debate not only in academia but also in other institutions of public and private sector. The published papers address current interdisciplinary challenges arising at the intersection of digital transformation, artificial intelligence, sustainability, enterprise information systems, finance, public policy, marketing, and management. Together, they demonstrate the growing importance of data-driven approaches, intelligent technologies, and sustainable solutions in addressing contemporary economic and societal challenges.

Keywords: artificial intelligence, digital transformation, information systems, economics and finance, marketing and consumer behaviour, sustainability

ABSTRAKT

Ve dnech 26. a 27. března 2026 uspořádala Provozně ekonomická fakulta Mendelovy univerzity v Brně již 28. ročník mezinárodní konference „Economic Competitiveness and Sustainability", kterého se zúčastnilo přes 100 účastníků z celého světa. V předloženém sborníku najdete 44 příspěvků, které byly doporučeny diskutanty konference a vybrány na základě recenzního řízení. Prezentované výstupy výzkumu přispívají a rozšiřují současný stav poznání a podněcují další diskusi nejen na akademické půdě, ale i v rámci veřejného a soukromého sektoru. Publikované příspěvky reflektují současné interdisciplinární výzvy vznikající na průsečíku digitální transformace, umělé inteligence, udržitelnosti, podnikových informačních systémů, financí, veřejné správy a hospodářské politiky, marketingu a managementu. Současně poukazují na rostoucí význam datově podloženého rozhodování, inteligentních technologií a udržitelných řešení pro zvládání aktuálních ekonomických a společenských výzev.

Klíčová slova: umělá inteligence, digitální transformace, informační systémy, ekonomie a finance, marketing a spotřebitelské chování, udržitelnost

Sanela Arsić, Aleksandra Radić, Đorđe Nikolić
ERP SYSTEM SECURITY THROUGH THE EYES OF USERS: EVIDENCE FROM SERBIAN ENTERPRISES.....10

Sanela Arsić, Aleksandra Radić, Anđelka Stojanović
CIRCULAR ECONOMY INDICATORS AND DIGITAL TRANSFORMATION: ASSESSING THE RELATIONSHIP BETWEEN ERP ADOPTION AND CIRCULAR ECONOMY PERFORMANCE ACROSS EU COUNTRIES..... 23

Jaromír Běláček
ECONOMETRIC MODEL OF GENERALIZED LINEAR REGRESSION FOR PREDICTING FUTURE NUMBERS OF INSURED PEOPLE BY GENERAL HEALTH INSURANCE COMPANY OF THE CZECH REPUBLIC – COMPARISON OF RESULTS AFTER FIVE YEARS 39

Matěj Cajthaml, Tomáš Pitner
OPEN-SOURCE DECEPTION SECURITY FOR AWS CLOUD ENVIRONMENTS: A COST-CONSCIOUS ARCHITECTURE FOR SMES..... 58

Dávid Chomo, Jaromír Landa
AN OPEN ETL AND RAG PIPELINE FOR UNIFIED ACCESS TO HETEROGENEOUS CLIMATE DATA..... 67

Kateřina Dobrovolná, Tomáš Pitner
SOCIAL ENGINEERING AND HUMAN-FACTOR VULNERABILITIES IN CRITICAL INFRASTRUCTURE: FINDINGS FROM A SIMULATED PHISHING CAMPAIGN..... 76

Tereza Frömmelová, Kateřina Pojkarová
TRUST BETWEEN INDIVIDUALS: AN EMPIRICAL ANALYSIS INSPIRED BY COVEY AND SINEK 84

Shovan Nasim Haider, Miloslava Chovancova
FROM VALUES TO ACTION: EXPLAINING THE INTENTION-BEHAVIOR GAP IN GREEN FASHION CONSUMPTION AMONG GENERATION Z IN BANGLADESH..... 98

Marek Hebelka, Radim Farana
ROBUSTNESS OF A FUZZY CONTROLLER VERSUS CLASSICAL PID CONTROL ON TYPICAL PLANTS..... 118

Sandesh Kumar Kafle, František Dařena
PREDICTIVE ANALYTICS IN NEPALESE AGRICULTURE 124

Kateřina Klátilová, Stanislav Mokrý, Jana Turčínková, Michal Pšurný
VISUAL ATTENTION TO INFLUENCER MARKETING CONTENT IN DECORATIVE COSMETICS: EVIDENCE FROM AN EYE-TRACKING EXPERIMENT..... 131

Lukáš Kleveta, Jaromír Landa
A TOOL FOR AUTOMATICALLY TRANSCRIBING IN-TERVIEWS WITH SPEAKER RECOGNITION..... 140

Martin Kmenta
INTEGRATION OF CHRONOS HIGH-SPEED CAMERAS INTO THE MERCURYRT DIGITAL IMAGE CORRELATION SYSTEM 149

Ivan Kolesnychenko, Radim Farana
WEB APPLICATION FOR VISUALIZATION AND COMPARISON OF MINIMUM-REDUNDANCY CODES 155

Lívia Kolibárová, Irena Baláková
ONLINE PLATFORMS AS TOOLS FOR ENVIRONMENTAL CONSUMER EDUCATION 160

Polina Komárková
CZECHUP: A WEB SYSTEM FOR PREPARING FOREIGNERS FOR THE CZECH LANGUAGE EXAM WITH ADAPTIVE VOCABULARY AND AI-GENERATED EXERCISES 169

Benjamin Křištof, Jana Dannhoferová
GENERATION Z AND AI-DRIVEN SHOPPING: COMPARING ONLINE AND IN-STORE CONSUMER EXPERIENCES 176

David Kubart, Zuzana Špendel
SERVICE MANAGEMENT FOR SMART BUILDING ADMINISTRATION USING VIRTUAL IOT SIMULATION AND PREDICTIVE ANALYTICS 186

Tomáš Lacina, Petr Jedlička
LIGHTNING NETWORK AS AN ALTERNATIVE PAYMENT INFRASTRUCTURE FOR SMALL AND MEDIUM-SIZED ENTERPRISES 194

Vojtech Mahdal, Oldřich Trenz
FINE-TUNING ROBERTA FOR SENTIMENT ANALYSIS OF SUPPORT CONVERSATIONS WITH PRIVACY-PRESERVING ANONYMISATION 204

Anna Martinková, Veronika Antošová, Lucie Veselá
INCOME, ATTITUDES, AND FOOD WASTE: SEGMENTATION OF CZECH HOUSEHOLDS 211

Jakub Meduna, Radim Farana
CONTROL OF INTEGRATING SYSTEMS USING A TWO-DEGREE-OF-FREEDOM PID CONTROLLER 220

Jan Mrákava, Radim Farana
PRACTICAL BLOCKCHAIN DEPLOYMENT FOR TICKET MANAGEMENT: A CASE STUDY WITH SOLIDITY SMART CONTRACTS 226

Gabriela Navrátilová, Veronika Antošová, Lucie Veselá
GREENWASHING IN THE MARKETING COM-MUNICATION OF SUSTAINABLE FASHION COM-PANIES: THE ATTITUDES OF GENERATION Y AND Z 232

Markéta Nedělová, Ivo Pisařovic
INTERACTIVE TRANSFORMATION OF LAYERS IN 3D TILES FORMAT WITHIN A WEB GIS APPLICATION 242

Ondřej Pavlík, Oldřich Trenz
A WEB-BASED FINANCIAL PLANNING APPLI-CATION FOR ENTERPRISE BUDGETING AND IN--VESTMENT MANAGEMENT: A DESIGN AND IMPLEMENTATION CASE STUDY 250

Jiří Pokorný, Michaela Veselá, Oldřich Trenz
AUTHORSHIP VERIFICATION OF ENGLISH TEXT: TRADITIONAL STYLOMETRY, TRANSFORMERS, AND LARGE LANGUAGE MODELS COMPARED AND DEPLOYED..... 257

Jan Přichystal
PRODUCT NAME MATCHING..... 264

Jan Přichystal, Tomáš Jakúbek, Roman Valovič
SUMMARY OF FINANCIAL ARTICLES..... 271

Jan Přichystal, Roman Valovič, Tomáš Jakúbek
NPL STUDIO 281

Lenka Procházková, Marek Záboj, Matěj Sysel
FACTORS INFLUENCING THE SELECTION OF CRAFT BEER: A FACTOR ANALYSIS AND GENERATIONAL SEGMENTATION IN THE CZECH MARKET..... 290

Petra Skořepová, Petra Marešová
CLOSING THE GREEN GAP: HARMONIZING PROJECT AND PROCESS APPROACHES IN HIGHER EDUCATION VIA SUSTAINABLE LEADERSHIP AND DIGITAL INTERNAL MARKETING..... 299

Magradléna Skřičilová, Oldřich Trenz
CLASSIFYING SCANNING ELECTRON MICROSCOPE DEFECTS WITH A CNN AND A CUSTOM FOURIER-TRANSFORM LAYER..... 308

Tomáš Sochor, Hana Sochorová
POST-QUANTUM CRYPTOGRAPHY: CURRENT STATE OF STANDARDIZATION AND DEPLOYMENT 316

Damian Sova, Oldrich Trenz
A UNIFIED FORMAT FOR AGGREGATING PRE-TRAINED POSE ESTIMATORS INTO CUSTOM HUMAN--SKELETON DATASETS 323

Petr Straka, Oldrich Trenz
TREESTORAGE: HIERARCHICAL STORAGE WITH VERSIONING AND INHERITED ACCESS CONTROL IN A UNIVERSITY INFORMATION SYSTEM..... 330

Martin Šulc, Ivo Pisařovic
WEB APPLICATION FOR GENERATING TECHNICAL DOCUMENTS USING LARGE LANGUAGE MODELS 338

Ankit Tripathi, Oldrich Trenz
AI METHODS FOR STOCK-PRICE FORECASTING IN THE INDIAN MARKET: THE EFFECT OF MACROECONOMIC FACTORS 348

Ankit Tripathi, Michaela Veselá, Oldrich Trenz
LARGE LANGUAGE MODEL AGENTS AS INVESTMENT ADVISORS: A BACKTESTED COMPARISON OF GPT-3.5, GPT-4 AND GPT-4-TURBO OVER INDIAN STOCK DATA 356

David Vlastník, Jaromír Landa
AN OFFLINE RAG SYSTEM PROTOTYPE FOR SAFE QUERYING OF UNSTRUCTURED DOCUMENTS 363

Jan Vrána, Stanislav Mokrý, Michal Pšurný
THE ROLE OF HAPTIC PRODUCT EXPERIENCE IN CONSUMER EVALUATION: AN EYE-TRACKING AND PUPILLOMETRY STUDY OF REUSABLE BOTTLES 372

Evans Yeboah
FOREIGN DIRECT INVESTMENT, OIL REVENUES, AND TRADE AS DETERMINANTS OF ECONOMIC GROWTH IN GCC COUNTRIES 387

Marek Záboj, Lenka Procházková, Karolína Malinová
FOMO EFFECT AND GENERATION Z'S SENSITIVITY TO MARKETING COMMUNICATION TOOLS IN THE CLOTHING INDUSTRY 396

Štěpán Zálešák, Radim Farana
VISUALIZATION OF ALGORITHMS FOR GENERATING PSEUDORANDOM NUMBERS 404

FOREWORD

Dear readers,

We are honoured to present you with the Conference Proceedings we are publishing as the one of the outputs of the 28th year of the International Conference Economic Competitiveness and Sustainability (ECOS 2026) that we organised in March 26–27, 2026. The conference traditionally takes place in the campus of the Mendel University in Brno, Czech Republic and provides a productive background for the development of new ideas not only for senior but also for young researchers.

We hosted two distinguished well-known speakers who contributed to the conference programme with their keynote speeches. Prof. Gonul Colak (Head of Department of Accounting and Finance, University of Sussex Business School, UK) gave a speech "Government Debt, Corporate Debt, and Financial Crowding-In: A Cross-Country Analysis". Prof. Tomas Jandik (Finance and Dillard's Chair in Corporate Finance, Sam M. Walton College of Business, University of Arkansas, USA) contributed to the discussion on the topic of "Networks Everywhere: How Connectedness Shapes Risk, Information, and Value".

We welcomed more than 100 researchers representing universities and research institutes from countries including Czech Republic, Albania, France, Germany, Hungary, Lithuania, Malaysia, Poland, Serbia, Slovakia, Ukraine and USA, who also served as the discussants of the papers and helped to improve the quality of the research results presented during both conference days.

Another useful opportunity to discuss the conference themes was a moderated Panel Discussion entitled "Ethics, Excellence, and Algorithms: AI's Future in Academia", with five invited experts: Gonul Colak, Tomas Jandik, Adéla Lekešová (AI consultant, BDO Czech Republic), Kamil Malinka (Department of Intelligent Systems, Brno University of Technology) and Tomáš Pitner (Professor at Mendel University in Brno).

The purpose of the conference is to enhance academic debate on current problems in the global socio-economic environment and dynamic technological development. Last but not least it is a chance to meet our research counterparts in person and critically review our perspectives on contemporary research problems.

In the presented Proceedings you can find 44 papers that were recommended by conference discussants and selected on the basis of a careful peer-review process. The papers cover a broad spectrum of topics including artificial intelligence, cybersecurity, information systems, finance, sustainability, public policy, consumer behaviour, and digital innovation. The published papers also underwent a thorough plagiarism detection process.

We would like to thank all the participants in the conference for their inspiring contributions. Furthermore, we are grateful to all the discutants, reviewers and the members of the scientific committee for their contribution to the scientific level of presented papers and organisation of this international scientific conference.

ECONOMIC COMPETITIVENESS AND SUSTAINABILITY 2026

CONTRIBUTIONS

ERP SYSTEM SECURITY THROUGH THE EYES OF USERS: EVIDENCE FROM SERBIAN ENTERPRISES

Sanela Arsić[1], Aleksandra Radić[1], Đorđe Nikolić[1]

[1]Engineering Management Department, Technical Faculty in Bor, University of Belgrade, Vojske Jugoslavije 12, 19210 Bor, Serbia

DOI: https://doi.org/10.11118/ecos2026_001

ABSTRACT

Enterprise resource planning (ERP) systems are integrated business platforms characterised by a high need for security, which is a fundamental requirement for efficient operation. Although many vendors include comprehensive security mechanisms in their ERP solutions, user perceptions of security are also crucial for adequate system adoption. This study examines users' perceptions of ERP system security and reliability in relation to demographic characteristics. Analysis of Variance (ANOVA) was used to identify differences in responses. The empirical research included a sample of 206 ERP system users from organisations in Serbia. The ANOVA results show that user perceptions differ significantly by users' level of education, the organisation's industry sector and economic activity, and the frequency and duration of system use. These findings provide valuable insights into variations in perceptions of ERP system security and help identify behavioural patterns influenced by demographic factors.

Keywords: ERP systems, security perception, user behaviour, ANOVA, demographic factors
JEL Code: M15, O33, D83

1 INTRODUCTION

Enterprise Resource Planning (ERP) systems are integrated information platforms that enable the planning, organisation, monitoring, and control of business processes in real time (Hussain *et al.*, 2024). In many industries, ERP is considered essential for conducting business activities, as it allows companies to achieve high performance. Additionally, ERP connects many participants simultaneously, uniting internal and external information flows, the integrity of which is crucial for the company.

Any unauthorised access to the vast amount of data stored in an ERP system can result in significant financial and reputational losses, making the security and reliability of ERP systems among the most critical issues in modern business (Michelberger & Horváth, 2017). A study by Salih *et al.*

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

(2021) found that ERP system security is the most widely cited impacting factor, indicating that security is a critical consideration in examining ERP implementation and usage success.

ERP insecurity can lead to costly consequences for organisations, affecting both operational performance and long-term stability. SAP, one of the leading ERP vendors, reported that in 2021, every 11 seconds a company was hit by ransomware, while the average cost of recovering from a ransomware attack was $1.85 million (SAP, n.d.). Recently, ERP security has changed dramatically as cybersecurity threats have become more sophisticated, with AI-powered attacks increasing by 95% in 2024 alone (Gangapatnam, 2025).

Current research addresses the technical aspects of ERP security (Boppana *et al.*, 2021; May *et al.*, 2024) and models for their improvement (Atakari, 2024; Annapareddy, 2025; Nittala, 2025). However, the human factor remains the most crucial element for successful system application and use (Almahamid, 2019). Specifically, user perceptions of ERP security are critical for effective system adoption. Perceived security refers to users' perceptions of the reliability and safety of ERP service delivery, the robustness of the system's technical architecture, and the adequacy of its security requirements. Users' attitudes and awareness significantly shape how ERP systems are adopted, used, and safeguarded in everyday operations. Research by Kuo *et al.* (2023) demonstrated that perceived security has a statistically significant impact on ERP users' continued usage intention, indicating that users regard ERP security as an important factor that leads to perceiving ERP as more useful and user-friendly.

Nevertheless, there is a clear need for further research to understand the behavioural issues of ERP users shaped by their psychological beliefs (Salih *et al.*, 2021). Therefore, additional studies are needed to explore the behavioural and cognitive aspects of ERP system introduction in business operations.

The main objective of the study is to provide contemporary insights into differences in ERP users' security perceptions, based on empirical evidence tested using a statistical approach. Such results and implications assist ERP vendors in the process of ERP solution development and ERP consultants in achieving more successful ERP system acceptance during implementation.

Considering all the above, this paper examines ERP security from the users' perspective, emphasising the impact of users' demographic variables on perceived ERP system reliability and security in specific country context. To achieve this aim, an Analysis of Variance (ANOVA) was used to identify differences in users' responses obtained through empirical research, which included a sample of 206 ERP system users from organisations in Serbia.

The paper is organised as follows. After the Introduction, the Theoretical Background reviews previous literature and highlights its main findings, while Hypoheses Formulation section proposes the study research hypotheses. The Research Method section details data collection and the methodology used to process the data. The Results and Discussion section presents the main results and considers the findings. The Conclusion summarises the implications and limitations and outlines the future research agenda.

2 THEORETICAL BACKGROUND

Integrating an enterprise's diverse functions through the implementation of a unique ERP system offers numerous benefits. However, during the implementation process, enterprises may encounter several challenges, such as customisation, data migration, user training, and integration with existing systems (Siraj *et al.*, 2024). From the user's perspective, ERP security is not merely a technical constraint but a functional framework. As users operate across departmental boundaries, including finance, procurement, production, and logistics, the seamless integration of access controls and approval workflows is vital (Shish, 2025). Due to the increasing complexity and sophistication of cyber threats, ERP implementation poses a greater risk to the integrity and functionality of these systems.

Among the most significant threats are malware attacks, which can lead to data corruption or theft, as well as phishing and other forms of social engineering that exploit human weaknesses to gain unauthorised access. Distributed denial-of-service attacks can overload a system and cause downtime, while zero-day attacks exploit vulnerabilities for which security patches are not yet available, making them particularly dangerous. The most sophisticated attacks enable perpetrators to maintain a long-term presence in the network and systematically compromise data (Efe, 2024). Insufficient preparation and neglect of security in ERP projects can lead to failure, as key controls are often left until the end of implementation, when budgets are exceeded and deadlines missed. This practice results in significant system vulnerabilities. Morrisson (2020) states that this approach stems from the misconception that ERP is exclusively an IT project, whereas in reality it requires the engagement and responsibility of the entire enterprise.

This highlights the need to emphasise the importance of the user's role in ERP security. The main goal of ERP security is to reduce the risk of information loss, ensure compliance with state and corporate standards, protect the privacy of internal data, and preserve its integrity (Efe, 2024). ERP system security can be defined as the degree to which the ERP system is regarded as secure for data storage, data exchange, and other business operations (Salih *et al.*, 2021).

Although technical protection measures are important, contemporary literature increasingly emphasises that the human factor is the weakest, yet most important, link in the chain of defence (Triplett, 2022). In current discourse on ERP systems, security is no longer viewed solely through the lens of cryptographic algorithms and technical barriers, but as a complex socio-technical phenomenon centred on human behaviour. Users' awareness and attitudes directly shape how ERP systems are used and protected in daily work (Kuo *et al.*, 2023; Shish, 2025).

Examining ERP system security within the context of Serbian enterprises is particularly important given the heterogeneous levels of digital maturity, organisational practices, and user competencies present in transition economies. While ERP systems are increasingly adopted by organisations in Serbia, variations in implementation quality, user training, and security governance may influence how security and reliability are perceived by end users.

Regarding ERP security in Serbian enterprises, this issue represents a unique combination of strict legal regulations, accelerated digitisation (such as the introduction of e-invoices), and a specific mentality regarding data storage. A particular feature of Serbia is the mandatory use of the Electronic Invoice System, which has introduced new security challenges, such as the requirement that the connection between the local ERP and the national portal must be encrypted and protected by digital certificates to avoid the risk of interception.

Traditionally, most Serbian enterprises have preferred on-premise ERP systems and local data storage. However, this trend is changing, with an increasing number of companies adopting cloud ERP. Nevertheless, research by Ivanović and Marić (2021) shows that over 50% of surveyed

companies wish to switch to cloud ERP technology, but 40% of respondents agree that the major barrier is the problem of data security and safety.

Such insights are essential for designing more effective security policies, user training programmes, and system configurations that align with users' expectations and behaviours. By focusing on the Serbian context, this study contributes to the broader literature by providing context-sensitive evidence on ERP system security from the users' perspective in an underexplored empirical setting.

3 HYPOTHESES FORMULATION

To support the study's objectives, this section presents hypotheses drawn from previous research. The hypotheses were primarily based on previous international ERP and information systems literature, as empirical studies specifically examining perceived ERP security and reliability in the Serbian context are limited. Furthermore, examining these relationships in Serbia is particularly relevant, as transition economies may differ from others. Consequently, the study replicates previous findings and contextualises them within a less frequently investigated context. The proposed hypotheses will be tested to determine how perceptions differ by users' age, education level, organisation's industry sector, economic activity, company size, user experience, and frequency of system use.

Previous research found that males dominate information systems adoption (Kusumaningtyas & Suwarto, 2015). McGill and Thompson (2021) reported that women perceive lower levels of information security and privacy than men, and behaviours requiring greater technical skill are less frequently adopted by female users. Therefore, this study proposes the following hypothesis H1:

H1. There is a statistically significant difference in perceived ERP security and reliability between males and females.

Earlier studies identified a positive relationship between age and information system use, as older individuals have limited knowledge and training regarding computers (Elsadaani, 2013). Additionally, younger people adopt IS more readily than older generations (Kusumaningtyas & Suwarto, 2015). Based on this, the study proposes the following hypothesis H2:

H2. There is a statistically significant difference in perceived ERP security and reliability based on ERP users' age.

Olatokun (2009) found that education level is linked to the technology usage gap. Previous literature (Davis, 1989; Venkatesh & Bala, 2008) indicates that perceptions of ERP system acceptance are often influenced by education. Therefore, this study proposes the following hypothesis H3:

H3. There is a statistically significant difference in perceived ERP security and reliability based on ERP users' education level.

Despite industry specialisation, manufacturing and services remain the two main economic activities and may represent distinct groups of ERP system users. Previous literature (Ruivo *et al.*, 2017) suggests that manufacturing and service industries have different processes and types of data stored in the system (Ifinedo, 2012), which can affect perceptions of ERP security. Based on this, the study proposes the following hypothesis H4:

H4. There is a statistically significant difference in perceived ERP security and reliability based on the industry sector background of ERP users.

Abu Afifa *et al.* (2023) examined users' attitudes towards ERP use in the financial, service, and manufacturing sectors. The study found that ERP adoption is higher in manufacturing than in service and financial sectors. Mishra *et al.* (2022) also highlighted distinct cybersecurity issues in enterprises with different economic activities. Thus, this study proposes H5:

H5. There is a statistically significant difference in perceived ERP security and reliability based on the economic activity background of ERP users.

Regarding company size, Johansson *et al.* (2015) demonstrated that as enterprise size increases, the complexity and scale of ERP security issues also become more significant. The challenge of managing user permissions, ensuring regulatory compliance, and maintaining system integrity is greater in larger enterprises (Mishra *et al.*, 2022). Accordingly, this study proposes the following hypothesis H6:

H6. There is a statistically significant difference in perceived ERP security and reliability based on the company size in which ERP users are employed.

Previous literature indicates that user experience with an ERP system positively affects perceived system security (Bhattacherjee, 2001). Greater exposure to an ERP system increases users' security awareness and understanding of their responsibilities in maintaining system security (Seneviratne & Colombage, 2023). Therefore, this study proposes the following hypothesis H7:

H7. There is a statistically significant difference in perceived ERP security and reliability based on ERP users' experience.

The frequency of ERP system use is positively correlated with users' perceptions of security and reliability (Workman *et al.*, 2009). It influences user familiarity with the system and the trust developed through repeated interaction. Based on this, the study proposes the following hypothesis H8:

H8. There is a statistically significant difference in perceived ERP security and reliability based on the frequency of system use by ERP users.

4 RESEARCH METHOD

To assess user beliefs and attitudes regarding the acceptance and use of ERP systems, an empirical survey was conducted in 2024. Data were collected through an online questionnaire created in Google Forms and distributed via the LinkedIn network to ERP users employed in enterprises operating in Serbia. Participants were informed that their responses would be used exclusively for scientific research purposes.

The distributed questionnaire consisted of two parts. The first part included users' demographic questions, while the second comprised items related to ERP system acceptance and use. The questionnaire used in this study was adapted from Sternad Zabukovšek *et al.* (2019) and covered a wide range of constructs examining drivers influencing ERP system usage, such as perceived usefulness, ease of use, work compatibility, and external factors. Respondents rated their agreement on a five-point Likert scale, where 1 indicated the lowest level of agreement and 5 the highest. Among the items included under external factors was a statement addressing the reliability and security of ERP systems, as this is an important aspect of successful ERP implementation and continued use.

Furthermore, to examine the influence of demographic factors on perceived ERP system security and reliability, an Analysis of Variance (ANOVA) was applied. This statistical analysis was performed using the IBM SPSS v.25 software package.

The ANOVA test is a statistical technique used to determine whether there are statistically significant differences between the means of multiple groups for a selected dependent variable. In this study, ANOVA is used to examine whether users' perceptions of ERP system security and reliability differ significantly across demographic groups such as gender, age, education, industry sector, economic activity, company size, user experience, and frequency of use with ERP systems.

By enabling the comparison of several groups, ANOVA provides an appropriate methodological framework for identifying demographic factors that may influence variations in perceived ERP system security and reliability, thereby directly supporting the study's objective.

5 RESULTS AND DISCUSSION

The sample of 206 ERP users from enterprises operating in Serbia has been gathered to conduct the analysis. Demographic characteristics of the research sample are described in Table 1.

This empirical research included more males (56.3%) than females (43.7%). Most ERP users were aged 30–39 (49.0%) and held either a BSc (43.2%) or MSc (42.2%) degree. In terms of industry sector, 65% of respondents were from manufacturing. The sample displayed industry heterogeneity regarding the distribution of enterprises' economic activity. Most ERP users work for large enterprises, defined as companies with more than 250 employees. Regarding user experience, most users have used ERP systems for 3 to 5 years (51.0%) and use them daily (83.5%).

This paper explores the impact of demographic variables of ERP users on perceived ERP system reliability and security. To achieve this, an ANOVA test was performed, and the results are presented in Table 2.

Tab. 1 Demographic characteristics of the research sample (N = 206)

Variables	Category	Percentage %
Gender	Male	56.3
	Female	43.7
Age	20-29	20.4
	30-9	49.0
	40-49	25.2
	50-59	5.3
Education	High school	7.3
	Higher education	5.8
	BSc	43.2
	MSc	42.2
	PhD	1.5
Industry sector	Manufacturing	65.0
	Service	35.0
Economic activity	IT industry	15.0
	Automotive industry	17.5
	Industrial selling	11.7
	Banking	2.4
	Pharmaceutical industry	9.7
	Production	11.7
	Aviation industry	1.9
	Mining and metallurgy	13.1
	Energy industry	2.4
	Telecommunications industry	3.9
	Food and beverage industry	10.7
Company size	Less than 10 employees	1.9
	11–50 employees	5.3
	51–250 employees	18.9
	Above 250 employees	73.8
User experience	Less than 1 year	5.8
	1–3 years	26.7
	3–5 years	51.0
	More than 5 years	16.5
Frequency of use	Every day	83.5
	1–2 times a week	4.9
	Several times a week	7.8
	1–2 times a month	3.9

Tab. 2 ANOVA test results

Hypothesis	Demographic characteristic	Significance		Hypothesis remark
		F	p	
H1	Gender	1.472	0.226$^{n.s.}$	Rejected
H2	Age	0.503	0.681$^{n.s.}$	Rejected
H3	Education	3.284	0.012**	Accepted
H4	Industry sector	6.041	0.015**	Accepted
H5	Economic activity	2.440	0.007**	Accepted
H6	Company size	2.084	0.103$^{n.s.}$	Rejected
H7	User experience	3.791	0.005*	Accepted
H8	Frequency of use	2.734	0.030*	Accepted

Notes: *statistically significant at 0.01 level; **statistically significant at 0.05 level; n.s. Not significant.

Gender has not been found to have a statistically significant impact on the perceived security and reliability of ERP systems, leading to rejection of H1. Earlier studies (such as Broos, 2005) emphasised that, in the past, women were more anxious about using computers. Some recent studies (Mohamed & Ahmad, 2012; McGill & Thompson, 2021) have indicated differences in information security and privacy perceptions regarding gender. However, these studies explored privacy concerns and security perceptions in relation to social networks or internet technologies used for personal purposes. Moreover, users, regardless of gender, may perceive that ERP vendors have already implemented adequate security measures, diminishing the salience of security concerns in their perceptions. In the context of Serbia, the absence of statistically significant gender differences can be explained by the fact that the use of ERP systems in organisations is largely mandatory and directly related to daily work tasks.

The results also suggest that respondents' age does not have a statistically significant effect on the perceived security and reliability of ERP systems. Therefore, proposed hypothesis H2 is also rejected. ERP system security mechanisms, such as access control, encryption, and activity logging, are mostly invisible to users and are built in at the system level (Atakari, 2024). Therefore, users, regardless of age, do not have direct insight into the technical aspects of security, but form their perceptions based on general work experience and system stability. Furthermore, digital competencies are increasingly less dependent on age, as ERP systems have been present in business practice for many years. This may reduce differences between younger and older users in terms of understanding and trust in information systems. Similarly, the absence of differences between younger and older users can be attributed to the fact that ERP systems have been present in Serbia for some time, particularly in small and medium, and large enterprises, which constitute the dominant structure in this study sample.

Users' education has been found to have a statistically significant impact on the perceived security and reliability of ERP systems, leading to acceptance of H3. The literature suggests that perceptions of ERP system acceptance are often influenced by individual user characteristics, including education (Davis, 1989; Venkatesh & Bala, 2008). Specifically, a higher level of education can lead to a better understanding of how security measures support system stability, which positively

affects perceptions of reliability and safety. For example, Sternad Zabukovšek *et al.* (2019) report that more educated users demonstrate a higher degree of trust in ERP systems, as they are better able to recognise long-term benefits and understand system complexity. To gain deeper understanding, an ANOVA post hoc test was employed. The results indicate that the most pronounced difference is observed between ERP users with a high school degree and those with an MSc degree (mean difference = -0.531; standard error = 0.188; p = 0.042). This finding supports existing literature in information security, which suggests that higher levels of education are associated with greater awareness of security risks and a more accurate understanding of protective mechanisms (Ifinedo, 2012). Furthermore, this difference may also be influenced by the increasing integration of ERP-related content and formal training in higher education curricula. However, this remark should be treated with caution, as the present study did not examine whether participants had undergone formal ERP, cybersecurity, or digital skills training, nor did it assess the quality or intensity of such training.

Industry sector has a statistically significant impact on the perceived security and reliability of ERP systems, resulting in acceptance of H4. The literature indicates that in manufacturing environments, ERP system outages are directly linked to financial losses and supply chain disruptions, reinforcing the importance of security and reliability in user perception (Wagner & Monk, 2008). In contrast, in the service sector, ERP systems often manage large volumes of customer data, financial transactions, and contractual relationships, where issues of confidentiality, privacy, and data protection are prioritised. Studies show that users in the service sector more frequently associate ERP system security with the protection of sensitive information and compliance with regulatory requirements, shaping a different perception of security compared to the manufacturing sector (Ifinedo, 2012).

Similarly, the economic activity of the enterprise has a statistically significant impact on perceived security. In industries such as finance, healthcare, or telecommunications, ERP systems process highly sensitive data such as financial transactions and personal and health information. Therefore, security requirements are stricter and user awareness of potential risks is greater (Trunina *et al.*, 2018). In contrast, in industries like mining and metallurgy, the focus is often on operational reliability and system availability, as even short-term ERP system outages can cause production downtime and supply chain disruptions (Ranitović *et al.*, 2013). Therefore, proposed hypothesis H5 is accepted.

The statistically significant impact of the industrial sector and economic activity is due to the structure of the Serbian economy, where manufacturing, metal processing, mining, and the energy sector continue to play a significant role, with ERP system interruptions directly causing operational and logistical problems. Conversely, the service sector, banking, telecommunications, and healthcare institutions are increasingly dependent on data protection and regulatory compliance, especially following the intensified digitalisation of financial and administrative processes.

Company size has not been found to have a statistically significant impact on perceived ERP system security and reliability, leading to rejection of H6. ERP users, whether they work in small, medium, or large companies, generally have a similar level of exposure to the system's security functionalities. Furthermore, company size is primarily relevant for managers' strategic and technical decisions regarding the selection and implementation of the ERP system (Lutfi *et al.*, 2024), while users form their perceptions of security and reliability based on their use of the system, its stability, and availability, rather than on wider organisational resources or structure.

User experience has a statistically significant impact on the perceived security of ERP systems, as supported by previous literature suggesting that, with longer use, users develop procedural

knowledge and routines, which reduce cognitive effort and increase the perception of system stability (Bhattacherjee, 2001). To gain deeper insights, post hoc analysis was performed, and results indicated that the most pronounced difference is between ERP users with 1–3 years of experience and those with more than five years of experience (mean difference = -0.436; standard error = 0.146; p = 0.027). This supports the statement that experienced users are more likely to evaluate the security of ERP systems based on repeated positive experiences and have greater exposure to various situations when working with an ERP system, including system errors, security procedures, access controls, and the organisation's response to incidents. Therefore, proposed hypothesis H7 is accepted.

The frequency of using an ERP system has a statistically significant impact on perceptions of security and reliability (H8 is accepted), as it influences user familiarity with the system and the trust that develops through repeated interaction. Furthermore, research in information security shows that greater exposure to the system increases users' security awareness and understanding of their role in protecting the system (Workman *et al.*, 2009). This also confirms the results of post hoc analysis, showing the most pronounced difference is between ERP users who use ERP every day and those who use ERP 1–2 times per month (mean difference = 0.797; standard error = 0.261; p = 0.014).

User experience and the frequency of ERP use are particularly relevant in Serbia, where the level of digital transformation among companies varies. The adoption of these systems has increased employees' exposure to digital business processes, which may lead to greater awareness of the importance of security, reliability and data protection in information systems. Furthermore, e-invoicing in Serbia should not necessarily be interpreted as directly improving users' digital competence. Rather, it has significantly increased employees' exposure to digital business processes and information systems, thereby making issues such as cybersecurity, data protection, and system reliability more visible in everyday work activities.

6 CONCLUSION

This study examined ERP system security from the users' perspective, focusing on how demographic characteristics influence perceptions of system reliability and security in Serbian enterprises. Using ANOVA on a sample of ERP users, the research provides empirical evidence of variations in perceived ERP security across different user groups, addressing an underexplored aspect of ERP security research.

The findings show that users' perceptions of ERP system security and reliability differ significantly according to education level, industry sector, economic activity, user experience, and frequency of system use. In contrast, no statistically significant differences were found for gender, age, or company size. These results suggest that perceptual differences are more strongly influenced by users' cognitive, experiential, and contextual characteristics than by basic demographic attributes. More educated and experienced users, as well as those who interact with ERP systems more frequently, tend to evaluate system security more positively, likely due to greater familiarity with system functionality and security mechanisms.

From a theoretical perspective, this study reinforces the view that ERP system security should be understood as a socio-technical construct rather than a purely technical attribute. From a practical standpoint, the results highlight the importance of tailoring security-related training, communication, and support mechanisms to different user groups. These implications are particularly relevant in the context of Serbia, where enterprises are undergoing digital transformation characterised by

heterogeneous levels of digital maturity, different level digital competencies, and varying levels of ERP adoption across industries. Therefore, improving continuous user support, and awareness--raising activities may be essential for strengthening confidence in ERP systems and reducing security-related uncertainty among employees.

Despite its contributions, this study has several limitations. The research is based on self-reported perceptions, which may not fully reflect actual system security or user behaviour. Future research could extend this work by incorporating longitudinal designs, comparative cross-country analyses, or additional behavioural variables such as security awareness, training intensity, or organisational security culture. Such approaches would provide a deeper understanding of how user perceptions of ERP security evolve over time and across different organisational and institutional contexts.

Acknowledgements

The research presented in this paper was carried out with the financial support of the Ministry of Education, Science and Technological Development of the Republic of Serbia within the framework of financing scientific research work at the University of Belgrade, Technical Faculty in Bor, in accordance with the contract with registration number 451-03-34/2026-03/ 200131.

REFERENCES

ABU AFIFA, M. M., VO VAN, H., LE HOANG VAN, T. 2023. Blockchain adoption in accounting by an extended UTAUT model: empirical evidence from an emerging economy. Journal of Financial Reporting and Accounting. 21(1), 5-44.

ALMAHAMID, S. 2019. The influence of ERP system usage on agile capabilities: examining the mediating role of users' psychological empowerment in Jordanian commercial banks. Information Technology & People. 32(6), 1633–1656.

ANNAPAREDDY, V. N. 2025. The Intersection of Big Data, Cybersecurity, and ERP Systems: A Deep Learning Perspective. *Journal of Artificial Intelligence and Big Data Disciplines*. 2(1), 45–53.

ATAKARI, C. 2024. A Deep Learning-Based Security Model for ERP-Integrated IoT in National Defense Manufacturing Environments. *International Journal of Emerging Trends in Computer Science and Information Technology*. 5(3), 90–98.

BHATTACHERJEE, A. 2001. Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*: 351–370.

BOPPANA, S. B., MOORE, C. S., BODEPUDI, V., JHA, K. M., MAKA, S. R., SADARAM, G. 2021. AI And ML Applications In Big Data Analytics: Transforming ERP Security Models For Modern Enterprises. *SSRN Electronic Journal*. 30(S2), 98-117.

BROOS, A. 2005. Gender and information and communication technologies (ICT) anxiety: Male self-assurance and female hesitation. *CyberPsychology & Behavior*. 8(1), 21–31.

DAVIS, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*: 319–340.

EFE, A. 2024. Risk modelling of cyber threats against MIS and ERP applications. *Pamukkale Üniversitesi İşletme Araştırmaları Dergisi*. 11(2), 502–530.

ELSAADANI, M. A. 2013. Exploring the relationship between teaching staff'age and their attitude towards information and communications technologies (ICT). *International Journal of Instruction*. 6(1), 2015–226.

GANGAPATNAM, K. 2025. Securing the Digital Core: Cybersecurity Challenges and Strategies in SAP ERP Systems. *Journal of Computer Science and Technology Studies*. 7(3), 270–276.

HUSSAIN, M. K., SIDDIQUE, A., KAZMI, H., OBAIDULLAH, N. 2024. Factors Affecting User Acceptance of ERP Systems in SMEs. *Bulletin of Multidisciplinary Studies*. 1(1), 01–13.

IFINEDO, P. 2012. Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security*. 31(1), 83–95.

IVANOVIĆ, T., MARIĆ, M. 2021. Application of modern Enterprise Resource Planning (ERP) systems in the era of digital transformation. *Strategic Management – International Journal of Strategic Management and Decision Support Systems in Strategic Management*. 26(4), 28-36.

JOHANSSON, B., ALAJBEGOVIC, A., ALEXOPOULO, V., DESALERMOS, A. 2015,. Cloud ERP adoption opportunities and concerns: the role of organizational size. In: *2015 48th Hawaii international conference on system sciences*. pp. 4211-4219. IEEE.

KUO, C. S., KANG, Y., YANG, H. L. 2023. Investigating the determinants of continuance intention on cloud ERP systems adoption. *Advances in Mechanical Engineering*. 15(4), 16878132231165973.

KUSUMANINGTYAS, N., SUWARTO, D. H. 2015. ICT adoption, skill and use differences among small and medium enterprises managers based on demographic factors. *Procedia-Social and Behavioral Sciences*. 169, 296-302.

LUTFI, A., ALQUDAH, H., AL-DAOUD, K. I., ZAQEEBA, N., ALRAWAD, M., ALMAIAH, M. A. 2024. Technology factors and ERP system efficiency in the Jordanian industrial firms: does company size matter? *Humanities and Social Sciences Communications*. 11(1), 1–13.

MAY, R., BIERMANN, C., KENNER, A., KRÜGER, J., LEICH, T. 2024. A product-line-engineering framework for secure enterprise-resource-planning systems. *Procedia Computer Science*. 239, 1619–1626.

MCGILL, T., THOMPSON, N. 2021. Exploring potential gender differences in information security and privacy. *Information & Computer Security*. 29(5), 850–865.

MICHELBERGER, P., HORVÁTH, Z. 2017. Security aspects of process resource planning. *Polish Journal of Management Studies*. 16(1), 142–153.

MISHRA, A. P., DUBLISH, M., KUMAR, D. 2022. Cyber security application in ERP implementation. *J. Pharm. Negat. Results*. 13, 2507-2522.

MOHAMED, N., AHMAD, I. H. 2012. Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior*. 28(6), 2366–2375.

MORRISSON, M. K. 2020. Best practice models for enterprise resource planning implementation and security challenges. *Journal of Business*. 8(2), 55–60.

NITTALA, E. P. 2025. Zero Trust Security Models in AI-Integrated ERP Platforms for Defense-Grade Business Continuity. I*nternational Journal of AI, BigData, Computational and Management Studies*. 6(2), 75–84.

OLATOKUN, W. M. 2009. Analysing socio-demographic differences in access and use of ICTs in Nigeria using the capability approach. *Issues in Informing Science & Information Technology*. 6, 479-496.

RANITOVIĆ, P., TEPIĆ, G., MATIĆ, B., SREMAC, S., VUKADINOVIĆ, V. 2013. Functional dependency between the logistics security system and the MySAP ERP in metallurgy. *Metalurgija*. 52(4), 545–548.

SALIH, S., HAMDAN, M., ABDELMABOUD, A., ABDELAZIZ, A., ABDELSALAM, S., ALTHOBAITI, M. M. *et al.* 2021. Prioritising organisational factors impacting cloud ERP adoption and the critical issues related to security, usability, and vendors: A systematic literature review. *Sensors*. 21(24), 8391.

SAP. 2024. ERP security in a cybercrime world. *SAP*. [Accessed: 24 December 2025] https://www.sap.com/resources/erp-security

SENEVIRATNE, S. C., COLOMBAGE, L. 2023. The impact of User-Characteristics and Organizational-Characteristics on End-user Satisfaction with Enterprise Resource Planning (ERP) systems. *International Journal of Financial, Accounting, and Management*. 5(1), 75-95.

SHISH, Z. H. 2025. Securing ERP Systems: The Role Of Information Security Analysts In US Textile And Manufacturing Enterprises. *International Journal of Business and Economics Insights*. 5(3), 459–493.

SIRAJ, T., HAQUE, R., CHOWDHURY, S., ISLAM, N., BISWAS, B., CHOWDHURY, K. H. 2024. Analyzing challenges in enterprise resource planning (ERP) implementation in a safety inspection company: An IVT2IF DEMATEL approach. *Optimality*. 1(2), 205–223.

STERNAD ZABUKOVŠEK, S., KALINIC, Z., BOBEK, S., TOMINC, P. 2019. SEM–ANN based research of factors' impact on extended use of ERP systems. *Central European Journal of Operations Research*. 27(3), 703–735.

TRIPLETT, W. J. 2022. Addressing human factors in cybersecurity leadership. J*ournal of Cybersecurity and Privacy*. 2(3), 573–586.

TRUNINA, I., VARTANOVA, O., SUSHCHENKO, O. and ONYSHCHENKO, O. 2018. Introducing ERP System as a Condition of Information Security and Accounting System Transformation. *International Journal of Engineering & Technology*. 7(4.3), 530–536.

VENKATESH, V., BALA, H. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*. 39(2), 273–315.

WAGNER, B., MONK, E. 2008. *Enterprise resource planning*. Course Technology Press.

WORKMAN, M. 2009. How perceptions of justice affect security attitudes: Suggestions for practitioners and researchers. *Information Management & Computer Security*. 17(4), 341–353.

Contact information

Sanela Arsić: e-mail: sarsic@tfbor.bg.ac.rs

Aleksandra Radić: e-mail: aradic@tfbor.bg.ac.rs

Đorđe Nikolić: e-mail: djnikolic@tfbor.bg.ac.rs

CIRCULAR ECONOMY INDICATORS AND DIGITAL TRANSFORMATION: ASSESSING THE RELATIONSHIP BETWEEN ERP ADOPTION AND CIRCULAR ECONOMY PERFORMANCE ACROSS EU COUNTRIES

Sanela Arsić[1], Aleksandra Radić[1], Anđelka Stojanović[1]

[1]University of Belgrade, Technical Faculty in Bor, VJ 12, 19210 Bor, Serbia

DOI: https://doi.org/10.11118/ecos2026_002

ABSTRACT

The shift towards a circular economy requires the systematic integration of sustainability principles into economic and business processes, with digital technologies playing an increasingly important role. Among these, enterprise resource planning (ERP) systems represent an important tool, as they enable the collection, integration, and analysis of data crucial for enhancing resource efficiency and implementing circular economy principles. This study aims to examine the relationship between enterprise-level digitalization, ERP system adoption, and circular economy performance at the country level by using circular economy (CE) indicators across European Union member states. The research adopts a two-phase methodological approach. In the first phase, the TOPSIS multi-criteria decision-making method was applied to rank EU member states based on a set of relevant CE indicators. In the second phase, the resulting TOPSIS scores were used to examine the relationships among circular economy performance, digitalisation, and ERP system adoption using regression analysis and a group-based comparative technique. The results confirm pronounced differences in the performance of the circular economy across the member states of the European Union, with innovative and industrially advanced countries occupying leading positions in the TOPSIS ranking. At the same time, the findings indicate that digitisation at the enterprise level, including the adoption of ERP systems and ICT technologies, is positively associated with circular economy outcomes at the macro-country level, but by itself is not sufficient to achieve high performance without adequate structural and innovation capacities. The findings should be interpreted as macro-level associations rather than direct causal effects, as the study relies on aggregated country-level indicators and proxy measures of enterprise digitalisation.

Keywords: circular economy, ERP systems, digital transformation, sustainability indicators
JEL: Q56, O33, M15, C44

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

The transition to a circular economy is one of the European Union's key strategic priorities in pursuing sustainable development, climate neutrality, and long-term economic resilience (European Commission, 2020; Kirchherr *et al.*, 2017). For a long time, the economy has followed a simple, linear path: extracting resources from the earth, manufacturing products, and discarding them after use. This linear model, often summarised as "take-make-dispose", assumes resources are infinite and waste is an inevitable byproduct. Unlike the traditional linear model, the circular economy focuses on efficient resource use, extending product life cycles, and recycling and reusing materials, thereby reducing the negative environmental impacts of economic activities (Geissdoerfer *et al.*, 2017; Ghisellini *et al.*, 2016). The shift from a linear to a circular economy is a fundamental change, requiring a rethinking not only of waste management but also of how we design, produce, and use products. A circular economy is designed to eliminate waste and keep resources in use for as long as possible. Instead of a straight line ending in landfill, it forms a closed loop. This transition is progressing well across Europe, with clear objectives and increasing momentum for change. To support this shift, the European Union has established a comprehensive monitoring framework to assess circular economy performance, covering a wide range of environmental, economic, and innovation-related indicators (European Commission, 2023).

Concurrently, digital transformation has increasingly been recognized as a key enabler of the shift towards more sustainable economic models. Digital technologies facilitate the collection, processing and analysis of large volumes of data, enhance the transparency of business processes, and support data-driven decision-making (Brynjolfsson & McAfee, 2014). In this context, integrated information systems, particularly enterprise resource planning (ERP) systems, play a pivotal role by linking diverse business functions and enabling the systematic monitoring of resource flows, waste generation and emissions (Seidel *et al.*, 2017).

Despite increasing scholarly interest in the relationship between digitalization and sustainability, much existing research remains focused on the micro level or treats digitalization and sustainability as broad, abstract concepts. Consequently, limited attention has been paid to examining the specific mechanisms through which digital transformation affects circular economy performance at the macro level (de Sousa Jabbour *et al.*, 2020; Rosa *et al.*, 2020). In particular, there is a notable lack of comparative empirical studies that simultaneously analyses circular economy performance, the level of digitalization, and the adoption of ERP systems across European Union member states. To address this research gap, the present study aims to rank European Union member states according to their level of circular economy development and to examine the role of digital transformation and ERP systems in explaining observed cross-country differences in performance. The indicators employed in this study primarily operate at the macro-country level and are intended for comparative assessment and benchmarking of circular economy performance across European Union member states. Accordingly, the analysis does not interpret circular economy indicators as direct measures of firm-level ERP performance, but rather examines whether higher levels of enterprise digitalization and ERP adoption at the national level are associated with stronger circular economy outcomes. Furthermore, the selected indicators predominantly reflect resource efficiency, recycling performance, and material flow dimensions of the circular economy monitoring framework, while acknowledging that certain broader circular economy dimensions, such as regenerative and product life-extension aspects, remain more difficult to capture through currently available harmonized cross-country data. In line with recent circular economy indicator taxonomies, the indicators used in this study primarily belong to the macro-level monitoring domain and are intended to

support benchmarking and comparative policy assessment across countries rather than firm-level operational monitoring. Consequently, the study does not claim that circular economy indicators are directly embedded into ERP architectures or enterprise data infrastructures. Instead, ERP adoption and enterprise digitalisation are treated as macro-level proxy variables reflecting the broader digital readiness of national economies and their potential alignment with circular economy outcomes. To achieve this objective, the study adopts a two-phase methodological approach. In the first phase, the multi-criteria decision-making method TOPSIS is used to rank EU countries based on their circular economy performance (Mardani *et al*., 2018). In the second phase, the resulting TOPSIS scores are used to analyses the relationship between circular economy performance, digitalization, and ERP system adoption using regression analysis and a group-based comparative technique (Afolabi, 2025; Karnaushenko, 2025; Gil-Lamata, 2024).

The findings contribute to a deeper understanding of the complex relationship between digital transformation and sustainability at the macro level. Specifically, the results indicate that digitalization and ERP systems are important enabling factors for the development of a circular economy, while their overall impact depends on broader institutional, economic, and regulatory conditions (Karnaushenko, 2025; Parida *et al*., 2019). In addition to its theoretical contribution, the study provides practical insights for policymakers shaping digital transformation strategies and circular economy policies within the European Union.

2 LITERATURE REVIEW

Recent circular economy literature increasingly distinguishes between different levels and purposes of circular economy indicators, including micro-, meso-, and macro-level indicators designed for monitoring, benchmarking, communication, and policy evaluation purposes. Existing taxonomies also differentiate indicators according to the dimensions of circularity they capture, such as slowing resource loops, closing material loops through recycling, extending product lifecycles, and regenerative resource use. Due to limitations in harmonised EU-wide data availability, the present study primarily focuses on macro-level indicators associated with resource efficiency, recycling performance, secondary raw material use, and sustainability pressures, while acknowledging that regenerative and product life-extension dimensions remain less represented. Recent research increasingly regards the circular economy (CE) as a multi-dimensional policy and performance domain, typically assessed by indicators measuring resource efficiency, waste management, secondary materials, and innovation outcomes. Given the inherently multidimensional nature of CE progress, a growing body of studies uses composite indicators and cross-country rankings to improve comparability and inform policy interpretation across EU member states (Grybaitė & Burinskienė, 2024).

In recent years, the use of multi-criteria decision-making (MCDM) methods has increased significantly in EU-focused circular economy (CE) assessments, as MCDM provides a transparent means of combining diverse indicators into a single performance score. For example, Stanković *et al.* (2021) propose an integrated PCA-PROMETHEE approach to develop a composite index of CE development for EU countries, showing how rankings and clusters can be generated from a structured set of indicators. Similarly, Grybaitė and Burinskienė (2024) use the SAW method to assess CE development across EU countries, offering an indicator-based ranking system that is easy to interpret and communicate in policy contexts. Robustness-oriented studies increasingly combine several MCDM techniques (PROMETHEE, SAW, TOPSIS) to evaluate the stability of rankings and identify which

criteria drive cross-country differences (Grybaitė & Burinskienė, 2024). Furthermore, the MCDM literature has broadened the evaluation scope by incorporating social dimensions of circularity. Kaya *et al.* (2023) focus on social factors within the CE concept for European countries, reinforcing the argument that CE performance is shaped by a broader socioeconomic context beyond purely material and waste indicators.

At the same time, the relationship between digital transformation and circular economy outcomes has attracted increasing academic interest. Macro-level evidence shows that digitalisation can advance the circular economy by enhancing monitoring, coordination, and data-driven decision-making. For example, Afolabi (2025) identifies digitalisation as a key driver of circular economy development in Europe, alongside eco-innovation and economic complexity, while highlighting heterogeneity in effects across countries. Related EU-focused research investigates whether digitalisation promotes circular economy trajectories and emphasises digital capacities as important preconditions for implementing circular practices and policies (Gil-Lamata, 2024). Within the broader digitalisation agenda, ERP systems are increasingly regarded as key infrastructures for

Tab. 1 Selected studies on circular economy indicators, MCDM methods, and digitalisation

Study (Year)	Data source	Circular economy focus	Methodological approach	Digitalisation / ERP dimension
Abobakr *et al*. (2025)	Literature review	Sustainability monitoring and reporting	Conceptual analysis	ERP systems
Anjaria, (2025)	Literature-based	CE-enabling practices	Conceptual framework	ERP systems
Grybaitė & Burinskienė, (2024)	EU monitoring framework	CE performance and criteria importance	SAW	Not included
Afolabi (2025)	Macro-level panel data	Drivers of CE advancement	Econometric analysis	Digitalisation
Gil-Lamata (2024)	Digital performance indicators + CE context	Digitalisation - CE nexus	Macro-level empirical analysis	Digitalisation (DESI-type indicators)
Kaya, Ayçin & Pamucar (2023)	Multi-source indicators	Social dimension of CE	Hybrid MCDM	Not included
Tighnavard Balasbaneh *et al*. (2025)	International sustainability indicators	CE performance assessment	Hybrid MCDM (ANP, TOPSIS)	Not included
Stanković *et al*. (2021)	Eurostat CE indicators (multi-year)	Overall CE development	PCA, PROMETHEE	Not included
Mardani *et al*. (2018)	Secondary indicators (cross-country)	Sustainability & CE evaluation	MCDM	Not included
Zavadskas *et al*. (2016)	Literature review	Sustainability issues	MCDM	Not included

sustainability management because they integrate operational and reporting data across business functions. Recent studies on sustainable ERP systems argue that ERP-enabled monitoring can support resource optimisation, lifecycle tracking, reverse logistics, and ESG reporting. However, realised benefits depend on data quality, organisational adoption, and governance practices (Anjaria, 2025; Abobakr *et al.*, 2025). Table 1 presents an overview of selected recent studies that examine circular economy indicators, apply multi-criteria decision-making methods, and explore the role of digitalisation and ERP systems in the context of the European Union.

As shown in Table 1, the existing literature on circular economy assessment has primarily focused on developing indicators and ranking countries using multi-criteria decision-making methods, while largely neglecting digital transformation and ERP system adoption as explanatory factors. More recent studies have begun to recognise the importance of digitalisation for advancing the circular economy. However, these studies typically examine digital performance in isolation and do not incorporate it into frameworks for evaluating circular economy outcomes. Furthermore, research explicitly addressing the role of ERP systems remains largely conceptual and is rarely analysed alongside macro-level circular economy indicators. In this context, the present study contributes to the field by integrating MCDM-based ranking of circular economy performance with empirical analysis of digital transformation and ERP system adoption across EU countries. This approach addresses a notable research gap and provides a more comprehensive perspective on the digital drivers of circular economy development.

3 METHODOLOGY

3.1 Data sources and sample

This research used a quantitative approach based on secondary data analysis, focusing on the member states of the European Union (EU-27) as the unit of analysis. The analysis covers data from 2020 to 2024, the most recent and comparable years available, ensuring the timeliness of the results and minimising the influence of short-term fluctuations. The data were obtained from the official Eurostat database.

3.2 Selection of circular economy indicators

The selection of circular economy indicators in this study is based on the EU circular economy monitoring framework and supported by recent empirical literature, which highlights the need for a structured and multidimensional approach to assessing circular economy performance at the macro level. To ensure analytical clarity and conceptual consistency, the indicators are grouped into five thematic categories, each representing a distinct dimension of circular economy development. Table 2 presents the evaluated criteria organised into these thematic categories of circular economy indicators.

The **Production and consumption** category includes indicators that capture resource efficiency and waste generation intensity. Resource productivity reflects an economy's ability to generate value with lower material inputs, while municipal waste generation per capita indicates the pressure that consumption patterns place on resource and waste management systems (Ghisellini *et al.*, 2016; Geissdoerfer *et al.*, 2017). Together, these indicators assess how closely production and consumption models align with circular economy principles. The **Waste management** category focuses on the effectiveness of systems designed to close material loops through recycling. Recycling rates of municipal waste and recycling rates of packaging waste by packaging type are

Tab. 2 List of categories and criteria

Category	Code	Criteria	Direction
Production and consumption	PC_1 PC_2	Resource productivity Generation of municipal waste per capita	MAX MIN
Waste management	WM_1 WM_2	Recycling rate of municipal waste Recycling rate of packaging waste by type of packaging	MAX MAX
Secondary raw materials	RM_1	Circular material use rate	MAX MAX
Competitivenes and innovation	CI_1 CI_2	Private investment and gross added value related to circular economy sectors Patents related to recycling and secondary raw materials	MAX MAX
Global sustainability and resilience	GS_1 GS_2 GS_3	Consumption footprint Greenhouse gases emissions from production activities Material import dependency	MIN MIN MIN

key measures of waste management performance and the reintegration of waste materials into economic processes (Ūsas *et al.*, 2025; Tighnavard Balasbaneh *et al.*, 2025). The **Secondary raw materials** category is represented by the circular material use rate, which measures the share of secondary materials in total material consumption. This indicator plays a central role in evaluating the degree of material circularity and is widely regarded as a core measure of circular economy progress (Stanković *et al.*, 2021; Kirchherr *et al.*, 2017). The **Competitiveness and innovation** dimension captures the economic and technological foundations of circular economy development. Private investment and gross value added related to circular economy sectors reflect the capacity of economies to foster innovation, scale circular business models, and enhance competitiveness. In addition, patents concerning recycling and secondary raw materials serve as a direct indicator of innovation activity and technological progress in the circular economy. Patent filings reflect the creation of new processes, materials, and technologies that improve resource efficiency, enable waste valorisation, and facilitate the replacement of primary raw materials (European Commission, 2023; Grybaitė & Burinskienė, 2024). Finally, the **Global sustainability and resilience** category encompasses indicators addressing broader systemic impacts. Consumption footprint, greenhouse gas emissions from production activities, and material import dependency provide insight into environmental pressures and structural vulnerabilities, highlighting the role of circular economy practices in supporting long-term sustainability and economic resilience (Stanković *et al.*, 2021; Grybaitė & Burinskienė, 2024; Ūsas *et al.*, 2025).

The selected indicators are consistent with recent empirical studies that use multi-criteria decision-making approaches to evaluate and rank circular economy performance across EU countries, highlighting the importance of integrating indicators within a unified analytical framework (Ūsas *et al.*, 2025; Grybaitė & Burinskienė, 2024). Their use also aligns with the European Commission's recommendations for monitoring progress towards a circular economy (European Commission, 2023). The optimisation direction for each criterion is specified in Table 2. Criteria for which higher values indicate better circular economy performance are considered benefit (MAX) criteria, while those for which lower values indicate more advantageous environmental or structural outcomes are classified as cost (MIN) criteria. This predefined approach ensures consistent normalisation and supports the application of multi-criteria decision-making methods.

3.3 Entropy-based weighting method

To objectively determine the relative importance of criteria in the model, this study applies the entropy-based weighting method. Entropy-based weighting was applied to objectively determine criterion importance based on data variability, following the information entropy concept introduced by Shannon (1948) and its subsequent application in multi-criteria decision-making methods (Wang & Lee, 2007). The underlying assumption of this approach is that criteria exhibiting greater dispersion across the observed units contain more information and thus contribute more to the evaluation process. Conversely, criteria with limited variation among countries possess lower discriminatory power and are assigned smaller weights.

The entropy weighting method is particularly suitable for macro-level comparative analyses, such as cross-country assessments within the European Union, where transparency and objectivity in weighting are essential.

Entropy-based weighting was applied to objectively determine criterion importance based on data variability, following the information entropy concept introduced by Shannon (1948) and its subsequent application in multi-criteria decision-making methods (Wang & Lee, 2007).

3.4 The TOPSIS method

To rank European Union member states by their level of circular economy development, this study applies the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method. TOPSIS is one of the most widely used multi-criteria decision-making techniques for evaluating complex systems, as it enables the ranking of alternatives based on their relative distance from ideal and anti-ideal solutions (Hwang & Yoon, 1981). The core concept of the TOPSIS method is to identify a positive ideal solution, representing the best performance across all criteria, and a negative ideal solution, representing the worst performance. An alternative is considered preferable if it is closer to the ideal solution and further from the anti-ideal solution. This approach provides an intuitive and transparent ranking mechanism in decision-making contexts involving multiple heterogeneous criteria (Hwang & Yoon, 1981).

In this study, the application of the TOPSIS method follows standard procedural steps, including data normalisation, criterion weighting using entropy-derived weights, determination of ideal and anti-ideal solutions according to the predefined MIN/MAX orientation, and calculation of the relative closeness of each alternative to the ideal solution (Hwang & Yoon, 1981). The resulting TOPSIS score serves as a composite indicator of circular economy performance and forms the basis for ranking EU countries.

TOPSIS is particularly suitable for macro-level comparative analyses, such as cross-country assessments within the European Union, due to its clarity, robustness, and ease of interpretation. Its extensive application in circular economy and sustainability studies further supports its appropriateness for the analytical framework adopted in this research. To improve methodological robustness, the TOPSIS-based ranking was additionally compared with rankings obtained under alternative weighting scenarios and normalization procedures. The comparative analysis showed a high degree of ranking stability among the leading and lagging countries, suggesting that the overall conclusions are not driven by a single specification of the TOPSIS model.

3.5 Analysis of digital transformation and ERP system adoption

After ranking EU countries by circular economy performance using the TOPSIS method, the second stage of the analysis investigates the relationship between circular economy performance and enterprise-level digitalisation, with a particular focus on ERP system adoption. The TOPSIS closeness coefficient (Ci^+) serves as the dependent variable, while digitalisation is evaluated through indicators of ERP and ICT usage in enterprises. Additionally, GDP per capita is included as a control variable to account for cross-country differences in economic development. The TOPSIS closeness coefficient (Ci^+), normalized within the [0, 1] interval, is treated as a continuous composite indicator of circular economy performance in the subsequent regression analysis.

The relationship between digitalisation-related variables and circular economy performance is first examined using regression analysis, facilitating an assessment of how much enterprise-level digital technologies contribute to explaining cross-country variation in TOPSIS-based circular economy outcomes across EU member states. This approach aligns with recent macro-level empirical studies that highlight digitalisation is associated with sustainable and circular economic development (Gil-Lamata, 2024; Afolabi, 2025). Furthermore, a group-based comparative analysis is conducted based on the TOPSIS results. EU countries are classified into three distinct groups according to their Ci^+ values using a natural breaks classification (Ju *et al.*, 2025). The natural breaks approach was selected because it enables the identification of relatively homogeneous country groups based on the empirical distribution of the TOPSIS scores, while preserving interpretability within the policy-oriented comparative framework adopted in this study. This classification enables comparison of average digitalisation indicators and TOPSIS scores across the groups, providing additional insights into structural differences in circular economy performance and the role of digitalisation and ERP systems in distinguishing leading, transitional, and lagging countries.

4 RESULTS

To objectively determine the relative importance of circular economy criteria, the entropy method was applied. This approach assigns weights based on the informational content of the data, so criteria with greater variability across countries receive higher weights due to their stronger discriminatory power. The relative importance of the circular economy criteria, as determined by the entropy weighting method, is shown in Figure 1.

The results indicate that patents related to recycling and secondary raw materials carry the greatest weight, highlighting the importance of innovation capacity in differentiating circular economy performance across EU countries. This outcome is expected, as the entropy method assigns greater weight to criteria with higher variability among countries, and innovation activity is a key factor distinguishing circular economy development within the EU. In contrast, indicators such as recycling rates or consumption footprint tend to be more uniform due to shared EU policies, while innovation capacities reflect structural differences in industrial development and investment in research and development. These findings are consistent with recent studies identifying innovation is associated with circular economy development in European countries (European Commission, 2023; Cagno *et al.*, 2025), confirming the methodological and theoretical relevance of applying the entropy method in this context.

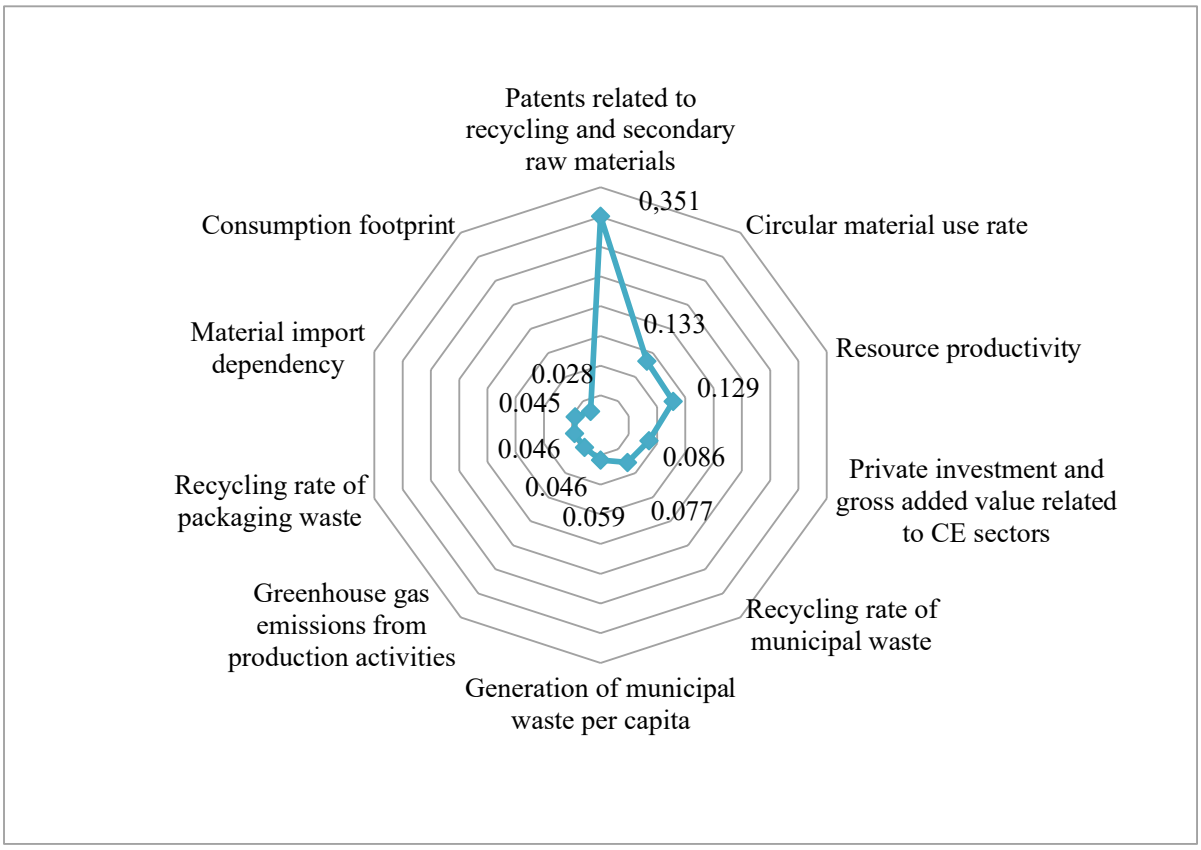


Fig. 1: Entropy-based weights of circular economy criteria

Using the entropy-derived weights and normalised criteria values, the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method was used to rank EU countries according to their circular economy performance. The final result is expressed as the closeness coefficient to the ideal solution (Ci+ score), with higher values indicating better performance. The TOPSIS-based ranking of EU countries by circular economy performance is shown in Table 3.

The TOPSIS results reveal significant heterogeneity among EU member states. Countries with strong industrial bases, advanced innovation systems, and well-established circular economy policies rank highest, while those with lower innovation capacity and structural challenges demonstrate weaker circular economy performance. Additional robustness checks confirmed that the relative positions of leading and lagging countries remained largely stable under alternative weighting specifications, indicating that the obtained ranking structure is not highly sensitive to a single TOPSIS configuration.

Figure 2 shows the TOPSIS-based ranking of EU countries by their circular economy performance, measured by the Ci^+ closeness coefficient. Countries are arranged from highest to lowest Ci^+ values, with colour coding reflecting a natural-breaks classification based on the distribution of Ci^+ scores. Specifically, countries with $Ci^+ \geq 0.30$ are classified as circular economy leaders, those with $0.11 \leq Ci^+ < 0.30$ as transitional performers, and countries with $Ci^+ < 0.11$ as lagging performers. This visualisation highlights pronounced structural differences across the three groups, revealing a compact group of high-performing countries, a broad transitional zone, and a homogeneous cluster of low-performing countries situated far from the ideal solution.

Tab. 3 TOPSIS ranking of EU countries by circular economy performance

Country	S^+	S^-	Ci^+	Rank
Belgium	0.21	0.07	0.25	8
Bulgaria	0.26	0.02	0.07	25
Czechia	0.24	0.04	0.14	14
Denmark	0.24	0.04	0.13	15
Germany	0.05	0.25	0.85	1
Estonia	0.25	0.05	0.15	12
Ireland	0.24	0.04	0.13	16
Greece	0.26	0.02	0.07	26
Spain	0.19	0.08	0.29	7
France	0.15	0.11	0.43	4
Croatia	0.26	0.02	0.08	24
Italy	0.13	0.13	0.49	2
Cyprus	0.26	0.02	0.07	27
Latvia	0.26	0.03	0.10	21
Lithuania	0.25	0.02	0.09	23
Luxembourg	0.25	0.06	0.18	10
Hungary	0.25	0.03	0.09	22
Malta	0.25	0.04	0.15	13
Netherlands	0.15	0.13	0.45	3
Austria	0.16	0.11	0.40	5
Poland	0.21	0.05	0.20	9
Portugal	0.24	0.03	0.11	19
Romania	0.23	0.04	0.17	11
Slovenia	0.25	0.03	0.11	20
Slovakia	0.25	0.03	0.11	18
Finland	0.17	0.10	0.38	6
Sweden	0.24	0.03	0.12	17

The clustering analysis revealed the existence of distinct country archetypes within the European Union. The first cluster included highly digitalised and innovation-oriented economies with strong circular economy performance, while the second cluster represented transitional economies characterised by moderate levels of digitalisation and circularity outcomes. The third cluster comprised countries with comparatively low digital readiness and weaker circular economy performance.

Following the ranking procedure, the relationship between enterprise-level digitalisation indicators and circular economy performance was examined. Specifically, ERP usage, ICT usage in enterprises, and GDP per capita (PPP) as a control variable were analysed. Table 4 presents the correlation coefficients between the CE score and the selected explanatory variables.

The results indicate that ERP system adoption has the most consistent association with circular economy performance, particularly when rank-based relationships are considered. ICT usage shows a moderate linear relationship, while GDP per capita plays a significant role in explaining the ranking structure of CE performance across EU countries.

To assess the relative contributions of digital technologies and economic development, a series of multiple linear regression models was estimated. All coefficients are reported as standardised beta values (Table 5), enabling direct comparison of effect sizes.

The regression results confirm that enterprise-level digitalization, particularly ERP and ICT usage, is a key driver of circular economy performance. While GDP per capita is positively related to circular economy performance at the ranking level, its direct effect becomes statistically negligible once digital variables are introduced, suggesting that digital transformation partially mediates the impact of economic development.

Tab. 4 Correlation between CE score and explanatory variables

Variable	Pearson r	Spearman ρ
ERP usage	0.341	0.433
ICT usage in enterprises	0.362	0.154
GDP per capita (PPP)	0.234	0.493

Tab. 5 Regression results (standardized β coefficients)

Model	ERP usage	ICT usage	GDP per capita	R^2
Model 1: CE ~ ERP	0.341	–	–	0.116
Model 2: CE ~ ERP + ICT	0.374	0.394	–	0.270
Model 3: CE ~ ERP + ICT + GDPpc	0.349	0.393	0.046	0.272

5 DISCUSSION

In recent years, there has been increased focus on developing the circular economy across the European Union, particularly regarding sustainability, resource efficiency, and long-term economic resilience (European Commission, 2023). Despite growing policy efforts and academic interest, significant differences in circular economy performance persist among EU member states, raising questions about the structural and technological factors underlying these disparities.

The findings of this study largely align with previous research on the development of the circular economy at the EU level, while also providing methodological and empirical improvements. Studies employing multi-criteria decision-making approaches to rank countries, such as Ūsas *et al.* (2025), Grybaitė and Burinskienė (2024), and earlier researchs (Mardani *et al.*, 2018; Zavadskas *et al.*, 2016), consistently show notable differences in circular economy development among EU member states, with more industrialised and innovation-focused countries frequently leading. These conclusions are fully consistent with this study's results, which identify Germany, Italy, the Netherlands, France, Austria, and Finland as top performers based on the TOPSIS ranking. Germany's leading position can be directly attributed to its high scores on innovation-related indicators, which, according to the entropy analysis, have the greatest impact on the final ranking. This finding is in line with previous studies showing that countries with robust industrial foundations, well-developed innovation systems, and long-term sustainability policies tend to achieve better circular economy outcomes (Mohajan, 2021; Grybaitė & Burinskienė, 2019). Conversely, countries at the lower end of the ranking (Cyprus, Greece, Bulgaria, Croatia, Lithuania, Hungary) consistently record lower scores across most criteria, especially in innovation, resource productivity, and waste management, indicating the presence of structural constraints that impede their transition towards a circular economy model. The classification of EU countries into leader, transitional, and lagging clusters highlights that differences in circular economy performance stem primarily from structural capacities and innovation readiness rather than from marginal variations in individual indicators.

However, unlike most of the studies mentioned, which focus solely on circular economy indicators and their relative significance, this research also examines the role of enterprise-level digitalisation in explaining cross-country differences in circular economy performance. Conceptual studies highlighting the importance of ERP systems in supporting sustainable and circular business practices (Abobakr *et al.*, 2025; Anjaria, 2025) provide a theoretical basis for selecting ERP and ICT variables in this study, while macro-level empirical analyses confirm that digitalisation is associated with progress in the circular economy (Afolabi, 2025; Gil-Lamata, 2024). In this context, the results of regression analysis indicate that ERP usage and ICT adoption are positively associated with circular economy performance, while including GDP per capita as a control variable does not significantly alter these relationships. This finding suggests that the digitalisation variables explain a substantial share of cross-country variation on differences in circular economy performance than the overall level of economic development. By selecting ERP and ICT variables, the analysis focuses on specific digital practices within companies that directly influence resource planning, material flow management, and the coordination of circular activities. These findings are consistent with recent studies emphasising digital technologies as key enablers of circular business models and enhanced resource efficiency (Neri *et al.*, 2023; Cagno *et al.*, 2025), while also suggesting that economic wealth alone is insufficient for achieving high circular economy performance. Accordingly, this study not only confirms existing evidence on the heterogeneity of circular economy development within the EU, but also advances previous research by combining multi-criteria ranking methods with enterprise-level digitalisation analysis in a single analytical framework.

This study contributes to the literature on circular economy development in the European Union in several important ways. First, it advances previous research by integrating a multi-criteria decision-making approach with an empirical analysis of enterprise-level digitalisation, moving beyond studies that focus solely on circular economy indicators and their relative importance. By combining TOPSIS-based country ranking with regression and group-based comparative analyses, the study provides a more comprehensive framework for understanding cross-country variations in circular economy performance. Second, the findings provide new empirical evidence on the role of operational digital readiness, demonstrating that ERP system usage and ICT adoption are linked with circular economy performance, even after controlling for economic development. This result refines existing insights by showing that digitalisation contributes independently to circular economy outcomes, rather than merely reflecting broader economic conditions. Third, the study shows that digital technologies, while important, are not sufficient on their own to ensure high circular economy performance. The identified leader, transitional, and lagging country clusters reveal that structural capacities and innovation readiness play a decisive role in shaping circular economy outcomes. In this way, the research deepens theoretical understanding of the interaction between digitalisation, innovation systems, and structural conditions in the transition towards a circular economy.

6 CONCLUSION

The integration of digital transformation processes into circular economy strategies has become an increasingly important issue in both academic research and policy discussions within the European Union. Hence, this study examined the relationship between circular economy performance and enterprise-level digitalisation in the European Union, with particular emphasis on the role of ERP systems. Using a two-phase methodological approach, the research combined a TOPSIS-based composite index and country ranking with regression and group-based comparative analyses to explore the drivers of cross-country differences in circular economy outcomes.

The results confirm substantial heterogeneity in circular economy performance across EU member states, with innovation-oriented and industrially advanced countries consistently achieving higher rankings. At the same time, the findings demonstrate that enterprise-level digitalisation, including ERP system usage and ICT adoption, is positively and significantly associated with circular economy performance. Importantly, these effects persist even when differences in economic development are taken into account, indicating that digitalisation of business processes plays a more decisive role than overall economic wealth alone. However, the analysis also shows that digital technologies are not sufficient by themselves to ensure high levels of circular economy performance. Countries that perform best combine digital readiness with strong structural and innovation capacities, while lagging countries face persistent constraints related to innovation, resource productivity, and waste management systems. In addition to its analytical contribution, the proposed composite CE index may serve as a policy-oriented benchmarking and monitoring tool for assessing cross-country differences in circular economy progress and digital readiness within the European Union.

Although the results offer significant insights into the connections between circular economy performance and digitisation in EU member states, the research has several limitations. The analysis relies on aggregated country-level data, which does not capture the heterogeneity of sectors and enterprises. Future research should extend the present macro-level approach by

integrating firm-level ERP log data and operational sustainability indicators in order to examine how granular resource-flow monitoring and digital process integration relate to circular economy outcomes within organisations. In this sense, the present study may serve as an initial macro-level step toward more detailed micro-level investigations of digital-enabled circular economy practices. The methodological framework, based on a combination of the entropy method and TOPSIS, could be extended by applying alternative or hybrid MCDM methods, such as PROMETHEE, VIKOR, or hybrid MCDM models, to assess the robustness and stability of the obtained rankings. Future studies could further strengthen the comparative analysis by applying clustering techniques such as k-means clustering combined with silhouette coefficient validation in order to identify more refined archetypes of digitalisation and circular economy performance across EU countries. Additionally, the set of indicators could be further expanded to include institutional frameworks, regulatory policies, and innovation capacities, enabling a more comprehensive assessment of circular economy development.

Acknowledgement

The authors wish to acknowledge the support by COST (European Cooperation in Science and Technology; www.cost.eu), in the framework of COST Action CA22110 "Cooperation, development and cross-border transfer of Industrial Symbiosis among industry and stakeholders" (LIAISE: www.liaise-action.eu) and Ministry of Education, Science and Technological Development of the Republic of Serbia, as part of the scientific research work at the University of Belgrade, Technical Faculty in Bor, according to the contract with registration number 451-03-34/2026-03/200131. The authors would also like to thank Sandra Vasković, English teacher, for proofreading the paper.

REFERENCES

Abobakr, M. A., Abdel-Kader, M., F. Elbayoumi, A. F. (2024). An experimental investigation of the impact of sustainable ERP systems implementation on sustainability performance. *Journal of Financial Reporting and Accounting*. https://doi.org/10.1108/jfra-04-2023-0207

Afolabi, J. A. (2025). Advancing the Circular Economy in Europe: The Role of Eco-Innovation, Economic Complexity and Digitalization. *Technology in Society*. 103027.

Anjaria, K. (2025). Circular Economy in S-ERP and Its Integration with Business Functions in Industry 4.0. In: *Sustainable Enterprise Resource Planning (S-ERP) for Industry 4.0: A Secure and Ethical Deployment Approach*. Singapore: Springer Nature Singapore, pp. 149-195.

Brynjolfsson, E., McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York, NY: W. W. Norton & Company.

Cagno, E., Neri, A., Negri, M., Bassani, C. (2025). Understanding how circular economy practices and digital technologies interact: A quantitative assessment. *Journal of Cleaner Production*. 430, 140012. https://doi.org/10.1016/j.jclepro.2024.140012

de Sousa Jabbour, A. B. L., Jabbour, C. J. C., Godinho Filho, M., Roubaud, D. (2020). Industry 4.0 and the circular economy: A proposed research agenda and original roadmap. *Technological Forecasting and Social Change*. 154, 119948. https://doi.org/10.1016/j.techfore.2020.119948

European Commission. (2020). *A new Circular Economy Action Plan*. Brussels: EC.

European Commission. (2023). *Monitoring framework for the circular economy*. Brussels: EC.

Geissdoerfer, M., Savaget, P., Bocken, N. M. P., Hultink, E. J. (2017). The Circular Economy – A new sustainability paradigm? *Journal of Cleaner Production*. 143, 757-768. https://doi.org/10.1016/j.jclepro.2016.12.048

Ghisellini, P., Cialani, C., Ulgiati, S. (2016). A review on circular economy: The expected transition to a balanced interplay of environmental and economic systems. *Journal of Cleaner Production*. 114, 11–32. https://doi.org/10.1016/j.jclepro.2015.09.007

Gil-Lamata, M. (2024). Does digitalization foster the transition towards a circular economy in the European Union? *Technological Forecasting and Social Change*. 198, 122951. https://doi.org/10.1016/j.techfore.2023.122951

Grybaitė, V., Burinskienė, A. (2024). Assessment of circular economy development in the EU countries based on SAW method. *Sustainability*. 16(21), 9582.

Hwang, C. L., Yoon, K. (1981). *Multiple attribute decision making: Methods and applications*. Berlin: Springer.

Ju, B., Chen, Z., Geng, Y., Xing, P., Han, L., Wang, L., Wei, W. (2025). Mapping regional disparities in the global shift toward decarbonized power systems. *Carbon Neutrality*. 4(1), 1-14.

Karnaushenko, A. (2025). Digital drivers of business model transformation in the circular economy paradigm. In: E*conomy in the era of digital transformation: trends, opportunities and perspectives*. Scientific Route OÜ, pp. 27–50.

Kaya, S. K., Ayçin, E., Pamucar, D. (2023). Evaluation of social factors within the circular economy concept for European countries. *Central European Journal of Operations Research*. 31(1), 73-108.

Kirchherr, J., Reike, D., Hekkert, M. (2017).Conceptualizing the circular economy: An analysis of 114 definitions. *Resources, Conservation & Recycling*. 127, 221-232. https://doi.org/10.1016/j.resconrec.2017.09.005

Mardani, A., Jusoh, A., Halicka, K., Ejdys, J., Magruk, A., U Ahmad, U. N. (2018). Determining the utility in management by using multi-criteria decision support tools: a review. *Economic Research-Ekonomska Istraživanja*. 31(1), 1666-1716.

Mohajan, H. K. (2021). Germany is ahead to implement sustainable circular economy. *Journal of Economic Development, Environment and People*. 10(2), 46-64.

Neri, A., Negri, M., Cagno, E., Franzò, S., Kumar, V., Lampertico, T., Bassani, C. A. (2023). The role of digital technologies in supporting the implementation of circular economy practices by industrial small and medium enterprises. *Business Strategy and the Environment*. 32(7), 4693-4718.

Parida, V., Sjödin, D., Reim, W. (2019). Reviewing literature on digitalization, business model innovation, and sustainable industry: Past achievements and future promises. *Sustainability*. 11(2), 391.

Rosa, P., Sassanelli, C., Urbinati, A., Chiaroni, D., Terzi, S. (2020). Assessing relations between Industry 4.0 and circular economy: A systematic literature review. *International Journal of Production Research*. 58(6), 1662-1687. https://doi.org/10.1080/00207543.2019.1680896

Seidel, S., Recker, J., vom Brocke, J. (2017). Sensemaking and sustainable practicing: Functional affordances of information systems in green transformations. *MIS Quarterly*. 41(4), 1273-1299.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*. 27(3), 379-423.

Stanković, J. J., Janković-Milić, V., Marjanović, I., Janjić, J. (2021). An integrated approach of PCA and PROMETHEE in spatial assessment of circular economy indicators. *Waste Management*. 128, 154-166.

Tighnavard Balasbaneh, A., Aldrovandi, S., Sher, W. (2025). A systematic review of implementing multi-criteria decision-making (MCDM) approaches for the circular economy and cost assessment. *Sustainability*. 17(11), 5007.

Ūsas, J., Balezentis, T., Streimikiene, D. (2025). Development and integrated assessment of the circular economy in the European Union: The outranking approach. *Journal of Enterprise Information Management*. 38(1), 243-260.

Wang, Y. J., Lee, H. S. (2007). Generalizing TOPSIS for fuzzy multiple-criteria group decision-making. *Computers & Mathematics with Applications*. 53(11), 1762-1772.

Zavadskas, E. K., Govindan, K., Antucheviciene, J., Turskis, Z. (2016). Hybrid multiple criteria decision-making methods: A review of applications for sustainability issues. *Economic research-Ekonomska istraživanja*. 29(1), 857-887.

Contact information

Sanela Arsić: e-mail: sarsic@tfbor.bg.ac.rs

Aleksandra Radić: e-mail: aradic@tfbor.bg.ac.rs

ECONOMETRIC MODEL OF GENERALIZED LINEAR REGRESSION FOR PREDICTING FUTURE NUMBERS OF INSURED PEOPLE BY GENERAL HEALTH INSURANCE COMPANY OF THE CZECH REPUBLIC – COMPARISON OF RESULTS AFTER FIVE YEARS

Jaromír Běláček[1]

[1]Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_003

ABSTRACT

This paper compares the results of predicting the number of insured persons of the General Health Insurance Company of the Czech Republic in 2018, based on a generalized-discounted linear regression model (GLM), which formalizes the projected number of insured persons in age groups through exogenous demographic, re-reregistration, and formal linear and quadratic trend components, with reality observed in 2018–2022. Compared to standard in-sample methods, derived population projections, or the traditional indexing method, the 2018 method, where the calibration of discount factors α meets the criteria for maximizing the determination coefficient R2(α), overall yields better results. Accurate predictions of future numbers of insured individuals have the potential to better calculate and plan future expenditures of the health insurance funds during current period of population aging, and for several years ahead, too.

Keywords: public health insurance, time series forecasting, derived population projection, exponential smoothing, coefficient of determination, weighted-discounted least squares
JEL Code: C53, I11, J11

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 DEVELOPMENT OF THE ISSUE

In 2016–17, the task of predicting the future number of its insured individuals was addressed at the General Health Insurance Company of the Czech Republic (hereinafter referred to as VZP ČR). In the paper (Běláček, 2018), the deterministic model for predicting the future number of VZP insured was presented, based on the upper variant of the demographic forecast by the Czech Statistical Office for the Czech Republic from 2013 and on additive adjustments of extrapolated quarterly re-registration balances (between VZP and other health insurance companies in the Czech Republic) – see the model (1) in Section 2.1 below. In the subsequent paper (Běláček, 2019), a model of generalized discounted linear regression was described, where the weights of retrospective observations might be directly proportional to exponentially decreasing discounts α (α < 1) assigned formally to time indices. In our review article, a universal algorithm described in (Cipra, 1986, p. 233) was used to estimate parameters, which also allows the choice of discounts α > 1 – this model is presented in section 2.2.

In this paper, we compare the results of the prediction published in (Běláček, 2019) from the threshold date of 12/31/2017 to the actual numbers of insured persons obtained for the years 2018–22 retrospectively from the VZP yearbooks as of 12/31 of each calendar year (Yearbooks of the VZP of the Czech Republic for the years 2018–22, 2023). The aim of this paper is to compare the numbers of VZP insured persons determined after five years in total and in age groups – a) with the values predicted from the threshold year 2017; – b) with the predictions based on in-sample minimization of squared deviations in the last five years of the pre-threshold period; – c) compared to the best fit with real data from 2018–22 found ad poste – see paragraphs 2.3 and 3. The discussion of the presented results will provide general conclusions and recommendations for the possibility of using this technologies in comparable applications.

2 DATA SOURCES AND METHODOLOGY

2.1 Data and deterministic models

To project the future number in VZP insured persons, the final quarterly time series in total and in five-year age groups and the similarly structured balances of re-registrations of insured persons between VZP and rest health insurance companies in the Czech Republic, taken from the internal databases of VZP, were used. The primary time series (in thousands of persons) are visualized by black dots in Figures 1–5 for the total of all insured persons ($x =$ '0+') and five-year age groups ($x =$ '0-4', ... , '85+'). An important part of these figures are three projection deterministic models generated in paper (Běláček, 2018) using the formula

$$\widehat{P}_x^t = \left(\widehat{P}_x^{t-1} + S_x^t \right) \cdot \left(D_x^t / D_x^{t-1} \right), \quad t = t_0, t_0 + 1, \dots , \tag{1}$$

where $\widehat{P}_x^{t0} \equiv P_x^{t0}$ denotes the known entry numbers of insured persons at age x in the threshold year t_0 (at the end of the entry quarter), $D_x{}^t$ is the mean number of living persons or interpolated per quarter from data officially projected by the Czech Statistical Office at time t (see Projekce obyvatelstva ČR do r. 2100, 2013) and $S_x{}^t$ denotes the balance of VZP re-registrations assumed (or extrapolated) for the period between years $t-1$ a t (see also in Běláček, 2019 for more details).

With the special choice $S_x^t = 0$ for all future times t model (1) is reduced to the common model of the so-called derived population projection (in Figures 1–5 marked as 'D-model' shown by light diamonds). But we can also consider the alternative of pure projection of cumulative re-registrations (in Figures 1–5 marked as 'E-model' represented by dashed connecting lines), which we obtain by choosing $D_x^t / D_x^{t-1} = 1$. This second development scenario is not realistic for most age groups, but it can be well used for the purposes of assessing the possible mutual synergistic influence of both explanatory components (demographic and re-registrations) on the future development of the number of insured persons. This is synthesized within the framework of the third deterministic model (in legends of Figures 1–5 marked as 'D+E-model' depicted by dark diamonds) created by directly applying formula (1).

2.2 Generalized linear models

All three deterministic models above were calculated in relation to the threshold date of 12/31/2017 toward the future (prospectively), but corresponding time series can also be calculated toward the past (retrospectively), as shown in Figures 1–5. These representations are important from the point of view in comparison with results for the following generalized linear regression models and with the actual numbers of insured persons determined later. They point out in what form mechanically generated demographic or reregistration variables can or must enter the subsequent weighted linear regression models as exogenous variables. Their formal significance in regression models can then be assessed using methods of statistic inference.

The most general linear regression model used later to predict the future number of VZP insured persons in each age group (see Běláček, 2019, p. 5) is based on a four-parameter linear regression model for difference

$$\Delta\left(P_x^t\right) = b_x + c_x \cdot t + d_x \cdot \Delta\left(D_x^t\right) + e_x \cdot S_x^t + \varepsilon\left(0, w_t \cdot \sigma_x^2\right), \quad t = 1, \ldots, t_0 \ , \qquad \textbf{(2)}$$

where $\Delta\left(P_x^t\right) = \left(P_x^t - P_x^{t-1}\right)$ denotes denotes the retrospective difference in the number of insured persons in group x between the ends of adjacent quarters, t is a time variable (with a projection threshold of $t_0 = 12/31/2017$), $\Delta(D_x^t)$ denotes the interpolated quarterly increments of the demographic variable (derived from historical data for the Czech Republic) and S_x^t is the balance of re-registrations from formula (1); further, let us denote by the symbol $\boldsymbol{\theta}_x = (b_x; c_x; d_x; e_x)$ the vector of linear regression parameters from (2) and $\varepsilon(0, w_t \,.\, \sigma_x^2)$ the error term – with a zero mean value and a variance in the form of the product $w_t \,.\, \sigma_x^2$ for the weight factor wt functionally independent of the system parameter σ_x^2. We consider important that all terms on the right-hand side in (2) are now part of the standard additive regression model, which includes the earlier deterministic models from (1) be his special cases (for fixed choices: $b_x = c_x = 0$ and $d_x = 1$ or $e_x = 1$).

We obtained the parameter estimates $\widehat{\theta_x} = \left(\hat{b}_x; \hat{c}_x; \hat{d}_x; \hat{e}_x\right)$ in model (2) by minimizing (by $\alpha_x \in [0.65; 1.0]$) the weighted sum of squares

$$\sum_{k=1}^{t_0} w_t \cdot \left[\Delta\left(P_x^k\right) - b_x - c_x \cdot k - d_x \cdot \Delta\left(D_x^k\right) - e_x \cdot S_x^k\right]^2, \qquad \textbf{(3)}$$

namely through a matrix formula

$$\widehat{\boldsymbol{\theta}}_x = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\Delta\left(P_x\right) = \left(\left(\boldsymbol{X}\boldsymbol{W}^{\frac{1}{2}}\right)'\left(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}\right)\right)^{-1}\left(\boldsymbol{X}\boldsymbol{W}^{\frac{1}{2}}\right)'\left(\boldsymbol{W}^{\frac{1}{2}}\Delta\left(P_x\right)\right), \quad \textbf{(4)}$$

where $\boldsymbol{X}$ is the matrix of the system of linear equations from (2), $\Delta(\boldsymbol{P}_x)$ its corresponding vector with components $\Delta\left(P_x^t\right)$ and $\boldsymbol{W}$ a square diagonal matrix with diagonal elements $w_t = \alpha_x^{-t}$ for the discount factor $\alpha_x > 0$. This means that for $\alpha_x < 1$ we assign an exponentially decreasing influence to the historical data (with gradual distance from the threshold value t_0), for $\alpha_x = 1$ the estimates $\widehat{\theta_x}$ of the parameter vector $\boldsymbol{\theta}_x$ correspond to the classical unweighted linear regression (all past values of the time series have the same weight) and for $\alpha_X > 1$ the historical values of the time series increase their weight exponentially. Substituting the optimized estimates $\widehat{\theta_x}$ from (4) into formula (3) for $\boldsymbol{\theta}_x$ we obtain estimates $\widehat{\sigma_x^2}$ for the variance parameters σ_x^2 and we are counting the coefficients of determination $\widehat{R}^2\left(\alpha_x\right)$, depending on the selected discount level α_x (otherwise also the percentages of variance of the dependent difference variable explained by each of the selected regression models in age group). The final prediction of the number of insured persons for times $t = t_0 + 1$, ..., T, where T denotes an unspecified projection horizon, was finally obtained by cumulative calculation

$$\begin{aligned}\widehat{P}_x^t\left(\alpha_x\right) &= \widehat{P}_x^{t-1}\left(\alpha_x\right) + \Delta\left(\widehat{P}_x^t\left(\alpha_x\right)\right) = P_x^{t_0} + \sum_{k=t_0+1}^{t}\left[\widehat{b_x} + \widehat{c_x}\cdot k + \widehat{d_x}\cdot\Delta\left(D_x^k\right) + \widehat{e_x}\cdot S_x^k\right] \\ &= \widehat{A_x} + \widehat{B_x}\cdot\left(t - t_0\right) + \widehat{C_x}\cdot(t - t_0)^2 + \widehat{D_x}\cdot\sum_{k=t_0+1}^{t}\Delta\left(D_x^k\right) + \widehat{E_x}\cdot\sum_{k=t_0+1}^{t}S_x^k,\end{aligned} \quad \textbf{(5)}$$

where $\Delta\left(\widehat{P}_x^t\left(\alpha_x\right)\right)$ is the prediction obtained from the difference model (2) after substituting $\widehat{\theta_x} = \hat{\theta}\left(\alpha_x\right)$ for components of $\boldsymbol{\theta}_x$ (and after omitting the error term); for the newly calculated parameters in model (5) the easily verifiable relations hold: $\widehat{A_x} = P_x^{t_0}$, $\widehat{B_x} = \widehat{b_x} + \frac{(t_0+1)}{2}\cdot\widehat{c_x}$, $\widehat{C_x} = \frac{1}{2}\cdot\widehat{c_x}$, $\widehat{D_x} = \widehat{d_x}$, $\widehat{E_x} = \widehat{e_x}$. The advantage of the calculation method based on the difference model is the possibility of setting the absolute term $\widehat{A_x}$ in the cumulative model (5) directly to the last known value in the threshold year t_0, so we do not have to estimate this formal term, although formally important for extrapolation.

2.3 Comparison of model predictions with reality

The final data as of 12/31 of each year from the VZP yearbooks are shown in Graphs 1–5 with large fulfilled black circles until 2022. In 2002–2017, they do not visually differ from the quarterly data from the internal VZP databases. However, in 2022, massive increases were recorded in most age groups due to the influx of immigrants from Ukraine, as VZP was the majority health insurance company registering these refugees in the Czech Republic at that time. Of course, the prediction models from 2017 could not take this unexpected entry into account. However, since Ukrainians are not separated from the portfolio of other insured persons in the age structures of the annual VZP reports, their numbers in each five-year age group were qualifiedly estimated based on the balance of foreign migration of the Czech Republic by age and citizenship – see Tables H.12 and H.14 in (Demographic Yearbook of the Czech Republic – 2022, 2023). The numbers of VZP insured persons after deducting the estimated balance of Ukrainians from 2022 are shown as large circles without filling in Figures 1–5.

Already as part of solving the original problem of predicting the numbers of insured persons, a user application was created in MS Excel at 2017 (currently see Excel application of the prediction model, 2025), which allowed the selection of optional user specifications: For example, the

historical evolution of the number of VZP insured persons in each age group could be alternated by time series from the database of all health insured persons in the Czech Republic (a more conservative, but later universally useful variant), the future development of VZP re-registration balances could be selected in three variants of extrapolation: a) constant zero; b) more conservative logarithmic – later selected as optimal one; c) more progressive logarithmic. Furthermore, the graphs visualize quarterly predictions for the three deterministic models above and selected GLM models (currently only for 'ABDE' and 'ABCDE') under the mode of several different options of choosing discount factors α_x (displayed in the header of graphs), all with a projection horizon of T=12/31/2032. GLM models are supplemented in the legends with values of determination coefficients $\widehat{R}^2(\alpha_x)$.

Hereafter, for the couple of discounted regression models mentioned above, the graphs preferentially display extrapolation according to the value of the individually selected discount factors α_x (within the "0 = ALFA by choice" mode). After archiving the $\widehat{R}^2(\alpha_x)$ values following the experimental choice of $\alpha_x \in [0.65; 1.0]$ with a step of 0.01 in each age group the optimum alpha has been specified, how it was used for prediction in 2018. The visualization of these curves can be automatically called up within the mode „1 or 2 = /optALFA for 'ABDE' or 'ABCDE'/", which is also the current display in Graphs 1–5. However, since the chosen optimal selection method for α_x was used rather intuitively and heuristically in 2018, now an attempt was made to compare it with the results of α_x selections using the commonly recommended version of the in-sample method (Bělohradský – Šolc, 2018), which, instead of discounting within all historical data, minimizes only the unweighted sums of squares in the last few pre-threshold years. Thus, in Excel, values of $\alpha_x \geq 0.65$ were step by step archived and after that found such for which was in age group x minimized the following sum of squares:

$$SS_{2012-16} = \sum_{k=t_0-5}^{t_0-1} \left[{P_x}^k - \widehat{P}_x^{\,k}(\alpha_x) \right]^2, \tag{6}$$

i.e. for the last five years before the threshold period. This α_x option can be initialized in the Excel application by setting mode "3 or 4 /= minALFA for 5year's old residua 'ABDE' or 'ABCDE'/" for automatic recalling such minimum alpha's (with the identification „-5Y's ALFA" in the header of created interactive charts).

Finally, we obtained the final variant of choosing discount factors α_x by minimizing throught formula

$$SS_{2018-22} = \sum_{k=t_0+1}^{t_0+5} \left[{P_x}^k - \widehat{P}_x^{\,k}(\alpha_x) \right]^2, \tag{7}$$

which is an analogy (6), only the addition index is now running the final balances of insured persons as of 12/31/2018-2022. It is therefore the square error of the prediction with respect to the real numbers of VZP insured persons in groups (adjusted in 2022 by subtracting citizens of Ukrainian nationality). This option can be initialized in the Excel application under the mode "5 or 6 /= finALFA after 5year's in 'ABDE' or 'ABCDE'/" for automatic recalling such final alpha's found after five years evolution. These final extrapolations are therefore the best which could be achieved by in-sample discounted-GLM method, if we were able to select the best α_x for the prediction in the following five years already in the threshold year $t_0 = 2018$ (but of course this can never be achieved exactly due to the stochastic nature of the modeled processes themselves). In headers of created interactive charts this option is displayed as "finALFA".

Based on the comparison of the square roots of the sums of squares ad (7), which we obtained by substituting the values $\widehat{P}_x^{\,t}(\alpha_x)$ into (7) for α_x corresponding to the choices of *optimum*, *minimum* or *final alpha* in each age group x, we can now objectively assess the quality of the models' prediction after five years. In summary, the models can be evaluated by the square roots of the sums of squares ad (7) over all age groups $x =$ '0-4', ... , '85+'. (Such prediction for all deterministic and in all 18 age groups marginally optimized GLM models 'ABDE' and 'ABCDE' in 2018 is shown in Figure 2, for comparison.) The square roots of the sums of squares ad (7) can be interpreted as formal Euclidean distances of predictions from reality observed in the years 2018–22. In Figures 1–5, they are displayed within the legends for all plotted models after the designation '√SS='. 3.

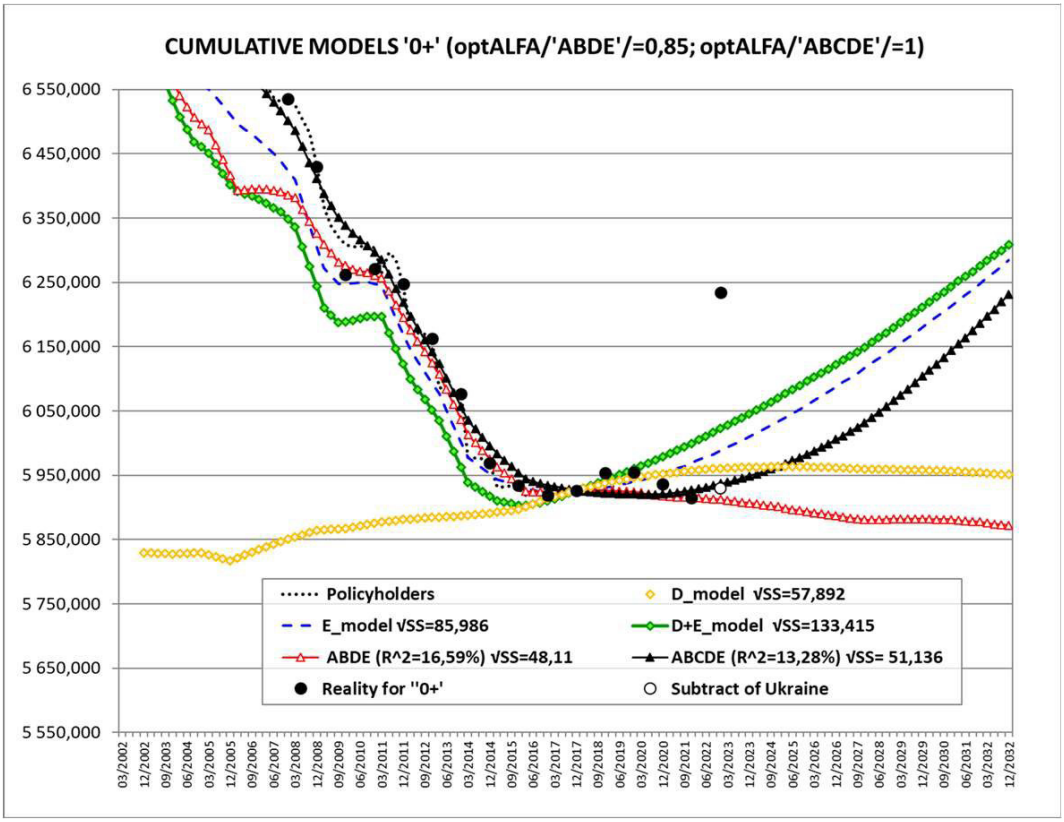


Fig. 1: Quarterly time series of the number of VZP CR insured people 2002-17 in group ‚0+' (estimated as the whole), their prospective and retrospective projections using three deterministic and two discounted regression models calibrated in 2017 in comparison with the real evolution in 2018-22 (see Legend inside graph)

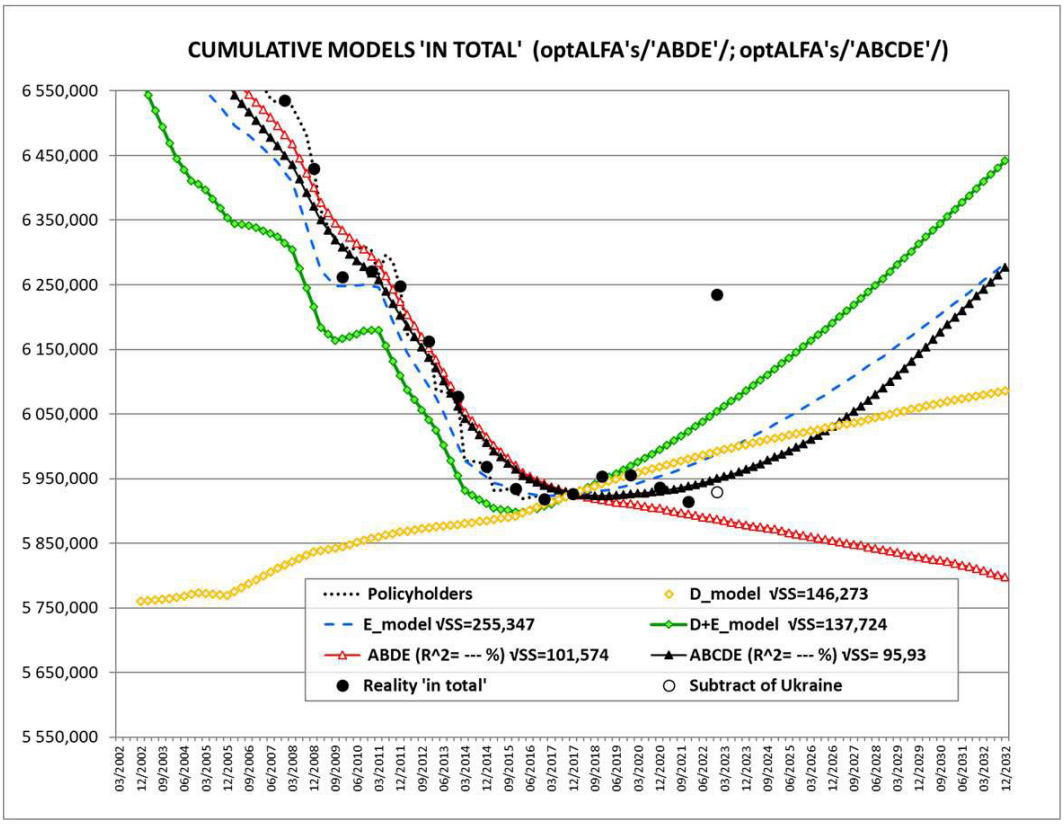


Fig. 2: Quarterly time series of the number of VZP CR insured people from 2002-2017 ‚in total' (sum of age groups), their prospective and retrospective projections using three deterministic and two discounted re-gression models calibrated separately in each group '0-4', ... , '85+' in 2017 in comparison with the real evolution in 2018-22 (see Legend inside graph)

3 RESULTS

The quality of data fitting by extrapolation curves can be assessed by visually comparison the model curves with the actual values among 2018–2022 in Graphs 1–5 marked by circles (the threshold year $t0 = 12/31/2017$ is located in the middle of the time axis of each graph). A more precise numerical evaluation of the quality of fitting for the variants of the models ‚ABCDE' and ‚ABDE' is found in Tables 1–3.

Table 1 takes the values of optimum alpha and their corresponding $R\text{-}2(\alpha x)$ coefficients decisive for the optimization of both generalized models from (Běláček, 2019). With the exception of extrapolations for the group of insured persons as the whole ('0+') and for the group of the youngest ('0-4'), where the coefficients of determination did not exceed 33.3% for any of the discounted linear models, the smoothing ability of the 'ABCDE' and 'ABDE' models is generally high in all other age groups (see columns 3 and 6 in Table 1). A more detailed analysis of the statistical significance of the model parameters showed (across age groups with the exception of '0+') that the demographic component is universally the most significant among the exogenous variables (see in Běláček, J., 2019 for more details). On the contrary, at the level of the 'ABCDE' models optimized in

Table 1, the least significant quadratic term by the time variable (coefficient Cx) appears to be. For most extrapolations, we could seemingly make do with the sub-model 'ABDE', whose linear trend component was statistically significantly different from zero ($p < 0.05$) in spectrum of age ranges '5-19' and '45+' and the re-registration component in the groups '10-14', '40-49' and '55-59'.

Tab. 1 Table of optimum discounts αx calibrated for weighted LSE estimators αx (using (2) at 2002-17 quarterly data in addition with criterions of maximization $R\text{-}2(\alpha x)$ as in part 2.2) for models 'ABDE' and 'ABCDE', their optimum $R\text{-}2(opt.\ \alpha x)$ values and the resulting $\sqrt{SS}2018\text{-}22$ deviations from reality 2018–22 (in thousands of people)

model:	***'ABDE'***	***'ABDE'***	***'ABDE'***	***'ABCDE'***	***'ABCDE'***	***'ABCDE'***
group x	$opt.\ \alpha x$	$R2(opt.\ \alpha x)$	$\sqrt{SS}2018\text{-}22$	$opt.\ \alpha x$	$R2(opt.\ \alpha x)$	$\sqrt{SS}2018\text{-}22$
'0+'	0.85	16.59%	48.110	1.00	13.28%	51.136
'0-4'	1.00	27.44%	32.757	1.00	32.14%	10.084
'5-9'	0.69	52.68%	21.611	0.69	52.68%	21.931
'10-14'	0.99	82.47%	25.009	0.99	82.75%	14.098
'15-19'	0.86	76.50%	7.272	1.00	64.91%	12.379
'20-24'	1.00	39.75%	11.559	1.00	41.69%	27.264
'25-29'	1.00	47.21%	28.359	1.00	49.12%	58.509
'30-34'	1.00	66.04%	10.095	1.00	66.18%	5.586
'35-39'	0.93	72.07%	22.766	0.94	73.45%	9.961
'40-44'	0.99	49.01%	41.759	1.00	49.47%	30.001
'45-49'	1.00	83.03%	24.798	1.00	83.23%	27.327
'50-54'	0.96	71.66%	6.121	0.96	71.77%	1.908
'55-59'	1.00	79.01%	4.867	1.00	79.09%	18.840
'60-64'	1.00	79.93%	10.527	1.00	79.98%	9.127
'65-69'	0.75	85.02%	3.794	1.00	83.65%	5.736
'70-74'	1.00	93.05%	2.448	1.00	93.15%	17.111
'75-79'	0.94	85.73%	15.014	0.94	85.82%	14.237
'80-84'	0.86	36.66%	21.566	1.00	82.29%	17.384
'85+'	1.00	64.78%	57.438	1.00	67.10%	32.429
'in total'	**---**	**---**	**101.574**	**---**	**---**	**95.930**

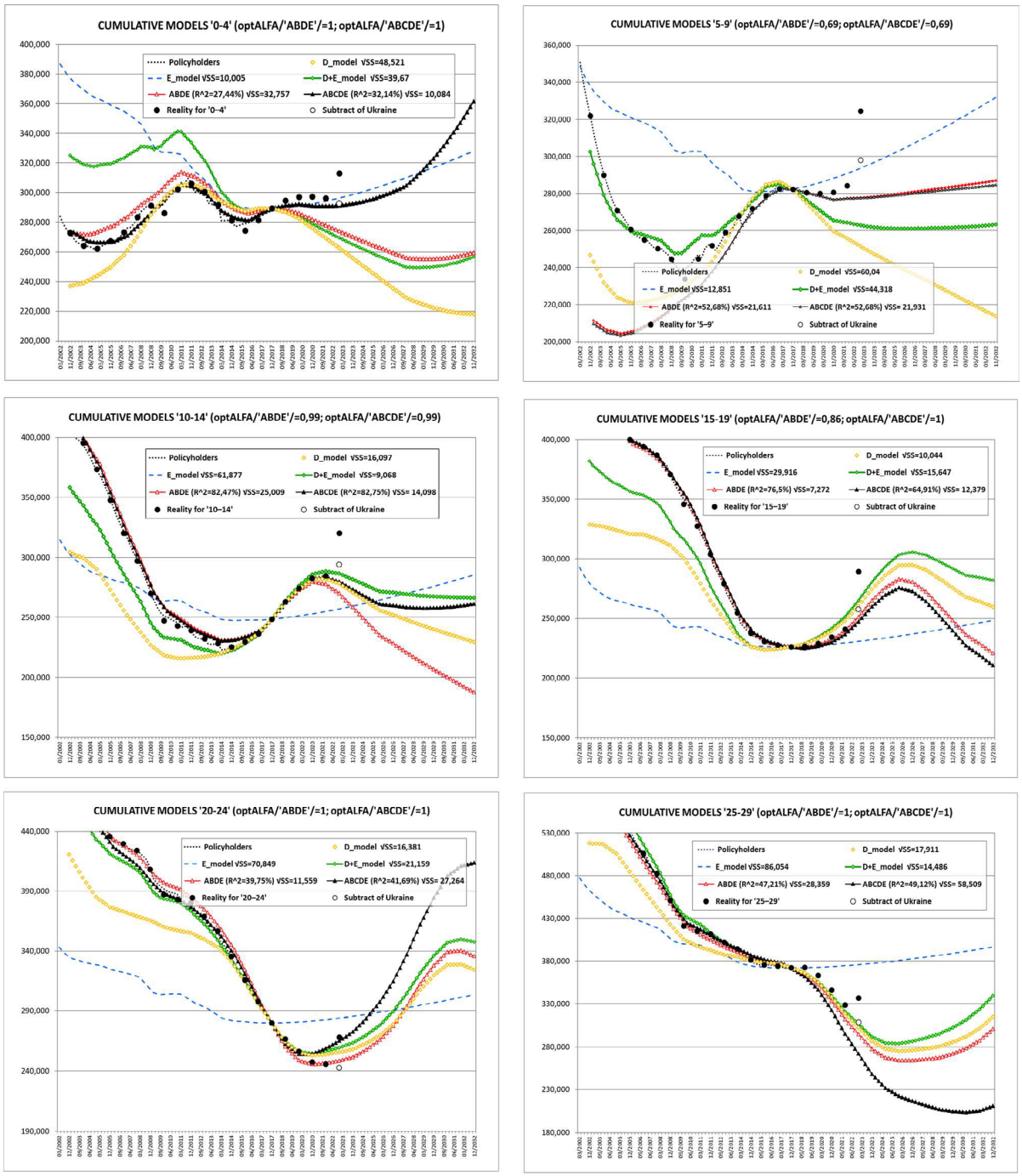


Fig. 3: a-f): Quarterly time series of the number of VZP CR insured people from 2002-17 in groups ‚0-4', ..., ‚25-29', their prospective and retrospective projections using three deterministic and two discounted re-gression models calibrated in 2017 in comparison with the real evolution in 2018-22 (see Legends inside graphs)

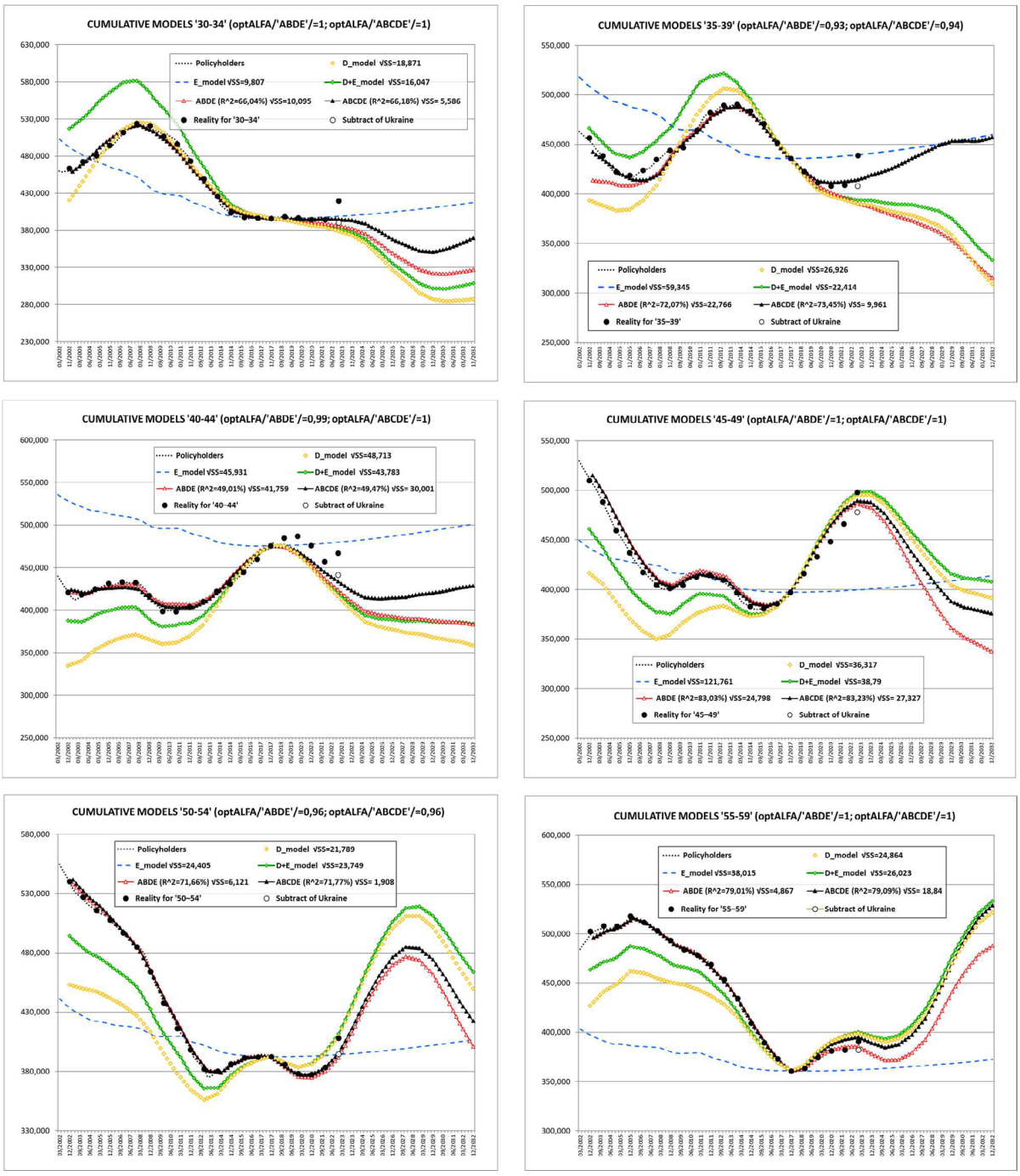


Fig. 4: a-f): Quarterly time series of the number of VZP CR insured people from 2002-17 in groups '30-34', ... , '55-59', their prospective and retrospective projections using three deterministic and two discounted regression models calibrated in 2017 in comparison with the real evolution in 2018-22 (see Legends in-side graphs)

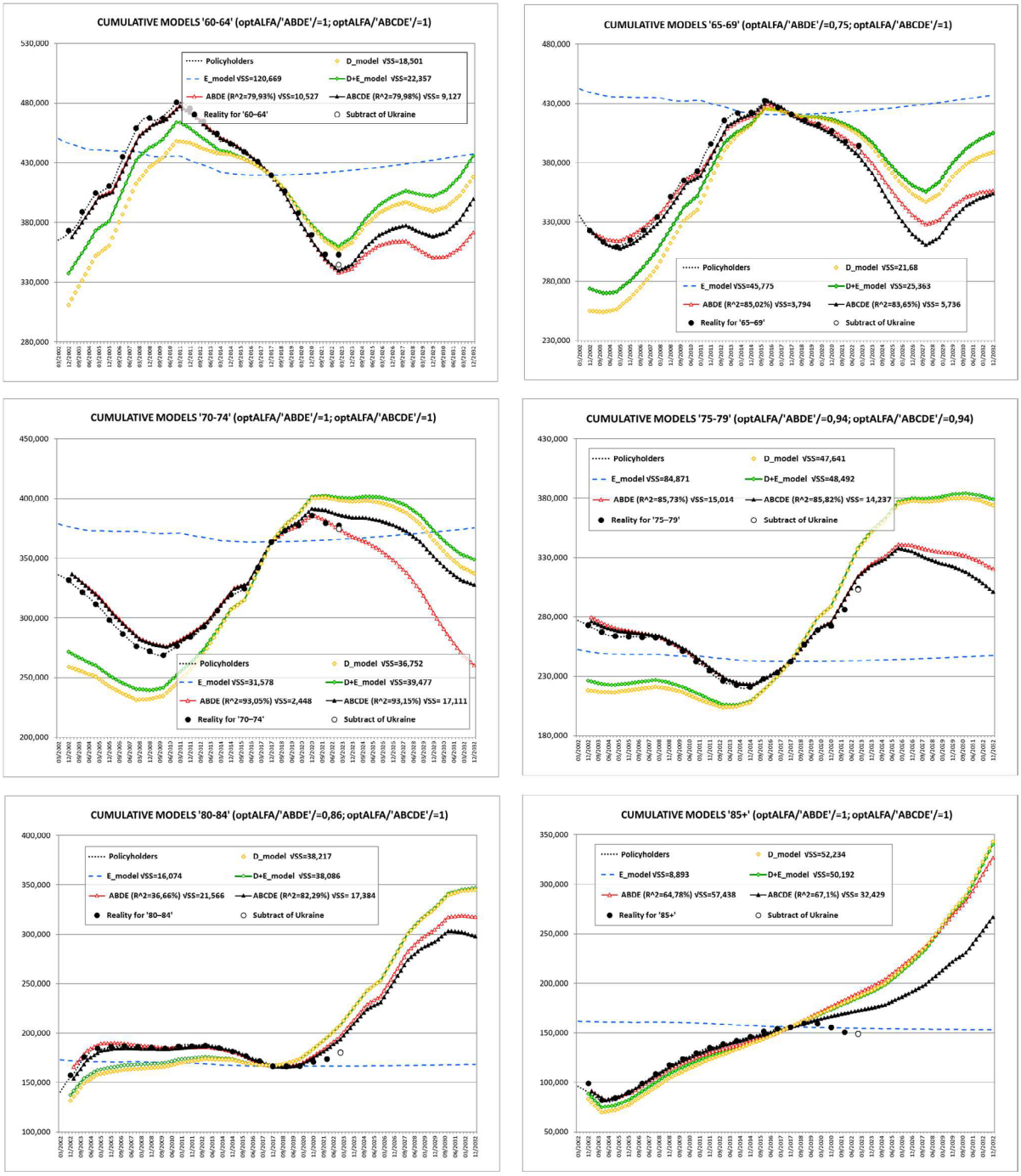


Fig. 5: a-f): Quarterly time series of the number of VZP CR insured people from 2002-17 in groups '60-65', ... , '85+', their prospective and retrospective projections from 2018 using three deterministic and two dis-counted regression models calibrated in 2017 in comparison with the real evolution in 2018-22 (see Leg-ends inside graphs)

However, we should be careful when interpreting these findings, because in some age groups, the explanatory role of formal polynomial terms is taken over by the currently steeper demographic or re-registration component inside some models. Overall, the higher level of statistical significance of the parameters in these models is proportional to the higher values of the coefficients of determination $R\text{-}2(\alpha x)$ shown in the Table 1.

The similarly structured Table 2 points to the significantly different results of the optimization of the 'ABDE' and 'ABCDE' models based on unweighted in-sample minimization in the last five years before the threshold period. It is not so much about how the minimum alpha values from Table 2 specifically differ from the previous optimum ones (in the minimizations ad (6) the choice $\alpha x > 1.00$ was also accepted), nor that the corresponding $R\text{-}2(\alpha x)$ values are generally lower than in Table 1 – this is a natural consequence of the fact that the models in Table 2 are parameterized

Tab. 2 Table of minimum discounts αx calibrated for unweighted LSE estimators αx (using (6) for 2012-16 annual data) for models ‚ABDE' and ‚ABCDE', their corresponding $R\text{-}2(min.\ \alpha x)$ values and resulting $\sqrt{SS}2018\text{-}22$ deviations from reality in 2018-22 (in thousands of people)

model:	***'ABDE'***	***'ABDE'***	***'ABDE'***	***'ABCDE'***	***'ABCDE'***	***'ABCDE'***
group x	$min.\ \alpha x$	$R2(min.\ \alpha x)$	$\sqrt{SS}2018\text{-}22$	$min.\ \alpha x$	$R2(min.\ \alpha x)$	$\sqrt{SS}2018\text{-}22$
'0+'	0.92	14.79%	82.662	0.95	15.12%	59.185
'0-4'	0.88	26.62%	14.618	1.05	43.77%	6.566
'5-9'	0.95	34.95%	72.639	0.90	29.36%	7.858
'10-14'	0.85	62.89%	19.176	0.85	75.50%	124.802
'15-19'	0.71	68.31%	3.414	0.89	76.44%	2.855
'20-24'	0.96	30.51%	7.433	0.98	36.40%	27.487
'25-29'	0.86	18.07%	40.537	0.83	24.81%	94.412
'30-34'	0.83	57.64%	11.822	0.89	60.01%	20.374
'35-39'	0.98	67.54%	21.672	0.92	72.95%	7.424
'40-44'	0.84	9.83%	86.747	0.89	27.96%	111.520
'45-49'	0.80	78.87%	30.436	1.03	85.06%	31.293
'50-54'	0.95	71.37%	5.574	0.96	71.77%	1.908
'55-59'	1.08	76.15%	5.091	0.99	77.73%	18.755
'60-64'	1.01	80.33%	11.323	0.82	17.23%	5.876
'65-69'	1.04	86.43%	23.133	1.04	86.60%	9.199
'70-74'	0.86	82.43%	7.087	1.04	92.28%	21.359
'75-79'	1.02	79.26%	19.329	1.07	75.69%	23.585
'80-84'	0.90	53.70%	26.010	1.00	81.83%	17.384
'85+'	0.61	0.34%	33.827	0.93	26.75%	19.216
'in total'	---	---	139.978	---	---	203.475

based on a different loss function than $R2(.)$. What is now essential is to compare the corresponding values in the columns $\sqrt{SS}$**2018-22** for both models, which measure the formal distances of the model values from reality for each age group. Only for a few age groups did the in-sample optimization method using minimum alpha in the five prethreshold years perform significantly better results than the method originally used in 2018; in most cases, the second in-sample method is worse than the later observed reality. This is documented also by the cumulative deviations $\sqrt{SS}$**2018-22** = 139.978 vs. 101.574 for the 'ABDE' model and $\sqrt{SS}$**2018-22**= 203.475 vs. 95.930 for the 'ABCDE' model, at the expense of the second in-sample variant.

The decisive data for comparing the results of both discount optimization variants for the 'ABDE' and 'ABCDE' models is contained in Table 3 with the data compared for the final discounts αx, which were found ad poste as those minimizing the distances between real data and modeled

Tab. 3 Table of final discounts αx found due to step by step minimization of (7) for models ‚ABDE' and ‚ABCDE', their corresponding $R\text{-}2(fin.\ \alpha x)$ values and resulting $\sqrt{SS}$2018-22 deviations from reality 2018-22 (in thousands of people)

model:	***'ABDE'***	***'ABDE'***	***'ABDE'***	***'ABCDE'***	***'ABCDE'***	***'ABCDE'***
group x	$fin.\ \alpha x$	$R2(fin.\ \alpha x)$	$\sqrt{SS}$2018-22	$fin.\ \alpha x$	$R2(fin.\ \alpha x)$	$\sqrt{SS}$2018-22
'0+'	0.82	16.42%	39.998	0.98	13.57%	49.291
'0-4'	0.76	26.22%	5.395	1.06	46.80%	6.511
'5-9'	0.54	35.52%	10.049	0.80	44.01%	5.456
'10-14'	0.80	56.40%	17.145	0.96	81.54%	7.196
'15-19'	0.66	61.32%	2.097	0.88	66.10%	2.646
'20-24'	0.45	9.73%	3.306	0.55	22.87%	8.736
'25-29'	0.92	33.40%	9.880	0.95	41.50%	12.913
'30-34'	0.97	62.62%	3.904	0.70	38.08%	3.140
'35-39'	0.71	41.05%	1.901	0.90	71.55%	6.380
'40-44'	0.69	14.74%	18.213	0.99	49.99%	20.448
'45-49'	1.01	83.78%	24.792	0.98	81.35%	26.800
'50-54'	0.87	58.66%	1.384	0.96	71.77%	1.908
'55-59'	0.64	55.33%	1.639	0.74	66.56%	7.906
'60-64'	0.90	45.33%	2.318	1.07	73.54%	3.048
'65-69'	0.66	82.68%	1.528	0.94	83.03%	1.779
'70-74'	1.09	84.87%	2.145	0.85	82.01%	2.114
'75-79'	0.95	85.60%	14.950	0.97	85.16%	13.878
'80-84'	1.00	81.83%	1.181	1.04	82.54%	5.173
'85+'	0.74	1.21%	15.627	0.60	2.38%	5.482
'in total'	---	---	44.590	---	---	43.514

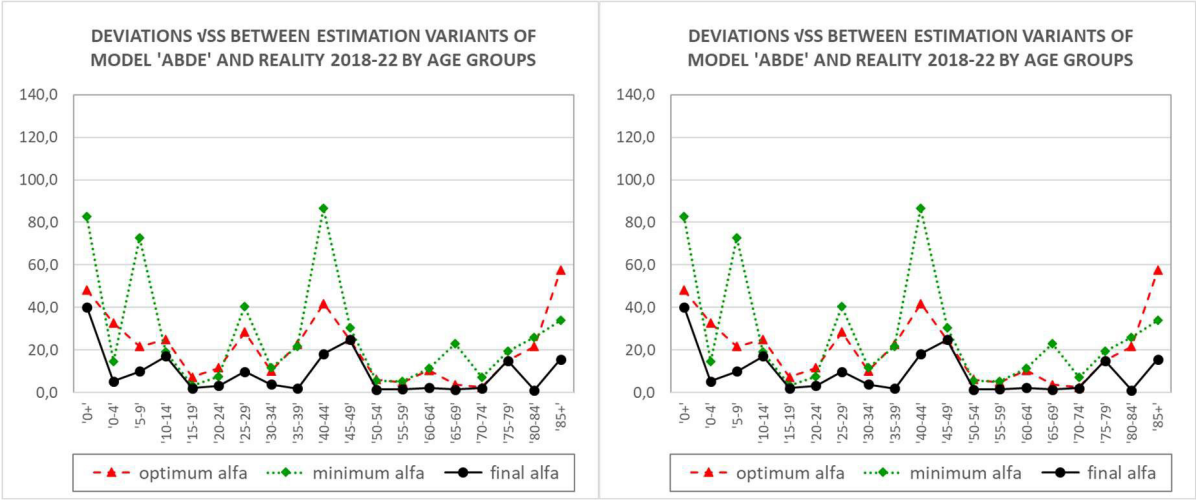


Fig. 6: a-b): Visual comparison of deviations between the real data 2018-22 against three calibration vari-ants (see Legends in the charts) of model 'ABCD' or 'ABCDE', respectively, by age groups

values for the years 2018–22 (compare the 4th and 7th columns of Table 3). This means that there is no such αx in any age group x for which a better fit could be achieved in the given class of generalized-discounted regression models in 2018! In summary, the decisive data from Tables 1–3 are visualized in Figures 6a–b) for both models. The significant deviations from reality are noticeable in some age groups, especially for in-sample variant from Table 2, more clearly for 'ABCDE' model.

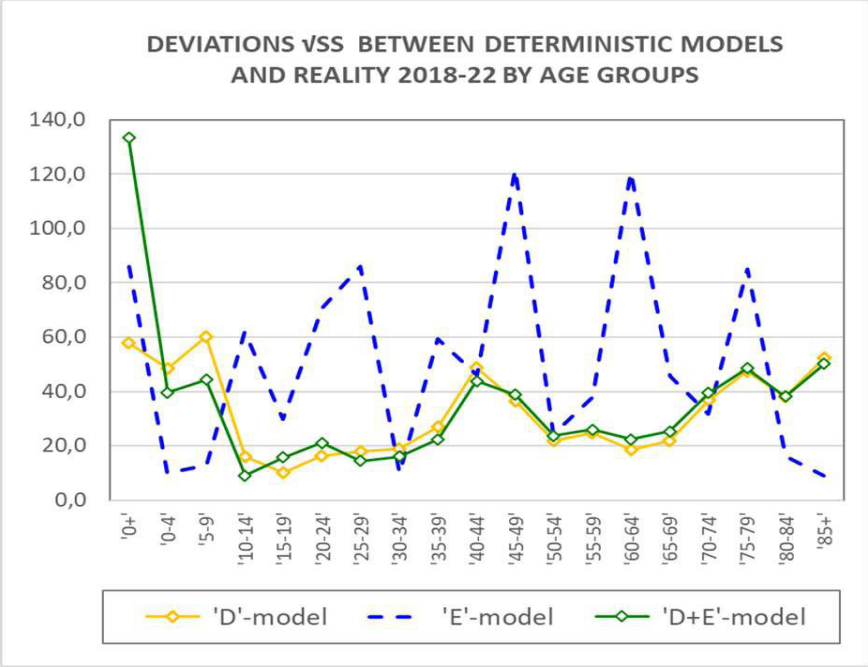


Fig. 7: Visual comparison of deviations between the real data 2018-22 against three deterministic models predictions (see Legend in the charts) by age groups)

For the variant from Table 1, the worst predictions were found in the groups '0+' (models with low smoothing and therefore predictive accuracy), '25-29' (we have found de facto inverse trend after 2017 here), '40-44' (underestimation of reality for both models) and '85+' (overestimation of reality for 'ABDE' model, where the linear trend component no longer had the ability to compensate for the losses of insured persons reflecting the COVID21 pandemic). In some age groups ('85+', '40-44' and '20-24'), we observed significant decreases in $R\text{-}2(\alpha x)$ relative to the 2018 model values with the transition to models with final discounts αx (see comparison of columns 3 and 6 in Tables 1 and 3). However, for most other age groups, the prediction based on optimum alpha's appears to be fairly successful (for some age groups, the predicted curves did not differ much from each other even with significant changes in alfa values).

Finally, we could be interested in comparing the prediction results with deterministic models from

Tab. 4 Table of resulting $\sqrt{SS}2018$- deviations due to the deterministic model predictions toward the reality in 2018–2022 (in thousands of people)

model:	***'D-model'***	***'D-model'***	***'D+E-model'***
group x	$\sqrt{SS}2018\text{-}22$	$\sqrt{SS}2018\text{-}22$	$\sqrt{SS}2018\text{-}22$
'0+'	57.89	85.99	133.42
'0-4'	48.52	10.01	39.67
'5-9'	60.04	12.85	44.32
'10-14'	16.10	61.88	9.07
'15-19'	10.04	29.92	15.65
'20-24'	16.38	70.85	21.16
'25-29'	17.91	86.05	14.49
'30-34'	18.87	9.81	16.05
'35-39'	26.93	59.35	22.41
'40-44'	48.71	45.93	43.78
'45-49'	36.32	121.76	38.79
'50-54'	21.79	24.41	23.75
'55-59'	24.86	38.02	26.02
'60-64'	18.50	120.67	22.36
'65-69'	21.68	45.78	25.36
'70-74'	36.75	31.58	39.48
'75-79'	47.64	84.87	48.49
'80-84'	38.22	16.07	38.09
'85+'	52.23	8.89	50.19
'in total'	146.273	255.347	137.724

(1). This purpose is served by Table 4, where the formal Euclidean distances of the actual numbers of VZP CR insured persons as of 12/31/2018-2022 from the numbers predicted by the derived population projection ('D-model'), the middle variant of extrapolation of the VZP re-registration balance ('E-model') and the combination of both methods ('D+E-model') are available. Visualization of these distances in Figure 7 shows, that the applicability of the re-registration balance itself holds sporadically only in some age groups ('0–9', '30–34' and '85+'). The good predictive ability of both deterministic models with demo-projection suits in some younger age groups ('10–34'). Here, these deterministic models were competitive with predictions based on our best generalized regression models 'ABDE' and 'ABCDE', respectively. However, as the summary sums in the last row of Table 3 show, the deterministic models cannot be used universally in all age groups with better results. The indicators described above identify approximately those age groups where demographic or re-registration components were statistically most significant.

4 SUMMARY OF THE MAIN EMPIRICAL RESULTS

The main difficulty of the model developed in 2016 at VZP was the no transparent specification of the assignment from the beginning. If we were interested in predicting the future number of insured persons for the next few years, it would be sufficient to base ourselves on one-year time series from historical VZP yearbooks. However, relatively early after starting works on the model construction, a requirement emerged for predicting quarterly or at least semiannual balance sheets of insured persons, which may be important for the preparation of the so-called health insurance plans approved by VZP once a year (see Health Insurance Plans – VZP ČR, 2025). However, this requirement entailed the need to use data from VZP's internal databases and interventions to conceptualization of the entire prediction model, e.g. adjustment of the model with the aim of taking the seasonal component into account.

The specification of the projection horizon is another problem in itself. The pragmatic need for one-year plans of an insurance company can still be met relatively well by the method of simple indexing, where we compare the annual growth/decline indices in groups with a possible expert correction to the expected trends (common practice in VZP). But from the strategic point of view we should assume a horizon of at least 5 or even more years? The five-year period makes it possible to use the population forecast of the Czech Statistical Office for the Czech Republic available in five-year age groups. For situation, when VZP had approximately 55% share of the public health insurance market in the Czech Republic, this application has proven to be successful.

The natural inclusion of demographic and re-registration variables into models means a departure from methods of simple indexing and from a simple application of derived population projection, where predicted curves are simply prolonged by the results of the most current demo forecasts to the threshold year of prediction. In the applications used above, the population forecast and the extrapolated re-registration component became part of more complex regression models, where in particular the linear trend term in the time variable also takes into account the trends of the historical development of the insured population of VZP, which had been taken from consistent time series since 2002. However, the qualification of this trend to the future proved to be not easy to estimate inside the threshold years 2016–17, because at that time the total decrease in the number of insured persons in VZP began to visibly stagnate and it was not clear in which direction the current trends would further go on (see large fulfilled black circles between 2017–2022 on Figures 1–2 for an overall overview). In combination with carefully performed discounting within

five-year groups, it was mainly the population projection variable that made it possible to specify the realistic framework of the trends of the optimized extrapolation curves (see ‚D-model' curves in Figures 3–5 for comparison with reality).

Of the two most important general regression models considered in this article, the 'ABDE' model, that one with only a linear term in the time variable, appears to be the more efficient. At comparable levels of the $R\text{-}2(\alpha x)$ coefficients shown in Table 1, it has the same or higher number of statistically significant parameters in most age groups than the 'ABCDE' model, that one with additional parabolic trend term! In terms of the best fit to historical data, as is seemed from the comparison of $R\text{-}2(\alpha x)$ and $\sqrt{\boldsymbol{SS}}\mathbf{2018\text{-}22}$ values in Table 3, the entire 'ABCDE' model also demonstrated serious smoothing ability in some age groups. However, for the conventional in-sample method in the last five pre-threshold years, it turned out (see the last column of Table 2 in the groups '10–14' and '40–44', for example) that only the formal application of the full model (with a non-zero parabolic term by the time variable) can be misleading for the purposes of realistic prediction of the number of insured persons. However, this is a known undesirable property of prediction models with an uncorrected quadratic term by the time variable when applied to a longer time horizon.

The model for predicting the number of health insured people in five-year groups can be easy recalculated for both sexes. However, when applying the 2018 methodology to special groups of health insurance payers (employees, self-employed, other payers), the models used would have to be calibrated completely separately. Then, high-quality regression models could also serve as a basis for creating annual health plans of the insurance company, if they were calibrated and recalculated at least once a year based on consistently supplemented internal data. Then, we would record more or less continuous changes in the parameters of the fitting models, including their discounts αx , and in groups with lower achieved levels of $R\text{-}2(\alpha x)$, we could make more predictable corrections. And if we had systematic data on evolution of average incomes and average payments for the relevant groups of health insurance payers, we could also bring calculations of future total incomes and expenditures of the health insurance company under greater control in the medium horizon (say, for the next 3–5 years). Unexpected sudden changes in the number of insured persons (as was the case from 2022, for example, in connection with immigrants from Ukraine) can be easily corrected in these models at the level of input demographic or re-registration exogenous regressors.

5 CONCLUSIONS AND RECOMMENDATION

The prediction method of discounted linear regression models, when the calibration of discounts α is based on the maximum values of the coefficient of determination $R2(\alpha)$, can be universally recommended based on the above experiences. The prerequisite for success is of course a sufficiently long history of the predicted time series (although when using discounts $\alpha \ll 1$ we could get by with an order of magnitude shorter). Furthermore, we must have a representative regression model with $R2(\alpha)$, values not lower than (say) 35% and reasonably predictable time series of exogenous regression variables (that is, if we are not limited only to formal polynomial trends as in standard exponential smoothing).

If one of the significant explanatory components of the model is a demographic variable, population forecasts or projections can be used in justified cases. The auxiliary demo-prognosis can be effectively used even if it later turns out to be overestimated or underestimated compared to reality – the estimated level of the corresponding regression coefficient is optimized in the context

of a multiparametric regression model, which itself corrects trend inaccuracies in the demographic component. In any case, this approach is more universal than performing a formally simple derived population projection, which implicitly assumes that there is a constant trend in the data. However, the methodology used is unsuitable for time series located near the threshold year in the phase of a stagnant trend (as was the case in 2017 in aggregated age group ‚0+'). A constant trend in the time series excludes the possibility of using a non-constant explanatory variable for smoothing the model, and therefore the values of discounted $R2(\alpha)$, which should be used when optimizing more complex regression models, will themselves be close to zero and therefore will not lead to a more effective parameterization of the prediction model.

The parameter estimation methodology based on a linear difference model of type (2), which, thanks to the additional calculation ad (5), allows placing the predicted curve directly into the last known value of the predicted time series in the threshold year, can also be recommended due to the efficiency of fitting the extrapolation curve – model curves fitted directly into primary data do not have this property. Parameter estimates in the difference model according to formula (4) are essentially just a trivial modification of the standard unweighted least squares method, when we multiply the experimental design matrix and the column of the explained regression variable in the basic model by the same weighting factors $wt = \alpha x\text{-}t/2$ $(t = 1, \ldots, t0)$ and perform the additional calculations described in paragraph 2.2. This simple algorithm works universally for any system of linear regression equations and is an alternation of formulas derived explicitly for exponential smoothing of polynomial regressions – given e.g. in (Cipra, 1986, pp. 57–72) or in (Exponential Smoothing, 2025). In contrast to those references, formula ad (4) allows to estimate parameters also for the choice of discount rates $\alpha > 1$, which may be also a justified choice in specific cases, as shown above.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

BĚLOHRADSKÝ, A., ŠOLC, Z. 2018. Predikce příjmů veřejného zdravotního pojištění. In: Metodické *compendium. Praha, Ministerst*vo financí ČR. https://www.mfcr.cz/cs/ministerstvo/kariera-a-vzdelavani/vzdelavani/odborne-studie-a-vyzkumy/2018/predikce-prijmu-verejneho-zdravotniho-po-33111

BĚLÁČEK, J. 2018. Predikce budou*cích počtů pojištěnců VZP ČR – d*ata, metodika a výsledky. In: Sborník příspěvků MEDSOFT 2018, pp. 7-19. Hotel Academic, Roztoky u Prahy, 3/20-21/2018. ht*tps://doi.org/10.35191/medsoft_2018_1_30_5_11*

BĚLÁČEK, J. 2019. Regresní model pro predikci budoucích počtů pojištěnců VZP ČR. In: Sborník příspěvků MEDSOFT 2019, pp. 4-8. Hotel Academic, Roztoky u Prahy, 3/19-20/2019. https://doi.org/10.35191/medsoft_2019_1_31_4_8

BĚLÁČEK, J. 2023. Predikce budoucích počtů pojištěnců VZP *ČR – ukázka modelů a* porovnání výsledků. Paper presented on ČDS conference 2023, Hradec Králové, 5/26/2023. https://www.czechdemography.cz/akce/konference/konference-cds-2023/prispevky-z-konference/

CIPRA, T. 1986. *Analýza časových řad s aplikacemi v ekonomii*. Praha, SNTL/Alfa.

CZECH STATISTICAL OFFICE. 2013. *Population projection of the Czech Republic until* 2100 [in Czech]. Czech statistical office. https://www.czso.cz/csu/czso/projekce-obyvatelstva-ceske-republiky-do-roku-21*00-n-fu4s64b8h4*

CZECH STATISTICAL OFFICE. 2014. Population projection in the regions of the Czech Republic - up to the 2050 [in Czech]. Czech statistical office. https://www.c*zso.cz/csu/czso/projekce*-obyvatelstva-v-krajich-cr-do-roku-2050-ua08v25hx9

CZECH STATISTICAL OFFICE. 2023. Demographic Yearbook of the Czech Republic – 2022. Czech statistical office. https://csu.gov.cz/produkty/demographic-yearbook-of-the-*czech-republic-2022*

*CZE*CH STATISTICAL OFFICE. 2026. Age Distribution of the Population - 2012, ..., 2016. 2017. Czech statistical *office. https://csu.gov.cz*/produkty/age-distribution-of-the-population-2016

VZP ČR. 2023. VZP ČR Yearbook *for 2018-22 [in Czech]. www.vzp.cz/o-nas/dokumenty/rocen*ky

VZP ČR. 2025. Health insurance plans. [in Czech] https://www.vzp.cz/o-nas/dokumenty/zdravotne-pojistne-plany

WIKIPEDIA. 2025. Exponential smoothing. https://en.wikipedia.org/wiki/Exponen-tial_smoothing

Electronic documentation:

Excel application of a predictive model. 2025. Available at https://zenodo.org/uploads/16986175

Contact information

Jaromír Běláček: e-mail: belacek@mendelu.cz

OPEN-SOURCE DECEPTION SECURITY FOR AWS CLOUD ENVIRONMENTS: A COST-CONSCIOUS ARCHITECTURE FOR SMES

Matěj Cajthaml[1], Tomáš Pitner[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_004

ABSTRACT

This paper presents the design and verification of an open-source deception security solution for Amazon Web Services cloud environments with a focus on small and medium-sized businesses and other budget-constrained organizational units. The study starts from the practical problem that commercial deception platforms provide centralized management, automated deployment, alerting, and visualization, yet their licensing and operational costs are often difficult to justify for smaller organizations. At the same time, cloud adoption continues to grow and expands the attack surface exposed through identities, storage services, virtual machines, and application interfaces. The proposed approach combines infrastructure as code, deceptive assets, and a lightweight monitoring pipeline to provide early warning signals without requiring a costly proprietary stack. The architecture integrates decoy cloud resources, identity-based honeytokens, event filtering, message queuing, log forwarding, and visualization components. The implementation was validated through a simulated insider-oriented attack path mapped to selected MITRE ATT&CK techniques. The results indicate that a reproducible and operationally realistic deception layer can be built from open-source tools and native cloud services while preserving constrained permissions and manageable running costs. The paper concludes that open-source deception can serve as a practical complementary layer in defense-in-depth strategies for cloud environments used by smaller organizations.

Keywords: deception security, cloud security, Amazon Web Services, honeytokens, honeypots, open-source security, threat detection, SME
JEL Code: O33, M15, L86



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Cloud computing has become a standard delivery model due to its scalability, flexibility, and reduced infrastructure costs (Mell and Grance, 2011; Qian *et al.*, 2009). Cloud adoption is increasing across organizations of all sizes, including small and medium-sized businesses (Eurostat, 2026a; Flexera Software, 2025). However, public cloud environments also introduce new security responsibilities. While providers secure the underlying infrastructure, customers remain responsible for identity management, access control, configuration, and data protection (Amazon Web Services; National Cyber Security Centre, 2022; Narayanasamy, 2025). For SMEs, these responsibilities are complicated by limited budgets, smaller security teams, and lack of operational capacity (ENISA, 2021; Kolli, 2025).

Cloud environments are exposed to risks such as misconfiguration, unauthorized access, identity abuse, insecure interfaces, and insufficient monitoring (Aljumah and Ahanger, 2020; Khoda Parast *et al.*, 2022; CSA, 2024; ENISA, 2025). These risks are amplified by multi-cloud architectures, cloud-native services, and shadow IT (Saxena, Gupta and Kumar Singh, 2021; Martseniuk *et al.*, 2024). Conventional security tools, including SIEM platforms, logging systems, and vulnerability scanners, are necessary but largely reactive because they often detect incidents only after attacker interaction with production assets has already occurred (Kratzke, 2022; Ban *et al.*, 2021).

Deception security approaches this problem differently by deploying deceptive assets such as honeypots, honeytokens, or decoy credentials that are not used during legitimate operations. Any interaction with such assets is therefore suspicious by design (Han, Kheir and Balzarotti, 2018; Dickinson, 2020; Javadpour *et al.*, 2024). These techniques can improve detection speed, provide contextual evidence of attacker behavior, and support earlier response during reconnaissance or lateral movement phases (Hutchins, Cloppert and Amin, 2011; The MITRE Corporation, [b.r.]). Recent studies also demonstrate the feasibility of lightweight deception mechanisms based on breadcrumbs and stealthy fingerprinting techniques (Msaad *et al.*, 2022; Awasthi, 2025).

Cloud platforms provide favorable conditions for deception-based monitoring because activities involving credentials, APIs, storage, and infrastructure services generate auditable events. In AWS, CloudTrail records management and API activity that can be integrated into monitoring pipelines (Amazon Web Services, [b.r.] a). Combined with open-source observability tools and log collectors (Udovicic, 2022; Vector, [b.r.]), these services allow organizations to build centralized event-processing workflows without extensive infrastructure overhead. However, logging visibility is not always complete, and attackers may attempt to avoid or minimize telemetry generation (Gietzen, 2018).

Commercial deception platforms provide advanced capabilities but are often financially inaccessible for smaller organizations (ENISA, 2021; Verizon, 2025). This paper therefore examines whether a practical deception-security layer for AWS can be implemented using open-source technologies and native cloud services. The objective is to design and evaluate a modular and reproducible solution capable of centralized event collection, automated deployment, and detection of interactions with deceptive assets, while remaining economically viable for smaller organizations.

2 BACKGROUND AND RELATED WORK

Deception security is a defensive approach based on deploying deceptive assets such as honeypots, honeytokens, decoy credentials, or cloud identities and monitoring interactions with them (Han, Kheir and Balzarotti, 2018; Dickinson, 2020; Javadpour *et al.*, 2024). Because these artifacts are not intended for legitimate use, any interaction with them can provide a high-confidence indicator of malicious activity.

Deception should not replace conventional security controls but complement them within a defense-in-depth strategy. Logging, identity governance, segmentation, and incident response remain necessary, while deception increases attacker visibility during reconnaissance, credential access, or lateral movement phases (Hutchins, Cloppert and Amin, 2011; The MITRE Corporation, [b.r.]). Recent work also highlights lightweight techniques such as breadcrumbs and stealthy honeytokens designed to reduce detectability by attackers (Msaad *et al.*, 2022; Awasthi, 2025).

Cloud environments are particularly suitable for deception-based monitoring because interactions with APIs, credentials, storage services, and infrastructure resources generate native telemetry. In AWS, CloudTrail provides audit visibility into management and API activity (Amazon Web Services, [b.r.] a), while open-source observability tools can forward and process these events in centralized analytics pipelines (Udovicic, 2022; Vector, [b.r.]). However, deceptive assets must remain credible and operationally safe, since poorly designed decoys can be easily fingerprinted or may introduce unnecessary risks.

Although the open-source ecosystem offers many individual deception-related tools, there is no widely adopted lightweight framework focused specifically on cloud-centric SMB deployments. This work therefore focuses on a practical subset of capabilities: deployment of cloud decoys, insertion of identity-based honeytokens, centralized event collection through AWS-native services, and integration with an open-source analytics stack. The objective is not to replicate enterprise deception platforms completely, but to evaluate whether an operationally useful and economically accessible solution can be built from open-source components.

3 METHODOLOGY

The study followed a design-and-build research approach focused on implementing and validating an open-source deception-security architecture for AWS environments. The initial phase analyzed deception techniques, cloud-security requirements, and architectural patterns commonly associated with commercial deception platforms (Han, Kheir and Balzarotti, 2018; Dickinson, 2020; Javadpour *et al.*, 2024). The objective was to identify capabilities realistically achievable through open-source components and native AWS services.

The proposed architecture combined multiple deception mechanisms, including honeypot virtual machines, decoy cloud resources, and IAM-based honeytokens. Event collection relied on AWS-native telemetry, particularly CloudTrail logging (Amazon Web Services, [b.r.] a), while centralized processing and visualization were implemented using open-source observability components (Udovicic, 2022; Vector, [b.r.]). Infrastructure provisioning was automated through Terraform, and the analytics environment was deployed using Docker Compose to ensure reproducibility and simplified redeployment.

The design also reflected operational constraints typical for SMB environments, including cost sensitivity, limited staffing, and the need for understandable deployment procedures (ENISA, 2021; Verizon, 2025). Additional emphasis was placed on minimizing permissions assigned to deceptive assets and on reducing the risk of easy attacker fingerprinting (Msaad *et al.*, 2022).

Validation was performed through a controlled attacker scenario executed exclusively within a dedicated AWS environment and aligned with AWS penetration-testing policies. The scenario simulated selected attacker actions involving credential access, resource discovery, and interaction with deceptive assets. Individual activities were mapped to relevant MITRE ATT&CK cloud techniques (The MITRE Corporation, [b.r.]) to verify whether the proposed solution generated meaningful and interpretable detection events.

Tab. 1 provides a condensed view of the design requirements that guided the implementation.

Requirement area	Interpretation for the prototype
Deceptive assets	At least one host-like element and several cloud-native artifacts had to be deployed so that the solution would not depend on a single detection method.
Identity-based detection	The design had to include honeytoken credentials suitable for immediate alerting and for controlled pivot simulation.
Operational realism	Artifacts had to use plausible naming, contextual placement, and seeded content that reflects everyday cloud administration patterns.
Collection and analysis	Events had to be captured through native AWS telemetry, transferred asynchronously, and indexed in an open-source analytical tool.
Reproducibility	The environment had to be deployable with limited manual intervention and documented well enough for reuse by third parties.
Risk control	Compromise of deceptive credentials or hosts had to remain constrained through least privilege and separation from legitimate assets.

4 PROPOSED ARCHITECTURE

The proposed architecture combines AWS-native services with a locally hosted analytics layer. This hybrid approach reduces operational costs while preserving access to cloud-native telemetry and centralized event analysis. The design follows principles commonly associated with cloud-security monitoring, observability, and defense-in-depth architectures (Amazon Web Services, 2024; Kratzke, 2022; CSA, 2025).

The cloud layer contains both supporting infrastructure and deceptive assets. Supporting services include CloudTrail logging, EventBridge event routing, SQS message queuing, and AWS Secrets Manager. Deceptive components include an EC2 honeypot instance, an S3 decoy bucket, a DynamoDB decoy resource, and IAM honeytoken identities with intentionally constrained permissions. The design reflects a realistic attacker path in which deceptive credentials discovered in cloud storage guide an attacker toward additional monitored assets.

The detection pipeline is event driven. CloudTrail records management and selected data events related to deceptive resources (Amazon Web Services, [b.r.] a). EventBridge filters relevant interactions and forwards them through SQS to a local observability stack, where Vector normalizes events before indexing in OpenSearch (Udovicic, 2022; Vector, [b.r.]). This separation reduces direct exposure of the analytics environment while maintaining centralized monitoring capabilities.

Infrastructure deployment is automated using Terraform and Docker Compose to ensure reproducibility and simplified redeployment. During provisioning, deceptive credentials are generated and automatically seeded into selected locations within the honeypot environment to increase realism and reduce manual configuration effort. Additional precautions were taken to minimize privileges assigned to deceptive identities and to reduce the risk of trivial fingerprinting by attackers (Msaad *et al.*, 2022).

5 IMPLEMENTATION

The implementation emphasizes reproducibility, modularity, and operational safety. AWS infrastructure was provisioned using Terraform, while the local analytics environment was deployed through Docker Compose. This combination provides portable deployment workflows and simplifies replication in other environments.

The Terraform configuration was divided into modules for networking, deception assets, and detection services. Deceptive components included an EC2 honeypot, an S3 decoy bucket, a DynamoDB decoy resource, and IAM honeytoken identities with intentionally restricted permissions. Supporting services included CloudTrail, EventBridge, SQS, and AWS Secrets Manager. The permission model followed least-privilege principles in order to minimize the operational impact of compromise (National Institute of Standards and Technology, 2020).

The EC2 honeypot was configured automatically during instance initialization through user_data scripts. These scripts retrieved honeytoken credentials from Secrets Manager and placed them into realistic locations such as AWS credential files and shell history entries. The objective was not to create a fully interactive research honeypot but a believable operational decoy capable of generating useful telemetry while remaining manageable in SMB-oriented deployments.

The detection pipeline relied on CloudTrail management and selected data events associated with deceptive resources (Amazon Web Services, [b.r.] a). EventBridge filtered relevant events and forwarded them through SQS into the local observability stack. Vector normalized and transformed the collected telemetry before indexing in OpenSearch (Udovicic, 2022; Vector, [b.r.]). OpenSearch dashboards then enabled chronological analysis of attacker interactions involving credentials, storage access, and host-level activity.

Several implementation constraints were identified during development. Terraform stores generated IAM access keys in state files, requiring careful handling of deployment artifacts and exclusion of state data from version control. In addition, maintaining realism while limiting operational complexity proved to be a significant trade-off. Expanding the number of decoys would increase coverage but also administrative overhead and the risk of configuration drift. The resulting implementation therefore prioritizes a smaller set of coherent deception elements that remain practical for resource-constrained SMB environments.

6 VERIFICATION AND RESULTS

Verification was based on a controlled insider-style attack scenario executed within a dedicated AWS environment. The scenario simulated reconnaissance, access to deceptive cloud resources, use of honeytoken credentials, and controlled lateral movement toward the EC2 honeypot. Individual activities were aligned with relevant MITRE ATT&CK cloud techniques to evaluate whether the architecture generated meaningful detection events (The MITRE Corporation, [b.r.]).

The attack sequence began with access to the decoy S3 bucket containing operationally plausible configuration data and credentials for the pivot honeytoken account. Use of these credentials generated CloudTrail management events, after which the attacker attempted Session Manager access to the EC2 honeypot. Additional traces inside the honeypot guided the scenario toward interaction with the decoy DynamoDB resource. This chained structure allowed verification across multiple deception layers instead of relying on a single isolated alert.

Tab. 2 Table 2 summarizes the key evaluation dimensions of the prototype.

Evaluation area	Observed outcome
Deployment reproducibility	Infrastructure and local services were repeatedly deployable through Terraform and Docker Compose with limited manual intervention.
Detection coverage	Interactions with deceptive identities, S3 objects, Session Manager access, and selected decoy resources were captured through the configured event path.
Operational realism	Seeded files, naming conventions, and constrained permissions created a plausible attacker path without exposing legitimate assets.
Alert interpretability	The collected events could be correlated in OpenSearch into a chronological sequence suitable for investigation.
Cost and accessibility	The hybrid design reduced dependence on continuously running cloud analytics resources and remained compatible with SMB constraints.

The verification confirmed that the detection pipeline operated as intended. CloudTrail captured management and selected data events associated with deceptive identities and resources (Amazon Web Services, [b.r.] a). EventBridge correctly filtered and forwarded relevant events through SQS, while Vector and OpenSearch enabled centralized processing and timeline visualization of attacker activity (Udovicic, 2022; Vector, [b.r.]). The architecture therefore demonstrated the practical feasibility of combining AWS-native telemetry with open-source observability tooling.

The results also highlighted the value of constrained deception identities. The blank honeytoken account immediately indicated suspicious credential use, while the pivot account provided richer context by linking credential access with attempted host interaction. Restricting permissions to the honeypot instance minimized operational risk while preserving realistic attacker behavior, consistent with least-privilege security principles (National Institute of Standards and Technology, 2020).

Overall, the evaluation demonstrated that open-source and cloud-native components can form a coherent and operationally useful deception workflow for SMB-oriented cloud environments. Although the prototype does not attempt to replicate enterprise deception platforms completely, it provides meaningful attacker visibility with substantially lower operational and financial requirements.

The modular structure of the solution, separation between cloud and local components, and use of widely adopted open-source tools make the architecture reproducible and suitable for experimentation or pilot deployment in SMB-oriented environments.

7 DISCUSSION

The results indicate that open-source deception in AWS is technically feasible and operationally useful for SMB-oriented environments. The main value of the approach lies in improving visibility during reconnaissance, credential misuse, and limited lateral movement rather than replacing conventional preventive controls.

The prototype also highlights several limitations. Deception mechanisms depend on attacker interaction with deceptive assets and therefore cannot guarantee detection in every scenario. In addition, long-term deployments require continuous maintenance to preserve realism and reduce the risk of fingerprinting. Expanding the number and complexity of decoys could improve coverage but would also increase operational overhead.

Compared with commercial deception platforms, the proposed solution provides a narrower feature set and reduced automation. However, it significantly lowers deployment costs while preserving core capabilities such as centralized event collection, telemetry correlation, and high-confidence alert generation. For smaller organizations, this trade-off may be acceptable when the alternative is the absence of deception capabilities entirely.

Although the implementation is AWS-specific, the broader design principles remain transferable to other public-cloud environments. Infrastructure as code, event-driven telemetry collection, and modular deceptive assets could be adapted to alternative cloud platforms with equivalent monitoring and routing services.

8 CONCLUSION

This paper presented an open-source deception-security architecture for AWS cloud environments designed for SMB-oriented deployments. The proposed solution combines deceptive cloud resources, constrained IAM honeytokens, AWS-native telemetry, and an open-source analytics stack into a reproducible monitoring workflow.

The verification scenario demonstrated that the architecture could detect and contextualize attacker interaction with deceptive storage, credentials, and host resources while maintaining limited operational complexity and cost. Although the prototype does not replace traditional cloud-security controls, it provides an additional high-confidence detection layer that can improve attacker visibility in cloud environments.

Future work should focus on stronger anti-fingerprinting techniques, broader automation, and adaptation to additional cloud providers. Even in its current form, the results suggest that open-source deception can represent a practical defense-in-depth enhancement for organizations operating under limited security budgets.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ALJUMAH, Abdullah, AHANGER, Tariq Ahamed. 2020. Cyber security threats, challen-ges and defence mechanisms in cloud computing. Online. *IET Communications*. 14(7), 1185–1191. ISSN 1751-8628. https://doi.org/10.1049/iet-com.2019.0040

AMAZON WEB SERVICES. 2024. Security Pillar: AWS Well-Architected Framework. *Amazon Web Services*. [Accessed: 2026-02-22] https://docs.aws.amazon.com/wellarchitected/latest/security-pillar/welcome.html

AMAZON WEB SERVICES, [n.d.] a. AWS CloudTrail features. *Amazon Web Services*. [Accessed: 2026-02-28] https://aws.amazon.com/cloudtrail/features/.

AMAZON WEB SERVICES, [n.d.] b. Shared Responsibility Model. *Amazon*. Amazon Web Services, Inc. [Accessed: 2026-02-17] https://aws.amazon.com/compliance/shared-responsibility-model/

AWASTHI, Kriti. 2025. Optimizing Deception Breadcrumbs for Endpoint Security Effectiveness. *Fidelis Security*. [Accessed: 2026-03-18] https://fidelissecurity.com/threatgeek/deception/deception-breadcrumbs-for-endpoint-security/.

BAN, Tao; SAMUEL, Ndichu; TAKAHASHI, Takeshi; INOUE, Daisuke. 2021. *Combat Security Alert Fatigue with AI-Assisted Techniques*. Association for Computing Machinery. ISBN 9781450390651. https://doi.org/10.1145/3474718.3474723

CSA. 2024. Top Threats to Cloud Computing 2024. *Cloud security alliance*. [Accessed: 2026-03-04] https://cloudsecurityalliance.org/artifacts/top-threats-to-cloud-computing-2024

CSA. 2025. Security Guidance for Critical Areas of Focus in Cloud Computing v5. *Cloud security alliance*. [Accessed: 2026-03-04] https://cloudsecurityalliance.org/artifacts/security-guidance-v5

DICKINSON, Kyle. 2020. *Implementer's Guide to Deception Technologies*. SANS INSTITUTE. [Accessed: 2026-03-09] https://www.sans.org/media/analyst-program/implementers-guide-deception-technologies-39390.pdf

ENISA. 2021. Cybersecurity for SMEs - Challenges and Recommendations. *European Union Agency for Cybersecurity*. [Accessed: 2026-22-02] https://www.enisa.europa.eu/publications/enisa-report-cybersecurity-for-smes

ENISA. 2025. ENISA Threat Landscape 2025. *European Union Agency for Cybersecurity*. [Accessed: 2026-22-02] https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025.

EUROSTAT. 2026a. Cloud computing - statistics on the use by enterprises. *Eurostat*. [Accssed: 2026-02-05] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cloud_computing_-_statistics_on_the_use_by_enterprises#Cloud_computing_as_a_service_model_for_meeting_enterprises.E2.80.99_ICT_needs

FLEXERA SOFTWARE. 2025. Share of small and medium-sized enterprises (SMB) using workloads and data in public cloud in 2025. *Statista*. Statista Inc.. [Accssed: 2026-02-05] https://www.statista.com/statistics/1473568/worldwide-smb-workload-data-in-public-cloud/

GIETZEN, Spencer. 2018. AWS IAM Enumeration 2.0: Bypassing CloudTrail Logging. Rhino Security Labs. [Accssed: 2026-03-22] https://rhinosecuritylabs.com/aws/aws-iam-enumeration-2-0-bypassing-cloudtrail-logging/. [cit. 2026-03-22].

HAN, Xiao; KHEIR, Nizar; BALZAROTTI, Davide. 2018. Deception Techniques in Compu-ter Security: A Research Perspective. *ACM Computing Surveys (CSUR)*. 51(4), 80. ISSN 0360-0300. https://doi.org/10.1145/3214305

HUTCHINS, Eric; CLOPPERT, Michael; AMIN, Rohan. 2011. Intelligence-Driven Compu-ter Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains. *Leading Issues in Information Warfare & Security Research*. 1(1), 80. [Accssed: 2026-03-22]: https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf

JAVADPOUR, Amir; JA'FARI, Forough; TALEB, Tarik; SHOJAFAR, Mohammad; BENZAÏD, Chafika. 2024. A comprehensive survey on cyber deception techniques to improve honeypot performance. *Computers & Security*. 140. ISSN 0167-4048. https://doi.org/https://doi.org/10.1016/j.cose.2024.103792

KHODA PARAST, Fatemeh; SINDHAV, Chandni; NIKAM, Seema; IZADI YEKTA, Hadiseh; KENT, Kenneth B. *et al*., 2022. Cloud computing security: A survey of service-based models. *Computers & Security*. 114, 102580. ISSN 0167-4048. https://doi.org/10.1016/j.cose.2021.102580

KOLLI, Nagaraju. 2025. Navigating the Cloud: Security, Compliance, and Risk Challenges in SME Adoption. *Journal of Information Technology and Digital World*. 7(3), 200-215. ISSN 2582-418X. https://doi.org/10.36548/jitdw.2025.3.001.

KRATZKE, Nane. 2022. Cloud-Native Observability: The Many-Faceted Benefits of Structured and Unified Logging—A Multi-Case Study. Future Internet. 14(10), 274. ISSN 1999-5903. https://doi.org/10.3390/fi14100274

MARTSENIUK, Yevhenii; PARTYKA, Andrii; HARASYMCHUK, Oleh; NYEMKOVA, Elena; KARPINSKI, Mikolaj P. 2024. Shadow IT risk analysis in public cloud infrastructure. In: *CEUR Workshop Proceedings*, pp. 22-31. ISSN 1613-0073. [Accessed: 2026-03-04] https://ceur-ws.org/Vol-3800/paper3.pdf

MELL, Peter; GRANCE, Timothy. 2011. *The NIST Definition of Cloud Computing*. Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.SP.800-145.

MSAAD, Mohamed; SRINIVASA, Shreyas; ANDERSEN, Mikkel M.; AUDRAN, David H.; ORJI, Charity U. *et al*. 2022. Honeysweeper: Towards Stealthy Honeytoken Fingerprinting Techniques. In: Reiser, H. P., Kyas, M. (eds.). *Secure IT Systems*. NordSec 2022. Lecture Notes in Computer Science, vol 13700. Springer, Cham. https://doi.org/10.1007/978-3-031-22295-5_6

NARAYANASAMY, Deneesh. 2025. Cloud Security 101: Understanding Shared Respon-sibility and Securing Your Data: Understanding Shared Responsibility and Securing Your Data. *International journal of computer engineering & technology*. 16, 704-716. https://doi.org/10.34218/IJCET_16_01_058

NATIONAL CYBER SECURITY CENTRE. 2022. Cloud security guidance: Cloud security sha-red responsibility model. Online. *National Cyber Security Centre*. [Accessed: 2026-02-17] https://www.ncsc.gov.uk/collection/cloud/understanding-cloud-services/cloud-security-shared-responsibility-model

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. 2020. Security and Privacy Controls for Information Systems and Organizations. NIST SP 800-53 Rev. 5. https://doi.org/10.6028/NIST.SP.800-53r5

QIAN, Ling; LUO, Zhiguo; DU, Yujian; GUO, Leitao. 2009. Cloud Computing: An Overview. In: JAATUN, M. G., ZHAO, G., RONG, C. (eds.). *Cloud Computing*. CloudCom 2009. Lecture Notes in Computer Science, vol 5931. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-10665-1_63

SAXENA, Deepika; GUPTA, Rishabh; KUMAR SINGH, Ashutosh. 2021. *A Survey and Com-parative Study on Multi-Cloud Architectures: Emerging Issues And Challenges For Cloud Federation*. ArXiv.org. https://arxiv.org/abs/2108.12831v1

THE MITRE CORPORATION. [n.d.]. Cloud Matrix. *MITRE ATT&CK*. [Accessed: 2026-03-04] https://attack.mitre.org/matrices/enterprise/cloud/#

UDOVICIC, Stela. 2022. Logstash, Fluentd, Fluent Bit, or Vector? How to choose the right open-source log collector. *CNCF*. [Accessed: 2026-03-30] https://www.cncf.io/blog/2022/02/10/logstash-fluentd-fluent-bit-or-vector-how-to-choose-the-right-open-source-log-collector/

VECTOR. [n. d.]. AWS SQS. *VECTOR*. Vector documentation. [Accessed: 2026-04-20] https://vector.dev/docs/reference/configuration/sources/aws_sqs/

VERIZON. 2025. 2025 *Data Breach Investigations Report, Small and Medium-Sized Business Snapshot*. Verizon. [Accessed: 2026-03-06] https://www.verizon.com/business/resources/infographics/2025-dbir-smb-snapshot.pdf

Contact information

Matěj Cajthaml: e-mail: xcajtham@mendelu.cz
Tomáš Pitner: tomas.pitner@mendelu.cz

AN OPEN ETL AND RAG PIPELINE FOR UNIFIED ACCESS TO HETEROGENEOUS CLIMATE DATA

Dávid Chomo[1], Jaromír Landa[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_005

ABSTRACT

Climate data are dispersed across numerous portals in incompatible formats, presenting a substantial barrier for researchers and analysts working on climate change and sustainability. This paper presents an open system that automates the processing of climate datasets end-to-end. The system covers the entire workflow, from the downloading of raw files through to format detection (NetCDF, GRIB, GeoTIFF, CSV, HDF5, and ZIP archives), and transformation into a text description with statistical characteristics. The system also creates vector embeddings and support natural natural language querying. The ETL pipeline is orchestrated by Dagster, the vector store is provided by Qdrant with an HNSW index, and response generation takes place in a multi-level retrieval-augmented generation (RAG) pipeline combining weighted Reciprocal Rank Fusion, Maximal Marginal Relevance diversification, and a Corrective RAG validation layer. The integration of a web interface, constructed on the foundations of Vue.js 3 and FastAPI, serves to facilitate the accessibility of this functionality to non-technical users. The system achieved an average composite score of 91% on a set of ten queries, with all tests passing the 80% threshold. Following a series of optimizations, the average query latency decreased from 120 to 180 seconds to approximately 26 seconds. The primary contribution of this work lies in the integration of components that currently function independently within existing climate RAG systems, thereby reducing the barriers to accessing standardized climate data for sustainability research.

Keywords: ETL pipeline, climate data, vector database, retrieval-augmented generation, semantic search, Qdrant, automation, Dagster, open data, embeddings
JEL Code: C81, C88, Q54

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Climate change represents one of the most significant challenges of the 21st century, and any quantitative analysis of its manifestations or economic impacts depends on standardized climate data. However, this data is dispersed across numerous portals and stored in incompatible formats, which hinders its processing and utilization. The most significant sources comprise reanalyses, such as ERA5 (Hersbach *et al.*, 2020), CMIP6 model simulations (Eyring *et al.*, 2016), and observational datasets, including E-OBS (Cornes *et al.*, 2018) and CRU TS (Harris *et al.*, 2020). Each source is characterized by its own temporal granularity, spatial resolution, and metadata. Many portals require prior registration, and the data is often accessible only through specialized APIs. For researchers in the fields of climate economics, energy policy, or adaptation strategies, a significant portion of their project is spent on registration on portals, downloading data, converting formats, and extracting relevant variables, even before they can proceed to the actual analysis.

Rolnick *et al.* (2022) identified machine learning as a key tool for future climate research; however, easy access to input data remains a prerequisite. Existing open-source tools address only specific steps. Climate Data Operators enable operations on NetCDF and GRIB files, the GDAL library handles the conversion of georeferenced formats, and the Python library xarray (Hoyer and Hamman, 2017) facilitates work with N-dimensional arrays. These tools do not, however, share a common runtime; combining them requires substantial expertise, and their command-line orientation excludes users without programming skills.

The ETL (Extract, Transform, Load) pattern has been identified as a suitable solution for processing heterogeneous sources (Vassiliadis, 2009). This methodology involves the initial extraction of data, followed by its transformation (cleaning, normalization and enrichment), and subsequent loading into the target system. The alternative ELT model, which moves the transformation to the target repository (Kimball and Caserta, 2004), is not practical for climate data because the NetCDF and GRIB binary formats cannot be stored directly in a vector database. Instead, they must first be divided into spatial-temporal blocks, transformed into text descriptions, and only then embedded. Frameworks such as Apache Airflow, Prefect, or Dagster are considered suitable for orchestrating the ETL pipeline, with the latter using a declarative software-defined assets paradigm that reduces configuration requirements compared to traditional DAG-based workflow systems (Deelman *et al.*, 2015).

The challenge of efficient search within large volumes of processed data is addressed by techniques based on vector embeddings—compact representations that preserve the essential features of the content. The Retrieval-Augmented Generation (RAG) technique, as formally defined in Lewis *et al.* (2020), combines vector search in a knowledge base with answer generation by a large language model. The development of specialized systems for the climate domain is an area of current research. ChatClimate (Vaghefi *et al.*, 2023) provides responses to inquiries regarding IPCC reports by employing GPT-4 with retrieval, while ClimateGPT (Thulke *et al.*, 2024) leverages ongoing pretraining from the Llama-2 model on climate texts. Furthermore, GeoGPT-RAG (Huang *et al.*, 2025) is designed for geoscience publications. A common limitation of these systems is that they are only compatible with text documents, rather than raw raster climate data. Despite the fact that the Spatial-RAG system (Yu *et al.*, 2025) extends RAG with spatial filtering, it is limited to the GeoJSON format and is component-closed.

The objective of this paper is to design and implement an end-to-end system that consolidates heterogeneous climate data sources into a common format and makes them accessible to users without programming skills via a web interface that supports natural language queries. The system has been designed to encompass the entire chain, from the downloading of raw files through to

automatic format detection, transformation, generation of text descriptions, and creation of vector embeddings, culminating in final RAG querying. The work poses three specific research questions: firstly, how to design an ETL pipeline for heterogeneous climate data formats; secondly, how to build a web interface with automatic update orchestration; and thirdly, whether vector embeddings derived from text descriptions of climate chunks are sufficiently accurate for semantic search over a real-world collection.

From the perspective of sustainability and climate policy, easier access to standardized data has a direct impact: a faster cycle between the formulation of a research question and its answer, easier replicability of analyses, and the opening up of climate resources to non-traditional users, such as students, journalists, local governments, or small analytical groups without their own data infrastructure. The work thus contributes to the ongoing discourse on open data infrastructures as a technical prerequisite for sustainability research.

The article is structured as follows. The section entitled "Methodology and Data" provides a detailed description of the research framework employed, the architectural design of the proposed system, the catalogue of the input dataset, and the evaluation methodology. The "Results" section presents a quantitative assessment of the RAG system's quality and the performance characteristics of the ETL pipeline. The Discussion and Conclusions chapter provides an interpretation of the results within the context of existing climate RAG systems, identifies limitations, and outlines directions for future work.

2 METHODOLOGY AND DATA

This work follows the Design Science Research (DSR) methodology of Hevner *et al.* (2004), which is designed for the creation and evaluation of IT artefacts. The framework fits this work because the output is not only an analysis of the current state but primarily the design and implementation of a software artefact addressing the identified problem. The process was divided into three phases: firstly, an analysis of existing solutions was conducted through a literature review; secondly, a modular design and implementation of the system was undertaken; and thirdly, a final quality evaluation was made on a running endpoint.

The analysis is initiated with a catalogue comprising 233 records, which collectively represent climate datasets derived from diverse sources. The catalogue is divided into three categories of data. The ERA5 (Hersbach *et al.*, 2020), ERA5-Land, and CERRA combine observations with a physical model, thereby ensuring consistent coverage of the entire atmosphere. CMIP6 (Eyring *et al.*, 2016) and EURO-CORDEX model simulations are utilized for the projection of future climate scenarios. The E-OBS (Cornes *et al.*, 2018) and CRU TS (Harris *et al.*, 2020) observational datasets are based on station measurements. The datasets are available in six main formats (NetCDF, GRIB, GeoTIFF, CSV, HDF5, and ZIP archives) and vary in dimensionality, spatial and temporal granularity, and access methods. Of these 233 records, 167 are currently fully ingested end-to-end, while the remaining 66 are available at the metadata embedding level from the zero phase.

A five-phase classification system was designed for the systematic processing of heterogeneous sources. Phase 0 processes all 233 original records via text embeddings from the catalog metadata without downloading data. Subsequent phases cover 98 sources with direct HTTP downloads, 25 sources requiring registration, 71 sources accessible via API portals with seven specialized adapters (Copernicus CDS, ESGF, NOAA, NASA Earthdata, Marine Copernicus, EIDC, and NCAR RDA), and 39 sources without automated access, which remain at the metadata level. Point identifiers in the

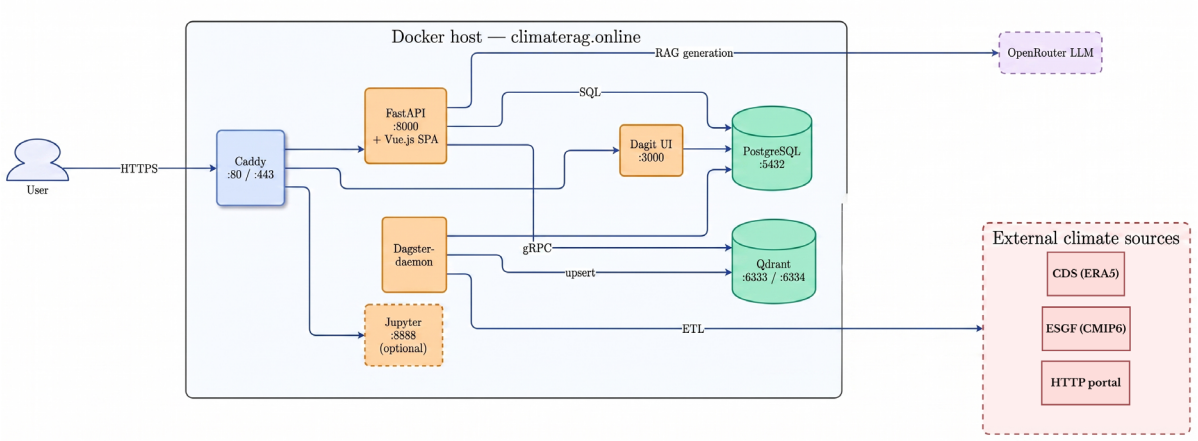


Fig. 1: The ClimateRAG system architecture. (Chomo, 2026)

vector database are generated deterministically from a UUID5 based on the source ID and chunk key, so that a repeated upsert from a higher phase overwrites the previous metadata representation without duplication.

The proposed ClimateRAG system is a distributed application with six main Docker services plus one optional service for interactive analysis. Figure 1 illustrates the overall architecture and interconnections of the components.

The entry point to the system is a Caddy reverse proxy with automatic HTTPS certificates. FastAPI handles the backend for the REST API and frontend serving, using asynchronous request processing. The ETL pipeline is orchestrated by Dagster using the software-defined assets paradigm, with scheduling via cron expressions and sensors for detecting updates. Relational metadata and processing statuses are stored in PostgreSQL, while vector embeddings are stored in Qdrant with an HNSW index (Malkov and Yashunin, 2020) with parameters $m = 16$ and $ef\_construct = 100$. The frontend is a single-page application in Vue.js 3 with nine views covering resource management, update scheduling, a chat with RAG responses, and monitoring of ETL runs.

The xarray library (Hoyer and Hamman, 2017) is used to load and transform climate grids, with generator-based streaming, which ensures that files hundreds of megabytes in size are never loaded entirely into memory but are instead chunked progressively according to their granularity (90 time steps and 40×40 spatial cells). For each spatial-temporal chunk, descriptive statistics are calculated, and a text description is generated containing the variable name, units, geographic area, time interval, and statistical characteristics according to CF Conventions. This text description is then encoded by the BAAI/bge-large-en-v1.5 model (Xiao *et al.*, 2023) into a 1024-dimensional vector and stored in a Qdrant collection with metadata in the payload. The choice of text embedding over a direct numerical vector is crucial because a purely numerical representation cannot distinguish temperature from precipitation—the same number 15 can mean both 15 °C and 15 mm.

The query pipeline combines several stages. LLM-driven query analysis extracts structured filtering attributes (dataset, variable, hazard, time window) while simultaneously generating variants of the expanded query. For each variant, vector search is performed in Qdrant via batch gRPC calls. The results are merged using the weighted Reciprocal Rank Fusion method (Cormack *et al.*, 2009), followed by diversification using the Maximal Marginal Relevance algorithm (Carbonell and Goldstein, 1998) with parameter $\lambda = 0.55$. The Corrective RAG layer uses a fast

LLM call to assess whether the found chunks are sufficient for a response, and if the results are insufficient, it rewrites the query and repeats the search. In the current version, the response is generated via the OpenRouter service using the Claude Sonnet 4.6 model.

A custom deterministic script was designed for evaluation, which sends test queries to a running endpoint and evaluates the response against expected facts using regular expressions and set-overlap, without any LLM involved in the scoring. The script measures five metrics: Context Relevance, Faithfulness, Answer Correctness, Retrieval Precision@5, and Source Diversity. The first four form a weighted composite score. Source Diversity is measured but not included in the composite because, for queries with a single authoritative source, forced diversity would degrade it without contributing to the quality of the answer. The script runs on a set of ten queries. These tests do not reflect what would "suit" the system, but rather questions that a climatologist might actually ask. The originally considered RAGAS framework (Es *et al.*, 2024) with an LLM-as-judge approach was not used because, in the specific climate domain, it is not guaranteed that the judge understands terms such as SPEI or lwe_thickness, and the variance between runs would complicate iterative optimization.

3 RESULTS

The catalogue currently features 167 sources (out of the original 233 records) that are ingested automatically from start to finish. The Qdrant vector collection contains around 1.53 million points with 1,024-dimensional embeddings. The HNSW index enables sub-second search queries, even with this volume of data.

The quality of the RAG system was measured iteratively. The baseline version, which combined HNSW search, cross-encoder reranking and generation via a large language model, achieved an average composite score of 78%. However, there were several major issues: a cross-domain query on the relationship between drought and wildfires failed due to the retrieval system's inability to return data from multiple thematic areas simultaneously; a CO_2 query was confused by the variables ‚co2_concentration' and ‚tcco' (total column carbon monoxide) in the embedding space; and latency hovered around 100 seconds per query.

The optimization process was executed across a total of eleven iterations, with each alteration being appraised utilizing the same test suite. Architecturally, several substantial modifications were introduced: multi-agent query decomposition generating three alternative variants and topic-focused variants for multi-topic queries; batch gRPC search instead of sequential calls; weighted RRF with primary weight 1.0, variant weight 0.6 and topic-focused query weight 0.9; MMR diversification with $\lambda = 0.55$; conditional data boost for multi-topic queries; a system prompt modification permitting linking of data to the query even with partial results; and a Corrective RAG validation layer. The cross-encoder reranker was eliminated from the primary processing pipeline due to its extended execution time of 40 to 50 seconds per query on the CPU, and its negligible impact on the composite score, contributing less than 1 percentage point. However, it was retained within the code as an optional fallback mechanism. The final deterministic run with a temperature of 0.0 achieved an average composite score of 91% and a success rate of 10/10, indicating that all tests passed above the 80% threshold.

Table 1 presents a comparison of the mean values of individual metrics between the baseline and the final deterministic run.

Tab. 1 Comparison of metrics is made between the baseline and the final deterministic run (Chomo, 2026)

Metric	Baseline	Final	Change (pp)
Context Relevance	89%	99%	+10
Faithfulness	86%	97%	+11
Answer Correctness	79%	81%	+2
Retrieval Precision@5	52%	86%	+34
Source Diversity	73%	74%	+1
Composite	**78%**	**91%**	**+13**

The most significant enhancement was observed in Retrieval Precision@5, with an increase of 34 percentage points. This finding corroborates the hypothesis that the majority of the observed quality gain was attributable to enhanced retrieval of pertinent segments, rather than to enhanced answer generation. A Context Relevance score approaching 100% indicates that, for the majority of tests, all returned chunks are deemed relevant. The accuracy of the answers remained at 81%, which can be attributed to the fact that the deterministic grader requires the exact occurrence of keywords in the answer, and the model occasionally uses a synonym (for example, “warm” instead of “warmest”), which counts as a mismatch even though the answer is semantically correct. The actual answer correctness is therefore higher than the nominally measured 81%. Concurrently, the discrepancy between the shift in Precision@5 and Answer Correctness indicates that the primary ongoing constraint of the system is no longer retrieval, but rather the generation component of the pipeline.

When interpreting the results, it is important to keep in mind that the test set contains only ten questions, so minor differences between adjacent iterations fall within the expected noise for such a small sample size. At n = 10, a single improved test shifts the overall score by 10 percentage points; differences of less than approximately 20 percentage points are not statistically significant in the sense of the Wilcoxon test. The highest scores are achieved by the per-test analysis for queries on drought (94%) and sea level (95%), where the system accurately identifies a single authoritative source. The lowest mean test scores were observed in queries concerning extreme heat (88%) and the atmosphere (89%), where the catalogue contains numerous datasets with analogous semantics and retrieval, thereby dispersing attention among them. Cross-domain queries (e.g. Canadian wildfires, Pakistani floods) achieved 89% and 90% accuracy, respectively, although they require linking data from multiple datasets.

The second tracked quantity is response latency. The baseline pipeline averaged 120 to 180 seconds per query, which is unacceptable for a web interface. After optimization, the median response time stabilized at 25.9 seconds across the ten test queries in the final deterministic run — an improvement of roughly 80%. Two changes drove most of the gain. First, batch gRPC search collapsed sequential variant queries into a single Qdrant call, cutting the search time from 60–100 seconds to about 5. Second, removing the cross-encoder reranker from the hot path saved 40–50 seconds per query.

To evaluate the performance of the ETL pipeline, four real-world datasets from the catalogue were tested. The processing of the CMIP6 dataset with three variables (tas, pr, psl) required approximately 95 seconds and yielded 3,124 chunks with a total size of approximately 50 MB. The CRU TS dataset was processed in approximately 2.3 minutes, yielding 5,718 chunks and 120 MB of output data. The E-OBS dataset, comprising four variables (tg, tn, tx, rr), yielded approximately 33,000

chunks, processed in batches that completed in approximately 32 seconds each. The processing of metadata for all 233 records in the zero phase was completed in under 30 seconds. Key optimizations in the ETL phase included batch GPU embedding, which delivered a 10–20-fold speed enhancement compared to sequential execution, and generator-based streaming, which maintains a constant memory footprint for processing regardless of the input file size.

4 DISCUSSION AND CONCLUSIONS

The implemented system encompasses the entire workflow, from the downloading of raw climate datasets through to automatic format detection, transformation into text-numeric descriptions, embedding generation and storage, and finally RAG querying via a web interface. The validation results on a test set of ten queries (average composite score of 91%, success rate of 10/10) demonstrate that the text embeddings of climate chunks are sufficiently accurate for this type of query.

A comparison of the proposed solution with existing climate RAG systems reveals significant distinctions, primarily in terms of the breadth of formats it supports and its ability to directly process raw climate data, in addition to text documents. ChatClimate (Vaghefi *et al.*, 2023) is a system that utilizes GPT-4 with retrieval to address queries regarding IPCC reports; however, it is exclusively compatible with PDF documents and is not open-source. ClimateGPT (Thulke *et al.*, 2024), which was trained via follow-up pretraining on climate texts, functions with scientific literature rather than raw datasets. GeoGPT-RAG (Huang *et al.*, 2025) is conceptually the closest, utilizing its own GeoEmbedding model built on Mistral-7B and a georeranker trained on spatial data. However, its RAG system targets geoscientific publications, not climate grids, and the system as a whole is not publicly available. Spatial-RAG (Yu *et al.*, 2025) is a system that explicitly addresses spatial filtering via geometric queries. However, it is limited to the GeoJSON format and is a closed-source system. The proposed system‘s key advantage lies in its combination of dimensions, as evidenced by the fact that none of the compared solutions cover the entire chain from downloading raw climate rasters through transformation and embedding to final web-based querying. Furthermore, the system is open-source.

From a technical standpoint, the most significant finding of this study is that text embeddings of climatic chunks outperform purely numerical representations. A statistical 8-dimensional vector (mean, standard deviation, minimum, maximum, percentiles) is incapable of distinguishing variables with a similar numerical scale, resulting in temperature and precipitation chunks occupying proximity in the latent space. The resolution of this issue is achieved through the provision of a textual description that encompasses units, attributes in accordance with CF Conventions, and statistical characteristics. This approach is substantiated by the observation that sentences such as “mean temperature 15.3 °C” and “total precipitation 15.3 mm” possess divergent semantics.

The work has several limitations. The BAAI/bge-large-en-v1.5 embedding model has been trained on English texts; consequently, queries in Slovak or Czech have lower accuracy. The multilingual alternative, BGE-M3, was not utilized due to incompatibility with ONNX Runtime. The test set of ten queries is inadequate for deriving statistically robust conclusions, and the deterministic keyword-based grader imposes penalties on semantically correct answers formulated using synonyms. Dependence on a single LLM backend (OpenRouter) is a single point of failure; on service outage the endpoint degrades to raw retrieval results, without a generated answer. Furthermore, Juhasz *et al.* (2024) also measured a policy violation rate of 27 to 30% on four tested models using climate RAG, indicating that the errors are not merely a local phenomenon of a specific implementation.

From the perspective of climate policy and sustainability research, easier access to standardized climate data can broaden who is able to use it: municipal analytics teams, journalists, educational institutions, and small academic groups without dedicated data infrastructure. Climatologists and researchers benefit from a shorter onboarding phase, students from simpler exploration in teaching, and downstream LLM applications in the climate domain from a ready-to-use data source. The open-source nature of the system, together with reproducible Docker Compose deployment, supports analysis replicability - important for the academic credibility of climate work.

The article demonstrates that an integrated open-source system for the automated processing of heterogeneous climate data with a RAG layer over a real-world dataset of over one million points is technically feasible and evaluably accurate within the framework of external validation. The paper explicitly identifies and quantifies compromises in language availability (English embedding model), the scope of evaluation (n = 10 queries), and response latency (26 seconds). For the majority of these compromises, a proposed next step exists, indicating that the system is not merely a demonstration of a working prototype, but also a starting point for real-world deployment and domain extensions.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

Several AI tools were used in the preparation of this article. They assisted with summarizing source material, translating passages between languages, and identifying typographical and stylistic issues in the text. All output produced by these tools was reviewed and verified. Summaries were checked against the original sources, translations were assessed for accuracy, and suggested corrections were evaluated before being accepted. The tools served as an aid to the writing process; the final content reflects the authors‘ judgment and remains the authors‘ responsibility.

REFERENCES

CARBONELL, J., GOLDSTEIN, J. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ,98)*. New York: ACM, pp. 335–336. https://doi.org/10.1145/290941.291025

CORMACK, G. V., CLARKE, C. L. A., BÜTTCHER, S. 2009. Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In: *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ,09)*. New York: ACM, pp. 758–759. https://doi.org/10.1145/1571941.1572114

CORNES, R. C., VAN DER SCHRIER, G., VAN DEN BESSELAAR, E. J. M., JONES, P. D. 2018. An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*. 123(17), 9391–9409. https://doi.org/10.1029/2017JD028200

CHOMO, D. 2026. *Automatizácia ETL procesov pre spracovanie klimatických údajov z webových zdrojov*. Master's thesis. Brno: Mendel University in Brno, Faculty of Business and Economics.

DEELMAN, E., VAHI, K., JUVE, G., RYNGE, M., CALLAGHAN, S., MAECHLING, P. J., MAYANI, R., CHEN, W., FERREIRA DA SILVA, R., LIVNY, M., WENGER, K. 2015. Pegasus, a workflow management system for science automation. *Future Generation Computer Systems*. 46, 17–35. https://doi.org/10.1016/j.future.2014.10.008

ES, S., JAMES, J., ESPINOSA ANKE, L., SCHOCKAERT, S. 2024. RAGAS: automated evaluation of retrieval augmented generation. In: *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL): System Demonstrations*. pp. 150–158.

EYRING, V., BONY, S., MEEHL, G. A. *et al*. 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6): experimental design and organization. *Geoscientific Model Development*. 9(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

HARRIS, I., OSBORN, T. J., JONES, P., LISTER, D. 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*. 7(1), 1–18. https://doi.org/10.1038/s41597-020-0453-3

HERSBACH, H., BELL, B., BERRISFORD, P. *et al.* 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. 146(730), 1999–2049. https://doi.org/10.1002/qj.3803.

HEVNER, A. R., MARCH, S. T., PARK, J., RAM, S. 2004. Design science in information systems research. *MIS Quarterly*. 28(1), 75–105. https://doi.org/10.2307/25148625.

HOYER, S., HAMMAN, J. 2017. xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*. 5(1). https://doi.org/10.5334/jors.148.

HUANG, F., WU, F., ZHANG, Z. *et al.* 2025. GeoGPT-RAG technical report. arXiv preprint. *arXiv:2509.09686*.

JUHASZ, M., DUTIA, K., FRANKS, H. *et al.* 2024. Responsible retrieval augmented generation for climate decision making from documents. arXiv preprint. *arXiv:2410.23902*.

KIMBALL, R., CASERTA, J. 2004. *The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data*. Indianapolis: Wiley.

LEWIS, P., PEREZ, E., PIKTUS, A. *et al.* 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: *Advances in Neural Information Processing Systems (NeurIPS)*.

MALKOV, Y. A., YASHUNIN, D. A. 2020. Efficient and robust approximate nearest neighbor using hierarchical navigable small world graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 42(4), 824–836. https://doi.org/10.1109/TPAMI.2018.2889473.

ROLNICK, D., DONTI, P. L., KAACK, L. H. *et al.* 2022. Tackling climate change with machine learning. *ACM Computing Surveys*. 55(2), 1–96. https://doi.org/10.1145/3485128.

THULKE, D., GAO, Y., PELSER, P. *et al.* 2024. ClimateGPT: towards AI synthesizing interdisciplinary research on climate change. arXiv preprint. *arXiv:2401.09646*.

VAGHEFI, S. A., WANG, Q., MUCCIONE, V. *et al.* 2023. ChatClimate: grounding conversational AI in climate science. Communications Earth & Environment. a*rXiv:2304.05510*.

VASSILIADIS, P. 2009. A survey of extract-transform-load technology. *International Journal of Data Warehousing and Mining*. 5(3), 1–27. https://doi.org/10.4018/jdwm.2009070101.

XIAO, S., LIU, Z., ZHANG, P. *et al.* 2023. C-Pack: packed resources for general Chinese embeddings. arXiv preprint. *arXiv:2309.07597*.

YU, D., BAO, R., NING, R. *et al.* 2025. Spatial-RAG: spatial retrieval augmented generation for real-world geospatial reasoning questions. arXiv preprint. *arXiv:2502.18470*.

Contact information

Dávid Chomo: e-mail: xchomo@node.mendelu.cz
Jaromír Landa: e-mail: jaromir.landa@mendelu.cz

SOCIAL ENGINEERING AND HUMAN-FACTOR VULNERABILITIES IN CRITICAL INFRASTRUCTURE: FINDINGS FROM A SIMULATED PHISHING CAMPAIGN

Kateřina Dobrovolná[1], Tomáš Pitner[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_006

ABSTRACT

Critical infrastructure organizations increasingly face cyberattacks that exploit human behaviour rather than technical vulnerabilities. Social engineering attacks, especially phishing and spear-phishing, remain among the most effective attack vectors because they target employees directly. This paper analyses the role of the human factor in cybersecurity within a critical infrastructure environment through a combination of threat intelligence analysis, questionnaire-based research, and a simulated phishing campaign implemented using the GoPhish framework. The study evaluates employee awareness, behavioural vulnerabilities, and organizational preparedness against social engineering attacks. The results confirm that phishing remains one of the dominant cyber threats and that technical protection alone is insufficient without systematic employee education and awareness-building. The paper proposes a set of organizational and technical countermeasures focused on strengthening the organizational "human firewall." The findings contribute to current discussions on cybersecurity resilience in critical infrastructure and highlight the importance of integrating continuous awareness programs with practical simulation exercises.

Keywords: cybersecurity, social engineering, phishing, critical infrastructure, human factor, threat intelligence
JEL Code: O33, M15, L86

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

The rapid digitalization of critical infrastructure has significantly increased organizational dependence on information and communication technologies. At the same time, cybercriminals increasingly exploit human psychology instead of purely technical vulnerabilities. Social engineering attacks manipulate victims into voluntarily disclosing sensitive information, opening malicious attachments, or performing actions that compromise organizational security.

Critical infrastructure organizations represent particularly attractive targets because disruption of their operations may have severe economic, social, and security consequences. Despite substantial investments into technological security mechanisms, the human factor frequently remains the weakest point in organizational cybersecurity.

Recent threat intelligence reports indicate a continuing growth of phishing-based attacks and socially engineered malware delivery campaigns (IBM, 2025). HTML phishing, malicious attachments, and impersonation techniques are among the most common threats observed in enterprise environments. Attackers increasingly rely on psychologically manipulative communication that creates urgency, fear, or trust.

This paper investigates the resistance of employees in a critical infrastructure environment to social engineering attacks. The research combines theoretical analysis, threat intelligence data evaluation, questionnaire-based research, and a simulated phishing campaign. The main objective is to identify the principal human-factor vulnerabilities and propose practical mitigation measures suitable for critical infrastructure organizations.

2 SOCIAL ENGINEERING AND THE HUMAN FACTOR

2.1 Social Engineering in Cybersecurity

Social engineering represents a set of manipulative techniques designed to exploit human behavior to bypass technological protection mechanisms (HADNAGY, 2018). Unlike conventional cyberattacks focused on software vulnerabilities, social engineering targets trust, emotions, routine behavior, and insufficient security awareness.

The most common forms include phishing, spear phishing, vishing, smishing, baiting, and impersonation attacks. Phishing attacks attempt to obtain credentials or sensitive information through fraudulent electronic communication. Spear phishing differs by focusing on a specific individual or organization using personalized information gathered from public or internal sources.

Vishing attacks exploit voice communication channels, while smishing uses SMS messages or social media communication. Baiting techniques rely on physical media or attractive digital content to trigger user curiosity and unsafe behavior.

Modern phishing campaigns increasingly imitate legitimate enterprise services such as Microsoft 365, OneDrive, SharePoint, banking systems, or delivery-service portals. Sophisticated attacks combine realistic visual imitation with psychologically persuasive language.

2.2 Human Firewall Concept

The concept of the "human firewall" describes employees who actively participate in cybersecurity defense through secure behavior and threat recognition (FORTINET, 2025). Unlike technical firewalls that filter network traffic, the human firewall relies on trained employees capable of recognizing suspicious communication and responding appropriately.

An effective human firewall requires:

- regular cybersecurity education,
- role-specific awareness training,
- simulated phishing exercises,
- incident reporting culture,
- management support,
- continuous reinforcement of secure behavior.

Well-trained employees can significantly improve organizational resilience by identifying suspicious emails, reporting anomalies, and preventing successful compromise before technical systems detect malicious activity.

3 RESEARCH METHODOLOGY

3.1 Research Design

The research was conducted as a mixed-method study combining:

- analysis of cybersecurity literature,
- evaluation of threat intelligence data,
- questionnaire-based employee survey,
- simulated phishing campaign.

The selected organization belongs to the critical infrastructure sector. Due to confidentiality requirements, all organizational data were anonymized and aggregated.

The study focused on three main research questions:

1. ***Which social engineering threats currently represent the greatest risk for critical infrastructure organizations?***
2. ***How resistant are employees to phishing-based manipulation?***
3. ***Which organizational and technical measures can improve resilience against social engineering attacks?***

3.2 Threat Intelligence Analysis

The research included analysis of threat intelligence telemetry data from 2025 provided within the organization's cybersecurity monitoring activities. The analysed data indicated that phishing and spam attacks represented the most frequent categories of observed threats.

Although many attacks were successfully blocked through technical security controls, the findings confirmed that attackers continuously target employees through increasingly sophisticated phishing campaigns.

The global threat intelligence trends also demonstrated a substantial growth of targeted phishing attacks involving malicious HTML attachments and fake Microsoft Office documents.

3.3 Questionnaire Survey

The empirical part of the research included a questionnaire survey involving 73 respondents working within the critical infrastructure environment. The survey collected data regarding:

- cybersecurity awareness,
- previous experience with phishing attacks,
- security training preferences,
- perceived organizational vulnerabilities,
- attitudes toward cybersecurity practices.

The respondents represented multiple organizational roles, including administration, management, technical operations, IT departments, and cybersecurity personnel.

3.4 Simulated Phishing Campaign

The practical experiment consisted of a simulated phishing campaign implemented using the GoPhish framework. The simulation aimed to evaluate realistic employee reactions to phishing communication without compromising operational security.

The campaign imitated a fraudulent Microsoft OneDrive login interface using the GoPhish framework (GOPHISH, 2025). Employees received a phishing email encouraging them to review a file allegedly containing salary information.

The landing page visually replicated the Microsoft OneDrive environment. After interaction with the phishing page, users were redirected to a harmless educational notification explaining the simulation and identifying suspicious indicators that should have been recognized.

The experiment intentionally avoided collecting passwords or sensitive personal information. Only behavioral interaction data were evaluated, ensuring compliance with organizational privacy requirements.

3.5 Threat Intelligence Findings

Threat intelligence analysis confirmed that phishing and spam attacks represented the dominant category of observed threats in the monitored environment (IBM, 2025). Web-based attacks and malicious email attachments were also frequently detected.

Global telemetry data demonstrated a substantial increase in targeted phishing attacks despite partial declines in mass spam campaigns. HTML phishing emerged as one of the most prevalent attack forms.

The findings support the assumption that attackers increasingly prioritize social engineering techniques because technological defenses alone are insufficient against human manipulation.

The analyzed telemetry data exhibited patterns consistent with publicly reported global threat intelligence trends, particularly regarding the dominance of phishing, spam, and socially engineered attacks (IBM, 2025). Similar trends were also identified in ESET Threat Reports H1 and H2 2025, which reported a substantial increase in targeted phishing campaigns, HTML phishing, and malicious email attachments as dominant attack vectors in enterprise environments.

3.6 Questionnaire Results

The questionnaire survey revealed relatively high awareness of phishing threats among respondents. Most participants indicated familiarity with phishing concepts and recognized suspicious emails as a major cybersecurity risk.

However, the results also identified several important weaknesses:

- inconsistent ability to recognize sophisticated phishing messages,
- insufficient practical training frequency,
- reliance on technical security mechanisms,
- varying levels of awareness across departments.

Respondents expressed strong interest in practical cybersecurity education and simulation-based training rather than purely theoretical instruction.

The results suggest that employees within the analyzed critical infrastructure environment demonstrated relatively higher awareness of phishing risks compared with patterns commonly reported in broader population-oriented phishing studies. Nevertheless, the simulated phishing campaign confirmed that awareness alone does not eliminate susceptibility to realistically designed social engineering attacks. Similar trends have been reported in ENISA Threat Landscape reports and Verizon Data Breach Investigations Reports, which identify phishing and human-factor exploitation among the most persistent attack vectors affecting both organizations and the general population.

3.7 Simulated Phishing Campaign Results

The phishing simulation demonstrated that even employees familiar with cybersecurity principles remained vulnerable to realistically designed phishing attacks.

The most significant observed vulnerabilities included:

- trust in visually legitimate interfaces,
- automatic interaction with urgent messages,
- insufficient verification of URLs and domains,
- reduced caution during routine activities.

The experiment confirmed that psychological manipulation and realistic visual imitation significantly increase attack success probability.

At the same time, the simulation showed that some employees successfully recognized suspicious indicators and avoided interaction, demonstrating the positive impact of prior cybersecurity awareness.

4 DISCUSSION

The research findings confirm that the human factor remains a critical component of organizational cybersecurity resilience. Although technical protection systems effectively mitigate many conventional attacks, social engineering techniques continue to bypass technological defenses by exploiting human behavior. The study also highlights several important practical implications.

First, cybersecurity awareness cannot rely solely on occasional mandatory training sessions. Effective security culture requires continuous reinforcement and practical exercises.

Second, simulation-based education provides substantially greater effectiveness than purely theoretical instruction. Employees learn more effectively through realistic attack scenarios that reproduce authentic phishing techniques.

Third, organizations should treat cybersecurity awareness as a long-term organizational process rather than a one-time compliance requirement.

The findings correspond with contemporary cybersecurity research emphasizing the growing importance of user-centric security models and behavioral defense mechanisms (WHITMAN and MATTORD, 2012).

5 RECOMMENDED COUNTERMEASURES

Based on the research findings, the following technical and organizational measures are recommended.

5.1 Organizational Measures

- Regular phishing simulation campaigns.
- Role-specific cybersecurity awareness training.
- Mandatory reporting procedures for suspicious communication.
- Security-awareness integration into onboarding processes.
- Continuous employee feedback and educational support.
- Strengthening management involvement in cybersecurity culture.

5.2 Technical Measures

- Multi-factor authentication.
- Advanced email filtering and anti-phishing systems.
- DNS and URL reputation filtering.
- Behavioral anomaly detection.
- Segmentation of critical systems.
- Threat intelligence integration.

The findings indicate that the most effective mitigation strategy combines technical protection mechanisms, including multi-factor authentication, advanced email filtering, and behavioral anomaly detection, with continuous awareness training and simulation-based education.

These findings correspond with contemporary human-firewall and defense-in-depth approaches emphasizing the integration of technical and behavioral security controls (FORTINET, 2025; WHITMAN and MATTORD, 2012).

6 LIMITATIONS

Several limitations affected the presented research. The simulated phishing campaign could not fully replicate highly sophisticated real-world attacks due to ethical and operational constraints. The analyzed organization also required anonymization and aggregation of telemetry data because

of critical infrastructure confidentiality requirements. Additionally, the threat landscape continuously evolves, meaning that attacker techniques observed during 2025 may change significantly in future periods.

7 CONCLUSION

This paper analyzed the role of social engineering in the cybersecurity of critical infrastructure organizations. The research combined threat intelligence analysis, employee survey data, and a simulated phishing campaign to evaluate organizational resilience against phishing-based attacks.

The findings confirm that phishing remains one of the most significant cybersecurity threats targeting critical infrastructure environments. Despite technological security investments, attackers continue to exploit human behavior successfully.

The research demonstrates that cybersecurity resilience requires not only technical protection mechanisms but also systematic investment into employee education, practical simulations, and organizational security culture.

Future research should focus on long-term behavioral changes resulting from repeated awareness programs and on the integration of artificial intelligence into phishing detection and employee training systems.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

BIRTHRIA, S. K., AHLAWAT, P., JAIN, A. K. 2025. Detection and prevention of spear phishing attacks: a comprehensive survey. *Computers & Security*. 151, 104317. https://doi.org/10.1016/j.cose.2025.104317

CIALDINI, R. B. 2021. *Influence: The Psychology of Persuasion*. Brno: Jan Melvil Publishing. ISBN 978-80-7555-135-1.

FORTINET, 2025. Human Firewall: Strengthening Enterprise Security Through Awareness. *Fortinet*. [Accessed: 2026-04-03] https://www.fortinet.com/resources/cyberglossary/human-firewall

GOPHISH, 2025. What is GoPhish? *Gophish user Guide*. [Accessed: 2026-04-04] https://docs.getgophish.com/user-guide/what-is-gophish

HADNAGY, C., 2018. *Social Engineering: The Science of Human Hacking*. Hoboken, New Jersey: John Wiley and Sons Inc. ISBN 9781119433385

IBM, 2025. What is Threat Intelligence? *IBM*. [Accessed: 2026-04-01] https://www.ibm.com/think/topics/threat-intelligence

JAKOBSSON, M., MYERS, S., 2007. *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*. Hoboken, New Jersey: John Wiley and Sons Inc. ISBN 9780471782459

KOLOUCH, J., BAŠTA, P., 2019. *CyberSecurity*. Praha: CZ.NIC, z.s.p.o. ISBN 9788088168317

NÁRODNÍ ÚŘAD PRO KYBERNETICKOU A INFORMAČNÍ BEZPEČNOST, 2023. *Výkladový slovník kybernetické bezpečnosti*. [Accessed: 2026-02-10] https://nukib.gov.cz/download/publikace/podpurnematerialy/Vykladovy-slovnik-kyberneticke-bezpecnosti-v6.pdf

STALLINGS, W., BROWN, L. 2024. *Computer Security: Principles and Practice*. Hoboken: Pearson Education, Inc. ISBN 9781292473291

ENISA. 2025. *ENISA Threat Landscape 2025*. European Union Agency for Cybersecurity. [Accessed: 2026-04-10] https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025

VERIZON, 2025. 2025 Data Breach Investigations Report. *Verizon*. [Accessed: 2026-04-10] https://www.verizon.com/business/resources/reports/dbir/

WHITMAN, M. E., MATTORD, H. J. 2012. *Principles of Information Security*. Boston: Thomson Course Technology. ISBN 9781337102063

ZERDALOGLU, S. 2019. *Fundamentals of Cybersecurity*. Brno: Mendel University in Brno. ISBN 9788075096487

Contact information

Kateřina Dobrovolná: e-mail: xdobrovo@mendelu.cz

Tomáš Pitner: e-mail: tomas.pitner@mendelu.cz

TRUST BETWEEN INDIVIDUALS: AN EMPIRICAL ANALYSIS INSPIRED BY COVEY AND SINEK

Tereza Frömmelová[1], Kateřina Pojkarová[2]

[1]Prague University of Business and Economics, Faculty of Informatics and Statistics, nám. W. Churchilla 1938/4, 130 67 Praha 3, Czech Republic

[2]University of Pardubice, Faculty of Transport Engineering, Studentská 95, 532 10 Pardubice 2, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_007

ABSTRACT

This paper examines trust as described in the works of Simon Sinek and Stephen M. R. Covey, which together form the theoretical framework for this study. Their approaches view trust as a relational and value-based process that develops through consistent behavior, integrity, and psychological safety. Based on this theoretical framework, the authors of the paper conducted a questionnaire survey among university students in 2025. The study aimed to explore how young adults understand trust, how they apply it in personal and team relationships, and whether their views support or challenge the concepts proposed by Covey and Sinek. The findings show that students' views on trust in teamwork reflect many elements of contemporary trust theory, particularly the emphasis on authenticity, reliability, and psychological safety. The paper contributes to a deeper understanding of how theoretical models of trust are translated into lived experiences among young adults.

Keywords: trust, interpersonal relationships, teamwork, survey research
JEL Code: D90, J24

1 INTRODUCTION

Trust is a fundamental element across all domains of human life. It can be understood as the willingness of all parties in a relationship to place their trust in one another. Initially, trust develops in early childhood, particularly in the relationship between children and their parents. Over time, it extends to teachers and other individuals in a person's social environment. In adulthood, trust becomes especially important in relationships with those with whom individuals share their everyday lives. A distinct and significant dimension of trust is also present in professional contexts, notably in the relationships between employees and their superiors and among members of work teams.



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

Trust must be continuously built and subsequently maintained. A crisis of trust is widely regarded as a critical factor in the deterioration of relationships. Once trust is lost, it is extremely difficult, if not impossible, to restore. When an organization loses its employees' trust, turnover may increase. Importantly, such a decline in trust may be triggered by a single unexpected event, potentially leading to the sudden departure of skilled workers and, in extreme cases, threatening the organization's very functioning. Moreover, it is not only the erosion of employees' trust in organizations that can generate instability; broader disruptions of trust may also have significant economic consequences.

Economic literature identifies crises of trust in key institutions as among the most serious threats to economic stability. A loss of confidence in the banking system may trigger mass withdrawals of deposits and bank runs, which can escalate into monetary and financial disruptions. Similarly, declining trust in government and other public institutions is commonly associated with rising public dissatisfaction and the erosion of social cohesion, developments that may adversely affect society's overall economic performance (Diamond & Dybvig, 1983).

In personal life, the loss of trust in a partner is often one of the primary reasons for the dissolution of intimate relationships. When a breakdown of trust leads to the disintegration of marriage, it is commonly recognized as grounds for divorce (Gottman, 2023; Höhne & Paloncyová, 2021). Similarly, the erosion of trust between parents and children may result in alienation and the severing of familial ties (Cherlin, 2004).

This article examines the link between the theoretical concept of trust, as articulated by Covey and Sinek (Covey & Merrill, 2006; Sinek, 2019), and empirical evidence from a questionnaire survey conducted among university students in the Czech Republic and other European countries. By combining established theoretical perspectives with original empirical observations, the study seeks to deepen the understanding of trust across interpersonal and institutional contexts.

The first part of the article reviews the literature on trust and its various forms. Particular attention is paid to different definitions of trust found primarily in the fields of managerial economics and psychology. This section further outlines how trust is conceptualized by Covey in The Speed of Trust (Covey & Merrill, 2006) and examines how this framework is complemented by Sinek's perspective presented in The Infinite Game (Sinek, 2019).

The empirical part of the article focuses on the interpretation of results from a questionnaire survey conducted among university students in the first half of 2025. It begins with the presentation of the research questions, followed by a brief description of the questionnaire used for data collection. Subsequently, the initial findings related to different patterns of trust observed among respondents are presented. The section then discusses the results of the questionnaire component dedicated to trust-related behavior.

2 THEORETICAL BACKGROUND

2.1 Definition and Dimensions of Trust

Research presented in this article is situated at the intersection of managerial and psychological perspectives on trust. To provide a solid foundation for understanding the research hypotheses and findings, it is essential to outline the concept of trust. It is also necessary to distinguish between how trust is defined within the fields of management and psychology.

Authors in the academic literature (Mayer *et al.*, 1995) define trust through various characteristics attributed to both the trustor and the trustee. Mayer (1995) defines trust as *"the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party"*. According to this definition, the emergence of trust between two parties requires a **willingness to be vulnerable**. This reflects the idea that trust is not about certainty, but rather about the readiness to accept risk. It also includes the **expectation of a particular behavior**. Importantly, trust implies taking that risk even in **the absence of the ability to monitor or control the other party**.

As previously mentioned, trustworthiness is a narrower construct than trust, referring specifically to the characteristics perceived by the trustor that influence their willingness to trust. According to Mayer *et al.* (1995), trustworthiness consists of three key components: **ability, benevolence, and integrity**.

Ability refers to the set of skills, competencies, and knowledge perceived by the trustor. The trustor, either consciously or unconsciously, evaluates whether the trustee is sufficiently capable of fulfilling the expected task or responsibility. **Benevolence** reflects the belief that the trustee is not solely motivated by self-interest but is genuinely concerned about the well-being of the trustor. **Integrity** represents the perceived moral character of the trustee, including adherence to principles, consistency in behavior, and the fulfillment of promises.

Another important dimension in the conceptualization of trust was introduced by McAllister (McAllister, 1995), who distinguished between cognitive and affective trust based on the degree to which **emotions and reasoning** are involved in the development of trust.

Cognitive trust is grounded in rational evaluation. It reflects the trustor's belief in the trustee's competence, reliability, and professionalism. It is typically developed through observation and experience and is more common in task-oriented or professional relationships.

Affective trust arises from emotional bonds and a sense of interpersonal care. It reflects the trustor's perception that the trustee genuinely cares about their well-being, shares mutual concern, and will act with goodwill. This form of trust is built through emotional closeness, expressions of empathy, and shared personal experiences. It often requires more time to develop and tends to be stronger in close and long-term relationships, such as family and close friendships (McAllister, 1995).

2.2 Covey's Concept of Trust: The Speed of Trust

In the context of this study, Stephen M. R. Covey's work, The Speed of Trust (Covey & Merrill, 2006), offers a relevant framework for understanding trust as a measurable and actionable construct. Covey presents trust as a strategic asset that directly influences the efficiency and quality of interactions, both in professional environments and personal relationships.

Covey (Covey & Merrill, 2006) conceptualizes trust as a function of two essential components: **character** and **competence**. This dual structure offers a practical and comprehensive understanding of how trust is built, lost, and restored in various relational settings.

Character-based trust refers to the internal moral foundation that an individual possesses. It is built on two elements: **integrity** and **intent**. **Integrity** involves honesty, consistency, and alignment between words and actions. **Intent** relates to a person's motivation, meaning whether their actions are driven by genuine concern and fairness, or by self-interest. In contexts such as family relationships, where emotional bonds and long-term mutual care are central, character-based trust tends to play a dominant role (Covey & Merrill, 2006).

Competence-based trust is built on a person's ability to perform effectively. It consists of two main components. The first component of competence-based trust is capability, which includes knowledge, practical skills, and the ability to adapt to various situations. The second component is **results**, referring to the person's history of successful performance. This form of trust plays a key role in teamwork and goal-oriented relationships, where individuals rely on others to contribute to shared outcomes (Covey & Merrill, 2006).

To establish a sustainable trust relationship, both dimensions of trust must be present. A person may have good intentions (character), but if they lack the ability to deliver (competence), trust will eventually erode. Conversely, someone who is highly skilled but whose motives are questionable may also fail to earn trust. Trust is fully established only when both character and competence are present (Covey & Merrill, 2006).

2.3 Sinek's Perspective: The Infinite Game

Simon Sinek views trust as a fundamental element of long-term cooperation and effective leadership. In his book, The Infinite Game (Sinek, 2019), he contrasts short-term, competitive "finite games" with long-term, purpose-driven "infinite games." Trust is an essential part of the "infinite" approach, as it cannot be forced or achieved through control. Instead, it must be **earned and maintained through consistent behavior, empathy, and integrity**.

Sinek (2019) argues that organizations and individuals often play a finite game when seeking to "win" against others through performance metrics, rankings, or short-term goals. An infinite mindset focuses on **continuity, adaptability, and endurance** both in career goals and relationships. Trust, in his concept, is something that allows and sustains collaboration even when outcomes are uncertain, because individuals believe in one another's intentions and share a common purpose.

A key component of Sinek's work is that **leadership means creating an environment of psychological safety**. He insists that people perform best when they feel protected, respected, and valued. In such conditions, team members are more willing to take risks, share new ideas, and admit mistakes. Team members are also taking voluntary actions that reinforce mutual trust within the team. He links this to the principle of **servant leadership**, where leaders prioritize people over results. As Sinek (2019) phrases it: ***"Leaders are not responsible for the results; they are responsible for the people who are responsible for the results."***

Trust is not transactional, but relational, which means it is b**uilt through small, consistent acts over time**, not through one-time demonstrations of competence. Trust grows when people observe alignment between words and actions, especially under pressure or adversity. On the other hand, **betrayal of trust has long-lasting negative effects**, because it undermines the sense of safety essential to cooperation. Therefore, trust must be intentionally nurtured through transparency, fairness, and shared purpose. In any cooperative group, such as student project teams or families, trust functions as a stabilizing force. When members feel safe and supported, they are more resilient to challenges and more willing to collaborate. This together creates **a culture of belonging and long-term commitment** (Sinek, 2019).

Although Simon Sinek, Roger Mayer, Daniel McAllister, and Stephen M. R. Covey (authors mentioned earlier) approach trust from different contexts, such as leadership philosophy, organizational psychology, and applied management, they share a common understanding of trust as a **relational, dynamic, and value-based process**. All four authors emphasize that trust develops gradually through ongoing interactions rather than being granted instantly. Trust must be built and sustained through consistent behavior, transparency, and moral integrity.

Research (Covey & Merrill, 2006) indicates that trust is transferred across relationships. That is because past experiences highly influence an individual's capacity to trust. Early trust experiences, particularly within families, form mental models of how safe or risky it is to rely on others. It can be assumed that trust is both personal and systemic, meaning that it is shaped by individual histories and experiences and later reproduced within new interactions within social and organizational structures.

2.4 The Influence of Social Environment and Family on the Development of Trust

Trust begins to form within the family, where children first experience being cared for, supported, and trusted. These early interactions create expectations about how safe it is to rely on others. When parents communicate confidence and allow autonomy, they allow their children to develop a perception that people are generally dependable. In contrast, excessive control or inconsistency can lead to cautious or defensive trust patterns later in life (Giulietti *et al.*, 2023).

Empirical evidence (Giulietti *et al.*, 2023) confirms that family influences the development of trust in complex ways. While parental trust does play a role, it is the overall family environment (such as shared norms, communication patterns, and emotional climate) that matters most. That means that trust is more socially cultivated than directly inherited.

Family-based trust formation is also linked to the transmission of values and social norms. Everyday interactions teach children what honesty, fairness, and responsibility mean in practice. Through the process of socialization, children develop an internal "trust compass" that guides their later perceptions and responses to others in school, friendship groups, and eventually in professional teams. In this sense, the family serves as both an emotional and moral foundation for the broader capacity to trust, providing the key environment in which the first patterns of trusting and being trusted are formed, which means that family provides patterns that shape how trust is later perceived, maintained, and transmitted across other relationships (Robbins, 2016).

3 SUMMARY AND FORMULATION OF THE RESEARCH QUESTIONS AND HYPOTHESES

Building on the theoretical framework and the empirical insights discussed above, this article formulates the following research questions aimed at examining the determinants of trust perception and trust-related behavior among university students.

The main research question (RQ1) is as follows:

Which factors affect how university students perceive and extend trust?

Sub-research question RQ2:

How is the relationship with parents (in terms of perceived parental trust) related to the willingness to extend trust to others?

Young people are influenced by various social and family factors, particularly by their parents and the environment in which they grow up. Therefore, it is relevant to examine whether their level of trust is significantly affected by the degree of autonomy they are granted in making independent decisions, or conversely, by parental restrictions that limit such independence.

Sub-research question RQ3:

Are there differences in the level of perceived and extended trust among students based on gender, age, employment status, living arrangements or financial independence?

The premise is that the level of trust differs among people of various ages, and that financially independent individuals may perceive and extend trust differently from those who are still dependent. It is also assumed that people with work experience and those who live independently from their parents (thus being more self-reliant) may approach or manage trust in different ways.

4 METHODOLOGY

The survey was conducted from late 2024 to April 2025. A pilot study was first carried out with students from the Prague University of Economics and Business and the University of Pardubice. Based on the results, two additional questions were incorporated into the questionnaire, which was then distributed to students from both Czech and international universities.

Respondents were predominantly from the EU countries (Czech Republic, Slovakia, France, Germany, Finland, Portugal, Hungary, etc.), with a small number from outside the EU, all of whom were studying at European universities at the time. In the main phase, 245 students participated, with a near-equal gender distribution (54% male, 46% female).

The largest age group was 21–25 years (49%), followed by those aged 20 and under (46%). Only a small proportion were over 25, reflecting the target group of university students. Half of the respondents were first-year students; representation declined with higher years of study, with postgraduate students forming the smallest group.

Regarding living arrangements, 46% lived with their parents, 36% in shared accommodation, 13% with a partner, and 5% alone. Most respondents (70%) were financially dependent on their parents.

The questionnaire entitled “Trust Between Individuals” was designed to explore interpersonal trust, perceptions of trustworthy behavior, and related attitudes among university students. The survey was administered online and took approximately 10–15 minutes to complete. It consisted of five main parts, containing both categorical and ordinal items. All questions were mandatory.

The final dataset consisting of the answers of each respondent includes variables that are corresponding to the questions in the questionnaire. There vere various types of questions asked in the questionnaire which generated various types of statistic variables that recorded questions of all respondents. Not all questions and related variables were considered for this very paper.

5 EMPIRICAL FINDINGS

5.1 Patterns of Trust Distribution: Who Is Entrusted with What

Over 90% of respondents reported having at least three people in their lives whom they could trust; however, only 62% indicated they would entrust someone with the PIN to their credit card. In comparison, 84% would share the password to their computer or mobile phone, and 86% would provide the keys to their home.

A small proportion (1%) would give their card PIN but not their home keys, while a quarter (25%) would do the opposite, provide home keys but not the PIN. Thirteen percent would not share either keys or PIN, and an identical share would not disclose either their computer/phone password or their card PIN. Three percent would share their card PIN but not their password, whereas 25% would share their password but withhold their card PIN.

These findings suggest that while most students maintain a relatively broad circle of trusted individuals, the level of trust varies substantially depending on the type of information or asset involved. Sensitive financial information, such as a card PIN, is guarded far more strictly than physical access (home keys) or digital access (device passwords). The asymmetries observed between these categories indicate that trust is context-dependent and shaped by perceived risk.

When students are trusted by their parents to make independent decisions, they, in turn, tend to trust others (including their parents) enough to share their card PIN. Conversely, when parents do not trust their children to make independent decisions, the students are less likely to trust others with their PIN.

Women are generally more likely than men to have someone in their circle they would entrust with their PIN.

The data also reveal that Slovak students demonstrate significantly higher levels of trust than Czech students (80% vs. 54% willing to share their PIN).

No significant effects were observed for the other examined factors. These results indicate that parental trust may play a key role in shaping students' broader trust relationships, and that cultural or national differences can substantially influence patterns of interpersonal trust.

Tab. 1 Factors Influencing Trust in Entrusting Others with Credit Card PIN Among University Students

Factors examined for their influence on trust	P value (given by chisquare test)	Conslusion
Relationship with parents	0,887	No
Ability to rely on parents for emotional support	0,630	No
Trust of parents to make independent decisions	0,054	YES
Gender	0,069	YES
Employment status	0,363	No
Living arrangements	0,240	No
Financial independence	0,270	No
Country of birth	<0,001	YES

Tab. 2 Factors Influencing Trust in Entrusting Others with Keys to their Home Among University Students

Factors examined for their influence on trust	P value (given by chisquare or Fisher test)	Conslusion
Relationship with parents	0.081	No
Ability to rely on parents for emotional support	0.176	No
Trust of parents to make independent decisions	0.006	YES
Gender	0.059	YES
Employment status	0.228	No
Living arrangements	0.182	No
Financial independence	0.846	No
Country of birth	0.084	YES

Similar results were observed regarding trust in entrusting others with the keys to one's home. A statistically significant influence was found for gender, parental trust in the student's ability to make independent decisions, and the country of birth. Specifically, women, Slovak students, and those who receive parental trust were more likely to identify someone in their social circle to whom they would entrust their home keys. Conversely, the closeness of the relationship with parents, reliance on parental emotional support, employment status during studies, and living arrangements showed no significant effect on this form of trust.

The results show that neither the quality of relationships with parents nor their perceived emotional support is associated with respondents' willingness to share their computer password. By contrast, perceived parental trust in respondents' ability to make independent decisions is positively associated with respondents' trust in others, though this association is based on subjective

Tab. 3 Factors Influencing Trust in Entrusting Others with Password Among University Students

Factors examined for their influence on trust	P value (given by chisquare or Fisher test)	Conslusion
Relationship with parents	0.081	No
Ability to rely on parents for emotional support	0.176	No
Trust of parents to make independent decisions	0.006	YES
Gender	0.059	YES
Employment status	0.228	No
Living arrangements	0.182	No
Financial independence	0.846	No
Country of birth	0.084	YES

perceptions. Statistically significant gender differences were identified, with women being more likely than men to entrust their password to someone in their close environment. Employment status, financial independence, and living arrangements showed no statistically significant association with password-sharing trust, as confirmed by Fisher's exact test. Cross-national comparisons indicate that while respondents in the Czech Republic are less willing to share a card PIN than those in Slovakia, their willingness to share a computer password does not differ.

5.2 Behavior of trust

Respondents evaluated their perceived ability to behave in a trustworthy manner using a 10-point scale.

Tab. 4 Descriptive Statistics of behavior of trust

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**average**
I keep promises (to myself, family, at school, at work ...).	3	6	8	4	10	11	29	80	**69**	**39**	7.81
I express my intentions openly and regularly.	**3**	**15**	23	15	28	35	44	43	34	19	6.40
The motives behind my actions are clear, honest, and straightforward.	3	7	12	11	18	14	34	57	70	33	7.43
My actions aim for mutual benefit.	2	13	13	9	31	28	38	56	38	31	6.91
I have the knowledge and skills necessary for what I do (field of study, field of work etc.).	**6**	**11**	14	10	29	19	47	49	48	26	6.88
I continuously seek to learn and improve.	3	4	11	8	26	18	31	44	50	64	7.61
I take responsibility for the results of my work.	4	4	10	2	10	9	19	30	**69**	**102**	8.37
Average	3.4	8.6	13	8.4	21.7	19.1	34.6	51.3	54	44.9	–
Maximum	6	15	23	15	31	35	47	80	70	102	–
Minimum	2	4	8	2	10	9	19	30	34	19	–

Tab. 5 Factors Influencing behaviors (extremes: low = 1 and 2 points, high = 9 and 10 points)

P value	Promises	Intentions	Motives	Actions	Skills	Improvement	Responsibility
Relationship with parents	0.57	0.856	0.796	0.260	0.353	0.776	0.682
Support from parents	0.685	0.573	1.000	0.764	1.000	0.675	0.201
Trust of parents	1.00	1.00	0.353	1.000	0.254	0.255	0.599
Age	0.584	0.597	0.772	1.000	0.360	0.836	0.698
Gender	0.305	**0.013**	**0.041**	0.154	0.927	0.125	1.000
Employment status	1.000	0.360	0.342	0.805	**0.089**	1.000	1.000
Living arrangements	0.161	0.504	0.160	0.328	0.787	**0.026**	0.607
Financial independence	0.268	**0.072**	0.277	0.504	**0.004**	0.246	0.439
Country of birth	1.000	0.660	0.679	0.106	0.329	0.590	0.354

Tab. 6 Factors Influencing behaviors (high, moderate, low scores)

P value	Promises	Intentions	Motives	Actions	Skills	Improvement	Responsibility
Relationship with parents	0.158	0.991	0.787	0.960	0.926	0.134	0.825
Support from parents	**0.075**	0.933	0.702	0.847	0.399	0.278	0.222
Trust of parents	0.173	0.741	1.000	0.214	**0.085**	0.184	0.755
Age	0.98	0.619	0.366	**0.144***	**0.061**	0.907	0.993
Gender	**0.019**	**0.085**	**0.032**	0.294	**0.068**	0.305	0.861
Employment status	0.377	0.145	0.708	0.741	0.120	0.226	0.206
Living arrangements	0.789	**0.072**	0.360	0.595	0.875	0.357	0.550
Financial independence	0.238	**0.050**	0.207	0.332	**0.005**	**0.024**	0.219
Country of birth	0.467	**0.070**	**0.066**	**0.076**	**0.056**	0.138	0.155

*(0.061 for group younger than 25 years and older than 25 years)
Note: Results lower than 0,1 means there is a significant influence

Respondents feel strongest in taking responsibility for outcomes and also fairly confident in keeping their promises. Conversely, they find it more difficult to openly express their intentions.

One approach to interpreting the results is to compare the number of respondents who selected extreme responses (those who feel either completely confident or entirely inadequate in the given behavior). An alternative approach is to categorize responses into three groups: high (respondents who believe they manage the behavior well, giving scores of 8–10), moderate (scores of 4–7), and low (scores of 1–3).

These results provide a basis for investigating whether the chosen factors have a significant effect on respondents' trustworthiness.

I keep promises (to myself, family, at school, at work ...).

The influence of two factors proved to be significant: gender and the ability to rely on emotional support from parents. Women and those who can depend on parental support are more likely to keep their promises.

I express my intentions openly and regularly.

Women tend to express their intentions more openly, with only 10% feeling weak in this skill compared to 20% of men. Similarly, financially independent students report fewer difficulties—only 3% in this group consider themselves poor at expressing intentions, whereas 9% of financially dependent students do. Slovak respondents, unlike their Czech counterparts (19%), generally do not struggle much with expressing intentions, with only 3% feeling weak in this area. Those who live alone express their intentions most confidently, with 57% reporting strong skills in this area. Conversely, respondents living with their parents or in shared housing most frequently indicated weakness in expressing intentions—14% among those in shared accommodation and 22% among those living with parents.

The motives behind my actions are clear, honest, and straightforward

Gender and country of birth are the only factors found to have a significant impact on respondents' perceived honesty of their motives. While only 4% of women report having impure motives, this figure rises to 12% among men. Regarding the opposite extreme, 83% of Slovak respondents consider themselves highly honest, compared to only 61% of Czech respondents.

My actions aim for mutual benefit.

The actions of Slovak respondents are more frequently motivated by mutual benefit compared to their Czech counterparts, with 63% of Slovaks perceiving themselves as sufficiently proficient in this behavior, compared to only 48% of Czechs. A significant difference emerged when comparing respondents aged 25 and above with those under 25. Older students were more likely to be motivated by mutual benefit than their younger counterparts. None of the other examined factors showed a significant influence on this behavior.

I have the knowledge and skills necessary for what I do (field of study, field of work etc.).

Students' sense of being sufficiently competent (that is, equipped with adequate knowledge and skills) was influenced by the greatest number of factors. These included parental trust in their ability to make independent decisions, age, gender, financial independence, and country of origin.

Higher levels of perceived competence were reported by those who felt trusted by their parents, respondents over the age of 25, financially independent individuals, and students born in Slovakia.

With respect to gender, women were less likely to rate their knowledge as weak (8% compared to 15% of men); however, men were slightly more likely to consider themselves strong in terms of knowledge (49% versus 46% of women).

Employment during studies was associated with more extreme responses, with working students reporting higher levels of perceived competence.

I continuously seek to learn and improve.

The only factor that significantly influenced responses to the question on continuous improvement was financial independence. Financially independent students reported striving for continuous development in 74% of cases, compared to only 56% among those who were financially dependent.

I take responsibility for the results of my work.

No statistically significant differences were observed between respondents in the area of responsibility with respect to any of the examined factors.

6 DISCUSSION

The findings of this study highlight several factors influencing respondents' self-perceived trustworthiness and related behaviors. Rather than being isolated, these factors appear to interact in shaping how individuals evaluate their own trust-related conduct. Gender consistently emerged as a significant factor: women are more likely than men to keep promises, express their intentions openly, and perceive their motives as honest, which supports existing research emphasizing communication clarity and reliability as central components of trust formation. Parental influence also plays a role, particularly the ability to rely on emotional support and perceived parental trust, which are associated with higher self-reported competence and reliability. This finding suggests that early interpersonal interactions may have lasting effects on individuals' confidence in fulfilling commitments and acting dependably in later life.

Age and financial independence influence specific behaviors, such as striving for continuous improvement and feeling competent in one's field, with older and financially independent respondents reporting higher levels. These patterns may reflect greater life experience, stability, and accumulated social or professional capital, which are commonly linked to stronger perceptions of personal efficacy and trustworthiness. Cross-national differences were observed, with Slovak respondents generally reporting stronger skills and motivations aligned with mutual benefit compared to Czech respondents. Although the study does not aim to explain these differences causally, they may point to broader cultural or institutional influences shaping norms of cooperation and trust-related behavior.

Living arrangements had some impact on the expression of intentions, with those living alone expressing themselves most confidently, while employment during studies was linked to higher perceived competence. Both findings suggest that autonomy and early labor-market exposure may enhance self-confidence in trust-related interactions. Finally, some areas, such as taking responsibility for work results, showed no significant differences across any factors, suggesting certain aspects of trustworthiness may be relatively stable regardless of demographic or contextual variables. This stability indicates that some components of trust-related behavior may be less sensitive to individual background characteristics and more strongly rooted in general norms or personal values.

Overall, these results underscore the complex interplay of gender, parental influence, age, financial independence, and cultural context in shaping self-perceived trustworthy behaviors, highlighting the importance of considering both individual life-course factors and broader social contexts when examining trust and trustworthiness.

An important consideration in interpreting the results concerns the subjective nature of the measures used in this study. The findings are based on self-reported assessments of trust-related behaviors and attitudes, which may not necessarily reflect actual behavior in real-life situations. For example, respondents' confidence in taking responsibility for their actions or their perceptions of being trusted by their parents may be influenced by self-esteem, social desirability, or individual self-perception rather than objective experience. Nevertheless, such subjective perceptions remain highly relevant, as an individual's perceived trustworthiness and sense of being trusted can directly shape their willingness to extend trust to others and to engage in trust-based relationships.

Another important factor relates to differences in life stages among the student population. Younger students, who are often in their first year of university studies, are typically in the early phase of becoming independent. For many, this period coincides with leaving the parental home for the first time and assuming responsibility for personal finances, living arrangements, and daily decision-making. These transitions may influence attitudes toward responsibility, material security, and trust-related behavior. In contrast, older students tend to have greater life experience, more stable financial and housing situations, and more precise professional trajectories, which may partly explain observed differences in trust-related perceptions and behaviors between younger and older respondents.

7 CONCLUSION

The results of this research confirmed statistically significant differences in trust across various segments of respondents. The findings indicate that women exhibit a stronger tendency to keep promises, suggesting a higher perceived level of trustworthiness than men. Moreover, trust appears to be more easily established with women, as they tend to communicate their intentions and motivations more clearly and consistently. These findings align with existing research (Covey & Merrill, 2006; Sinek, 2019) that emphasizes the role of transparent communication and reliability as key determinants of trust formation. The results also support the idea that trustworthiness is closely linked to integrity and consistency in behavior, which Mayer *et al.* (1995) identify as important components of trustworthiness.

Another factor contributing to an individual's trustworthiness and tendency to trust is whether they were able to rely on their parents during childhood and learned to develop mutual trust within close relationships. Early experiences of reliable parental support appear to shape later capacities for building, developing, and maintaining trust. This finding is consistent with previous research suggesting that trust develops through early social interactions and family environments (Giulietti *et al.*, 2023; Robbins, 2016). It also supports the argument that experiences of trust in childhood influence trust-related attitudes in later relationships (Covey & Merrill, 2006).

Furthermore, relationships with financially independent individuals tend to be associated with higher perceived trustworthiness than those involving financially dependent individuals. Financial independence may signal stability and predictability, characteristics that are commonly linked to trust formation. A possible explanation is that financially independent individuals are perceived as more responsible and capable of fulfilling commitments, which corresponds with the concepts of ability and competence discussed by Mayer *et al.* (1995) and Covey and Merrill (2006).

The results of this study should not be interpreted as implying that men who lacked a supportive upbringing in childhood or who are not financially independent are inherently untrustworthy. Trustworthiness is not a fixed personal attribute. It can be developed under appropriate motivation, guidance, and contextual conditions.

It is nevertheless important to acknowledge that individuals and their personal characteristics differ across a range of dimensions, including gender, childhood experiences, financial behavior, and cultural background. These factors influence an individual's predisposition toward trust and trustworthiness. Life partners, effective leaders, and mentors must therefore be aware of such differences and actively work with them. While these factors may be considered during team selection processes, they can also be addressed and leveraged throughout ongoing relationships to foster and strengthen mutual trust among team members and in other interpersonal contexts.

REFERENCES

Covey, S. M. R., Merrill, R. R. (2006). *The SPEED of Trust: The One Thing that Changes Everything*. New York: Simon and Schuster, 354p. ISBN:978-0-7432-9730-1

Diamond, D. W., Dybvig, P. H. (1983). Bank Runs, Deposit Insurance, and Liquidity. *Journal of Political Economy*. 91(3), 401–419. https://doi.org/10.1086/261155

Giulietti, C., Rettore, E., Tonini, S. (2023). The chips are down: The influence of family on children's trust formation. *Journal of Population Economics*. 36(1), 211–233. https://doi.org/10.1007/s00148-022-00921-1

Gottman, J. (2023). *What predicts divorce?: The relationship between marital processes and marital outcomes*. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9781003429807/predicts-divorce-john-gottman

Höhne, S., Paloncyová, J. (2021). Péče a výživa dětí po rozchodu rodičů podle výsledku opatrovnických řízení. *Demografie*. 63(1), 3–21.

Cherlin, A. J. (2004). The Deinstitutionalization of American Marriage. *Journal of Marriage and Family.* 66(4), 848–861.

Mayer, R. C., Davis, J. H., Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*. 20(3), 709–734. https://doi.org/10.2307/258792

McAllister, D. J. (1995). Affect- and Cognition-Based Trust as foundation for interpersonal cooperation in organizations. *Academy of Management Journal.* 38(1), 24–59. https://doi.org/10.2307/256727

Robbins, B. G. (2016). What is Trust? A Multidisciplinary Review, Critique, and Synthesis. *Sociology Compass.* 10(10), 972–986. https://doi.org/10.1111/soc4.12391

Sinek, S. (2019). *The Infinite Game.* New York: Penguin, 251 p. ISBN: 9780525538837

Contact information

Tereza Frömmelová: e-mail: tereza.frommelova@vse.cz

Kateřina Pojkarová: e-mail: katerina.pojkarova@upce.cz

FROM VALUES TO ACTION: EXPLAINING THE INTENTION-BEHAVIOR GAP IN GREEN FASHION CONSUMPTION AMONG GENERATION Z IN BANGLADESH

Shovan Nasim Haider[1], Miloslava Chovancova[2]

[1]Faculty of Management and Economics, Tomas Bata University in Zlin, Mostní 5139, 760 01 Zlín, Czech Republic

[2]Institute of Management and Marketing, Faculty of Management and Economics, Tomas Bata University in Zlin, Mostní 5139, 760 01 Zlín, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_008

ABSTRACT

This study aims to examine how green consumer values are translated into actual green fashion apparel purchasing behavior among Generation Z consumers in Bangladesh, with particular focus on explaining the persistent intention–behavior gap within a developing, apparel-producing economy. The study adopts a quantitative research design grounded in a value–intention–behavior framework. Drawing on the Theory of Consumption Values and the Theory of Planned Behavior, the proposed conceptual model incorporates green purchase intention as a central mediating mechanism, supported by post-intention constructs including implementation intention and differentiated self-efficacy (action and coping). Survey data are collected from Generation Z consumers in Bangladesh and analyzed using partial least squares structural equation modeling (PLS-SEM) to examine direct, mediated, and moderated relationships among the constructs. Building on the theoretical and empirical gaps identified in the literature, the study focuses on clarifying the mechanisms through which multidimensional consumer values influence green purchase intention and how intentions are subsequently enacted under contextual constraints. Particular attention is given to the roles of implementation intention and self-efficacy in strengthening intention–behavior consistency in sustainable fashion consumption. By moving beyond intention-centric explanations, this study is positioned to advance theoretical understanding of green consumption behavior and provide contextually grounded insights relevant to developing producer economies. The findings are expected to inform future research and support practitioners and policymakers in designing interventions that facilitate the translation of sustainability values into actual purchasing behavior among young consumers.

Keywords: Green fashion consumption; Value–Intention–Behavior; Theory of Consumption Values; Theory of Planned Behavior; Implementation intention; Self-efficacy; Generation Z; Bangladesh

JEL Code: M30, M31

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION AND RESEARCH MOTIVATION

Human consumption has long been central to economic development, yet its environmental and social consequences have become increasingly pronounced within globalized production systems (Zhang *et al.*, 2023). Among consumer industries, fashion apparel is widely recognized as one of the most resource-intensive and environmentally damaging sectors. The industry consumes vast quantities of water and energy, generates substantial greenhouse gas emissions, and produces significant waste throughout its lifecycle (Hasanbeigi, 2010; Niinimäki *et al.*, 2020; Shirvanimoghaddam *et al.*, 2020). In parallel, its global supply chains are frequently associated with labor rights violations and unsafe working conditions, particularly in buyer-driven value chains operating in developing economies (Islam & van Staden, 2021).

In response, sustainable or green fashion emphasizing eco-friendly materials, ethical production, and reduced resource use has emerged as a critical alternative (Elgammal *et al.*, 2024; Zhang *et al.*, 2023). However, despite increased awareness and positive consumer attitudes, adoption remains uneven. A central obstacle lies in the persistent discrepancy between consumers' environmental values, their stated intentions, and their actual purchasing behavior.

This challenge is especially pronounced among Generation Z consumers. Born between the mid-1990s and early 2010s, this cohort is frequently characterized as environmentally conscious, socially aware, and willing to support sustainable brands (Seyfi *et al.*, 2023; Lopes *et al.*, 2024). Empirical evidence suggests that a majority of Gen Z consumer's express preferences for sustainable products and a willingness to pay price premiums for ethical alternatives (Slepian *et al.*, 2024; Elgammal *et al.*, 2024). Yet, their actual consumption patterns often remain dominated by fast fashion, revealing a paradox between ethical orientation and behavioral outcomes (Carrington *et al.*, 2010; Caruana *et al.*, 2016).

This paradox motivates the present study. While prior research has documented the intention–behavior gap extensively, particularly in developed economies, limited attention has been paid to how this gap operates in developing producer economies where economic constraints, market structure, and cultural norms differ substantially.

2 ENVIRONMENTAL AND SOCIAL CHALLENGES OF THE FASHION APPAREL INDUSTRY

The fashion apparel industry's environmental footprint is substantial and multifaceted. Textile production requires intensive use of water, chemicals, and energy, while generating significant waste at multiple stages of the supply chain (Deshmukh *et al.*, 2024; Lorente-Leyva *et al.*, 2024; Qi *et al.*, 2021). Fiber production alone contributes considerably to global warming through airborne emissions and chemical runoff (Flanagan & Priyadarshini, 2021). Projections indicate that global textile waste could increase by over 60 percent by 2030, further intensifying environmental degradation (Niinimäki *et al.*, 2020; Shirvanimoghaddam *et al.*, 2020).

Beyond environmental harm, the industry is marked by persistent social externalities. Labor exploitation, wage suppression, and unsafe working conditions are structurally embedded in global apparel supply chains, particularly in export-oriented developing economies (Islam & van Staden, 2021). These challenges underscore the urgency of sustainable transformation not only at the production level but also at the consumption level, where demand patterns reinforce existing production practices.

3 THEORETICAL FOUNDATIONS OF GREEN CONSUMPTION RESEARCH

3.1 Values, Attitudes, and Intention Formation

Ethical and sustainable consumption has been a subject of scholarly inquiry since the late 1980s (Witkowski & Reddy, 2010). Early theoretical contributions were shaped by the Theory of Reasoned Action (Ajzen & Fishbein, 1970, 1980) and later the Theory of Planned Behavior (Ajzen, 1991), which positioned behavioral intention as the most immediate predictor of action. TPB has been widely applied in green consumption contexts, including organic food purchasing, fair--trade consumption, and environmentally responsible behaviors (Al-Swidi *et al.*, 2014; Paul *et al.*, 2016; Zerbini *et al.*, 2019).

While TPB provides valuable insight into intention formation, its explanatory power weakens in complex consumption domains such as fashion, where symbolic meaning, identity expression, and social visibility play central roles (Harris *et al.*, 2015; Gazzola *et al.*, 2020).

The Theory of Consumption Values (Sheth *et al.*, 1991) complements intention-based models by emphasizing the multidimensional nature of consumer values, including functional, emotional, social, epistemic, and conditional values. Recent studies increasingly apply TCV to sustainable fashion contexts, demonstrating how competing values influence green purchase intention (Kang *et al.*, 2013; Cho *et al.*, 2015; Sharma *et al.*, 2024). However, TCV-based studies often stop at intention formation and offer limited insight into behavioral enactment.

3.2 The Intention–Behavior Gap and Post-Intention Processes

Despite extensive application of TPB and TCV, empirical research consistently documents a gap between ethical values, intentions, and actual behavior (Carrington *et al.*, 2010; Caruana *et al.*, 2016). This intention–behavior gap is especially pronounced in fashion due to affordability constraints, limited availability of sustainable alternatives, and entrenched fast-fashion norms (Vermeir & Verbeke, 2006; Gazzola *et al.*, 2020).

Implementation intention theory addresses this limitation by emphasizing concrete planning regarding when, where, and how a behavior will occur (Gollwitzer, 1993, 1999). Empirical evidence suggests that implementation intentions strengthen intention–behavior consistency, yet their application in green fashion research remains limited (Carrington *et al.*, 2014; Tawde *et al.*, 2023).

Similarly, self-efficacy plays a critical role in behavioral enactment. Distinguishing between action self-efficacy and coping self-efficacy is particularly relevant in constrained environments where consumers must overcome situational barriers such as price and access (Bandura, 1997; Schwarzer, 2008). These distinctions remain underexplored in sustainable fashion research.

4 GENERATION Z AND THE BANGLADESHI CONTEXT

4.1 Generational Values and Behavioral Tendencies

Generational Cohort Theory posits that individuals who come of age during the same historical epoch develop shared value orientations and behavioral dispositions through exposure to common social, technological, and political conditions (Mannheim, 1952; Strauss & Howe, 1991). Generation Z, defined as individuals born between the mid-1990s and early 2010s, has matured during a period characterized by heightened environmental discourse, digital hyper--connectivity, and growing awareness of social and institutional accountability. This formative context distinguishes Generation Z from preceding cohorts and is associated with relatively stronger orientation toward environmental concern, ethical brand evaluation, and the expression of personal values through consumption decisions (Seyfi *et al.*, 2023; Elgammal *et al.*, 2024). Empirical evidence suggests that a substantial proportion of Gen Z consumer's express preferences for sustainable products and a willingness to accept price premiums for ethically sourced alternatives, a pattern documented across multiple national contexts (Elgammal *et al.*, 2024; Lopes *et al.*, 2024).

Despite these broad characterizations, generational value orientations do not translate uniformly into consumption behavior across contexts. Behavioral enactment remains contingent on structural, economic, and socio-cultural factors that vary substantially across national and regional settings (Mammadli, 2023; Jung *et al.*, 2021). In developing economies where purchasing power is constrained and sustainable product availability remains limited, the relationship between generational values and actual consumption behavior is subject to additional mediating pressures that the cohort literature frequently underspecifies. Accordingly, characterizations of Generation Z as inherently sustainability-oriented consumers require contextual qualification before theoretical models grounded in that assumption can be applied with confidence.

4.2 The Bangladeshi Socio-Economic Context

In Bangladesh, Generation Z constitutes approximately 40 percent of the national population and represents an increasingly urbanized and digitally engaged demographic cohort (Islam *et al.*, 2021). Their consumption behavior is, however, shaped by socio-economic conditions that differ substantially from those prevailing in high-income economies. Household income constraints, limited access to certified sustainable apparel, and a retail environment dominated by fast-fashion products combine to restrict the effective translation of environmental concern into purchasing behavior (Biswas & Roy, 2015; Hasan *et al.*, 2022). Although Bangladeshi Gen Z consumers express concern for environmental and labor rights issues, particularly given their proximity to the garment manufacturing sector, empirical evidence consistently documents a persistent value-action gap driven by structural and affordability barriers (Sobuj *et al.*, 2021).

Bangladesh occupies a distinctive position as the world's second-largest apparel exporter (BGMEA Report, 2024). This dual identity as both a major production economy and an emerging consumer market creates a specific social and perceptual landscape for sustainability. Consumers in this setting may possess a more direct awareness of production externalities than their counterparts in purely import-dependent economies, yet this awareness does not automatically translate into purchasing power or access to credible sustainable alternatives.

These structural realities reinforce the central theoretical argument of this study: that purchase intention alone is insufficient to predict behavioral outcomes and that volitional mechanisms, including implementation intention and differentiated self-efficacy, require explicit theoretical attention if the full intention-behavior pathway is to be understood.

4.3 Collectivist Cultural Norms, Peer Influence, and Social Identity

A dimension of the Bangladeshi context that receives insufficient attention in existing green consumption research is its collectivist cultural orientation. Bangladesh ranks among the more collectivist societies globally, characterized by strong in-group loyalty, adherence to group-level norms, and a tendency to prioritize relational identity over individual autonomy in consequential decisions (Hofstede, 2001). In such cultural environments, consumption decisions are rarely purely individual acts; they are embedded in relational networks that encompass family structures, peer groups, and community expectations. Fashion consumption in particular carries significant social and symbolic meaning, and purchasing decisions are frequently evaluated in terms of their alignment with group norms rather than solely through individual value assessment. This contextual reality fundamentally shapes the operational dynamics of consumer behavior models derived from individualist theoretical traditions.

This cultural reality has direct implications for how the Theory of Planned Behavior operates within the Bangladeshi context. The subjective norms component of TPB, which captures perceived social pressure from significant others to engage in or abstain from a given behavior, assumes particular explanatory salience in collectivist societies where the opinions and behavioral patterns of family members, peers, and social reference groups exert strong normative pressure on individual consumption decisions (Al-Swidi *et al.*, 2014). Generation Z consumers in Bangladesh may hold positive attitudes toward green fashion and have formed genuine purchase intentions, yet may nonetheless defer to group consumption norms that prioritize price conformity, trend alignment, or family expectations regarding spending behavior. The result is a socially mediated form of the intention-behavior gap that compounds the economic and structural barriers already identified in the literature and that existing frameworks have insufficiently theorized.

Peer influence constitutes a particularly potent mechanism among Generation Z consumers, whose identity formation is closely tied to digital social networks and peer-validated consumption practices (Chaturvedi *et al.*, 2020; Seyfi *et al.*, 2023). In collectivist cultural contexts such as Bangladesh, the social visibility of fashion consumption amplifies the role of peer norms in shaping actual purchasing behavior. Green fashion adoption may be facilitated where peer groups positively validate sustainable choices as markers of status or moral positioning, but may be inhibited where fast-fashion consumption functions as the dominant social currency through which belonging and peer acceptance are communicated. This dynamic suggests that the volitional mechanisms through which implementation intention and self-efficacy enable behavioral enactment must be understood not merely as individual psychological resources but as behavioral capacities fundamentally shaped by the social environments within which consumption decisions are embedded.

From a theoretical standpoint, these contextual dynamics underscore the need for an integrative framework that moves beyond the individualist assumptions often implicit in standard TPB applications. Social identity, group belonging, and normative reference group pressure are not peripheral variables in the Bangladeshi green fashion context; they are constitutive forces that shape the translation of values into behavioral plans and ultimately into purchasing acts.

By situating Generation Z green fashion consumption within this culturally embedded and structurally constrained context, the present study contributes a contextually grounded extension of existing theoretical accounts and positions its findings to speak meaningfully to the growing body of literature that calls for context-sensitive models of sustainable consumption in developing and producer economies (Dermody *et al.*, 2018; Zou & Chan, 2019; Ngo *et al.*, 2023).

5 RESEARCH GAPS AND STUDY POSITIONING

The literature reveals several interrelated gaps. First, theoretical models remain overly intention-centric and under-specify post-intention mechanisms. Second, consumer values are often treated narrowly, overlooking their multidimensional and sometimes conflicting nature. Third, empirical research is heavily biased toward developed economies, limiting contextual generalizability. Finally, methodological approaches rarely model mediated and moderated behavioral pathways simultaneously.

Table 1 summarizes the primary theoretical, mechanistic, contextual, and empirical gaps identified in the literature.

5.1 The persistent intention–behavior gap in green fashion consumption

The intention–behavior gap has been widely documented in green consumption research, yet its underlying mechanisms remain insufficiently specified. Studies consistently show that consumers who express strong environmental values and positive attitudes toward sustainable fashion often fail to act in accordance with those intentions (Carrington *et al.*, 2010; Caruana *et al.*, 2016; Song *et al.*, 2020; Tandon *et al.*, 2023). This gap is particularly pronounced in fashion consumption due to the sector's symbolic significance, social visibility, and emotional attachment (Harris *et al.*, 2015; Gazzola *et al.*, 2020).

A critical limitation of existing models lies in their insufficient attention to post-intention processes. Research has largely overlooked the self-regulatory and volitional mechanisms required to translate intention into action, particularly in contexts characterized by affordability constraints, limited access to sustainable alternatives, and strong fast-fashion norms. Implementation intention

Tab. 1 Table 1: Preliminary identification of the research gaps in contemporary research landscape

Gap	Reference
Theoretical gap	Sharma et al., 2025; Jaiswal and Kant, 2018; Ajzen, 1991; Sheth *et al.*, 1991
Mechanistic gap	Carrington *et al.*, 2014; Gollwitzer, 1999; Tawde *et al.*, 2023; Johnstone and Tan, 2015a
Contextual gap	Jung, Kim, and Oh, 2021; Lundblad and Davies, 2016; Sharma et al., 2025; Dermody *et al.*, 2018; Zou and Chan, 2019
Geographical gap	Seyfi *et al.*, 2023; Nguyen *et al.*, 2023; Mammadli, 2023; Ericksson, 2009; Srinivasan, 2012; Thangavel, Pathak, Chadra, 2021
Empirical gap	Tanrikulu, 2021; Wu & Lee, 2025

theory suggests that specifying concrete plans related to when, where, and how a behavior will be performed significantly increases the likelihood of behavioral enactment (Gollwitzer, 1993; Gollwitzer, 1999; Carrington *et al.*, 2010). However, empirical investigations of implementation intentions within green apparel consumption remain limited.

Similarly, self-efficacy has been applied inconsistently and often generically. Distinctions between action self-efficacy, which reflects confidence in initiating behavior, and coping self-efficacy, which reflects confidence in overcoming situational barriers, are rarely made explicit despite their relevance in constrained consumption environments (Bandura, 1997; Schwarzer, 2008). The absence of such differentiation limits understanding of why consumers with similar intentions demonstrate divergent behavioral outcomes (Khare & Sadachar, 2017; Wiederhold & Martinez, 2018; Tawde *et al.*, 2023).

5.2 Theoretical Fragmentation and conceptual limitations

Beyond the intention–behavior gap itself, the literature is characterized by substantial theoretical fragmentation. Many studies employ single-theory frameworks, such as the Theory of Planned Behavior or the Theory of Consumption Values, in isolation. While research grounded in the Theory of Consumption Values highlights the multidimensional nature of consumer values, including functional, emotional, social, and green values, it often stops at intention formation and does not adequately explain how these values are reconciled during actual purchasing decisions (Sheth *et al.*, 1991; Kang *et al.*, 2013; Cho *et al.*, 2015; Sharma *et al.*, 2024).

At the same time, intention-focused models frequently adopt a reductionist view of motivation, emphasizing environmental concern while underestimating the interplay of economic constraints, social norms, and market realities. This conceptual narrowness is especially problematic in fashion consumption, where symbolic meaning and social identity play a central role (Harris *et al.*, 2015; Gazzola *et al.*, 2020). Consequently, existing models struggle to explain the behavioral paradox in which consumers endorse sustainability values while continuing to engage in fast-fashion consumption (Maichum *et al.*, 2016; Hasan *et al.*, 2022).

Although some recent studies have attempted to integrate multiple frameworks, including combinations of TPB, MOA, and social identity perspectives, such efforts remain limited and rarely tailored to the socio-cultural realities of developing economies such as Bangladesh (Seyfi *et al.*, 2023; Nguyen *et al.*, 2019). As a result, the relationship between multidimensional values, intentions, and actual green apparel purchasing behavior remains under-explored within cohesive and context-sensitive models.

5.3 Contextual, Geographical, and Gaps

The explanatory limitations of existing models are compounded by a pronounced contextual and geographical bias. The majority of green consumption studies are rooted in Western, developed--economy contexts, raising concerns regarding the generalizability of findings to developing and producer economies (Dermody *et al.*, 2018; Zou & Chan, 2019). There is a notable lack of empirical research incorporating the socio-economic constraints, cultural norms, and market structures characteristic of countries such as Bangladesh (Islam *et al.*, 2021; Ngo *et al.*, 2023; Mammadli, 2023).

Bangladesh represents a distinctive context as the world's second-largest apparel exporter (BGMEA Report, 2024). Consumers in this setting are directly exposed to the environmental and social consequences of garment production, which may shape sustainability perceptions

differently from those in consumer-only economies (Islam & van Staden, 2021). At the same time, affordability pressures, limited availability of sustainable apparel, and strong fast-fashion norms create structural barriers that intensify the intention–behavior gap (Vermeir & Verbeke, 2006; Gazzola *et al.*, 2020; Hasan *et al.*, 2022).

The assumption of globally homogeneous generational behavior further constrains existing research. While Generational Cohort Theory suggests that shared historical experiences shape collective values (Mannheim, 1952; Strauss & Howe, 1991), the translation of these values into behavior is fundamentally mediated by local economic realities and cultural norms. For Generation Z consumers in Bangladesh, price sensitivity, accessibility, and social influence often exert a stronger effect on purchasing decisions than abstract environmental concern (Biswas & Roy, 2015; Islam *et al.*, 2021; Chaturvedi *et al.*, 2020). In collectivist societies such as Bangladesh, group norms and social identity may supersede individual attitudes, yet these dynamics remain under-examined (Hofstede, 2001).

5.4 Methodological and Synthesis Gaps

Methodologically, prior studies frequently rely on simplified predictive models and cross-sectional designs that are ill-equipped to capture the conditional and mediated pathways through which values are translated into behavior (Testa *et al.*, 2020). The dynamic roles of implementation intentions as mediators, as well as the moderating functions of action and coping self-efficacy within culturally and economically constrained environments, remain insufficiently examined (Essiz *et al.*, 2022).

Taken together, the literature reveals a fragmented landscape marked by over-reliance on intention-centric explanations, conceptual narrowness regarding consumer values, contextual bias toward developed economies, limited attention to digital and social mediation, and methodological constraints. These interconnected deficiencies restrict understanding of how Generation Z consumers navigate the transition from values and intentions to actual green fashion purchasing behavior, particularly within developing producer economies.

5.5 Contribution of the Present Study

In synthesis and research positioning, the literature reveals a fragmented research landscape marked by four interrelated deficiencies:

(1) over-reliance on intention-centric models lacking post-intention mechanisms (Sheeran, 2002);
(2) conceptual narrowness that underplays the multidimensional and often conflicting nature of consumer values (Sheth *et al.*, 1991; Kang *et al.*, 2013; Cho *et al.*, 2015);
(3) contextual bias toward developed economies with limited applicability to producer nations (Carrete *et al.*, 2012); and
(4) methodological approaches that fail to model complex behavioral enactment processes.

Addressing these gaps necessitates an integrated, context-sensitive framework that simultaneously explains how values shape intentions and how specific volitional mechanisms enable or inhibit the translation of intention into action under real-world constraints.

This study responds to these challenges by proposing and empirically testing a Value-Intention-Planning-Behavior framework that integrates value antecedents (TCV), intention formation (TPB), and post-intention mechanisms (implementation intentions and differentiated self-efficacy).

Grounded in the lived realities of Generation Z consumers in Bangladesh, a critical yet under-researched apparel producer economy, the study advances theoretical clarity, contextual relevance, and methodological rigor. Methodologically, it employs variance-based structural equation modeling (PLS-SEM) to explicitly test mediated and moderated pathways, moving beyond intention-only explanations toward a more holistic account of sustainable fashion consumption enactment.

This study addresses these gaps by proposing an integrated value–intention–planning–behavior framework that combines TCV, TPB, implementation intention, and differentiated self-efficacy within a developing producer-economy context.

6 RESEARCH OBJECTIVES

The study aims to examine how green consumer values and purchase intentions are translated into actual green fashion purchasing behavior among Generation Z consumers in Bangladesh. Specifically, it seeks to:

- *Examine the influence of multidimensional consumer values on green fashion purchase intention.*
- *Assess the effect of purchase intention on actual green fashion purchasing behavior.*
- *Investigate the mediating role of implementation intention.*
- *Examine the moderating roles of action and coping self-efficacy.*
- *Evaluate how contextual constraints shape the persistence of the intention–behavior gap.*

7 METHODOLOGICAL OVERVIEW

7.1 Research Design

The study employs a quantitative, cross-sectional research design using a structured self-administered online survey instrument. This design is appropriate for testing theoretically derived hypotheses regarding the structural relationships among consumer values, green purchase intention, implementation intention, phase-specific self-efficacy, and actual green purchasing behavior. Quantitative survey methodology enables the standardized measurement of latent constructs across a sufficiently large sample to support variance-based structural equation modeling. A cross--sectional design provides an efficient framework for capturing attitudinal, intentional, and behavioral measures within the target population at a single time point. The limitations associated with cross-sectional causal inference are acknowledged and discussed in Section 8.

7.2 Target Population, Sampling Strategy, and Sample Size

The target population comprises Generation Z consumers in Bangladesh, operationally defined as individuals born between 1997 and 2012 and residing in major urban centers, including Dhaka, Chittagong, and Sylhet. Urban Generation Z consumers constitute the primary sampling frame because they exhibit relatively higher exposure to sustainable fashion discourse, greater engagement with digital retail environments, and stronger awareness of environmental and ethical consumption issues compared with rural populations (Hasan *et al.*, 2022; Chaturvedi *et al.*, 2020). These urban characteristics also facilitate structured online data collection, which is consistent with the digital engagement patterns characteristic of the target cohort.

A non-probability purposive sampling strategy is employed and supplemented by snowball sampling to extend reach within the target demographic. Survey instruments are distributed through university networks, social media platforms, and peer referral mechanisms, consistent with established methodological approaches in comparable studies of Generation Z sustainable consumption behavior in developing economies (Sobuj *et al.*, 2021; Tandon *et al.*, 2023). A minimum target of 350 completed usable responses is established. This target is determined in accordance with the minimum sample size guidelines for partial least squares structural equation modeling, which specify that the sample should be at least ten times the maximum number of structural paths directed at any single latent construct in the model (Hair *et al.*, 2019). Given the model's structural complexity, which includes serial mediation pathways and two phase-specific moderation effects, a target of 350 responses provides adequate statistical power for detecting medium effect sizes at the conventional 0.05 significance threshold.

7.3 Measurement Instruments and Scales

All constructs are operationalized using established multi-item reflective scales adapted from prior validated instruments in the green consumption, social cognition, and behavioral planning literatures. Items are rated on a seven-point Likert scale anchored at 1 (strongly disagree) and 7 (strongly agree), a scaling format demonstrated to improve measurement sensitivity and distributional variance relative to narrower five-point alternatives.

Green Consumer Value (GCV) is measured using items adapted from the Theory of Consumption Values (Sheth *et al.*, 1991) and extended by Kang *et al.* (2013) and Cho *et al.* (2015) to encompass the functional, emotional, social, epistemic, and conditional value dimensions relevant to sustainable apparel consumption. GCV is operationalized as a second-order reflective construct, with each of the five value dimensions treated as first-order reflective components. Green Purchase Intention (GPI) is measured using items adapted from Ajzen (1991) and Paul *et al.* (2016), capturing respondents' self-reported likelihood of purchasing green fashion apparel within a specified near-term horizon. Implementation Intention (II) is operationalized using items adapted from Gollwitzer (1999) and Carrington *et al.* (2014), assessing the specificity of respondents' behavioral plans regarding when, where, and how they intend to execute a green fashion purchase. Action Self-Efficacy (ASE) and Coping Self-Efficacy (CSE) are measured using instruments derived from Schwarzer (2008) and Bandura (1997), with ASE items assessing confidence in initiating green fashion purchasing and CSE items assessing confidence in overcoming situational barriers including price constraints and limited product availability. Actual Green Purchase Behavior (GPB) is captured using behavioral frequency items adapted from Paul *et al.* (2016) and Tandon *et al.* (2023), reflecting the recency and frequency of respondents' actual green apparel purchases over a specified retrospective recall period.

To ensure content validity, all adapted items are reviewed by two independent experts in consumer behavior prior to pilot testing. A pilot study involving 30 respondents drawn from the target population is conducted to assess item clarity, response distribution, and preliminary reliability indicators before full-scale data collection commences.

7.4 Integration of TCV and TPB Constructs in the Empirical Model

The empirical model integrates the Theory of Consumption Values and the Theory of Planned Behavior through a theoretically coherent sequential pathway that preserves the explanatory logic of both frameworks while addressing their respective limitations. TCV provides the

multidimensional value substrate that motivates sustainable consumption orientation. The five value dimensions identified by Sheth *et al.* (1991) are operationalized as reflective components of a second-order Green Consumer Value (GCV) construct, which represents the latent consumer value orientation predisposing individuals toward green fashion purchasing. GCV thus constitutes the theoretically grounded antecedent through which TCV is integrated into the behavioral pathway.

TPB then supplies the mechanism through which value orientation is translated into behavioral motivation. GCV is hypothesized to predict Green Purchase Intention, the individual's motivational readiness to engage in green fashion purchasing, consistent with TPB's positioning of intention as the most proximal cognitive antecedent of behavior (Ajzen, 1991). Within the Bangladeshi collectivist context elaborated in Section 4.3, the subjective norms dimension of TPB is expected to assume particular explanatory weight, reflecting the influence of peer groups, family expectations, and community-level consumption norms on the formation of purchase intentions. This integration ensures that TCV and TPB are not merely juxtaposed but are connected through a coherent causal chain: value formation drives attitude and normative beliefs, which jointly shape purchase intention, which in turn activates the post-intention enactment process.

7.5 Role and Positioning of Self-Efficacy Within the Framework

Self-efficacy is positioned as a phase-specific moderating variable operating at two critical behavioral transition points within the post-intention pathway. This phase-specific positioning draws on the distinction between action self-efficacy and coping self-efficacy elaborated by Schwarzer (2008) within the Health Action Process Approach, which has received limited empirical application in sustainable fashion research despite its demonstrated relevance in constrained consumption environments (Tawde *et al.*, 2023).

Action Self-Efficacy (ASE) moderates the relationship between Green Purchase Intention and Implementation Intention. This moderation captures the process by which consumers who possess greater confidence in their capacity to initiate green purchasing behavior are more likely to convert their motivational states into concrete behavioral plans specifying when, where, and how the purchase will be made. Consumers with low ASE may maintain positive purchase intentions without translating them into specific implementation commitments, thereby contributing to the intention-behavior gap at the planning stage. Coping Self-Efficacy (CSE) moderates the relationship between Implementation Intention and actual Green Purchase Behavior. CSE reflects confidence in managing the situational obstacles, including price premiums, limited retail availability of sustainable apparel, and social norm incongruence, that can prevent behavioral plans from being executed. Consumers with higher CSE are hypothesized to more effectively activate their implementation plans even in the presence of such barriers.

The full structural pathway proposed by the study is therefore as follows: GCV predicts GPI (direct effect); GPI predicts Implementation Intention, moderated by ASE; Implementation Intention predicts GPB, moderated by CSE. The GCV to GPB relationship is thus mediated serially through GPI and Implementation Intention, with self-efficacy regulating the strength of the mediated pathway at phase-specific transition points. This structural arrangement directly addresses the reviewer's request for explicit specification of self-efficacy's positioning and is visually represented in the conceptual model diagram accompanying the original manuscript.

7.6 Construct Validation Procedures

Construct validity is assessed through a systematic two-stage analytical procedure. In the first stage, confirmatory factor analysis is conducted on each latent construct prior to structural model estimation. Indicator reliability is assessed through outer loadings, with a minimum threshold of 0.70 for acceptable item retention. Internal consistency reliability is evaluated using Cronbach's alpha and composite reliability (CR), with values exceeding 0.70 and 0.80, respectively, indicating adequate reliability. Convergent validity is established through average variance extracted (AVE), with values above 0.50 confirming that each construct explains more variance than measurement error.

Discriminant validity is assessed using the heterotrait-monotrait (HTMT) ratio of correlations criterion, with HTMT values below 0.90 considered indicative of adequate discriminant validity (Henseler *et al.*, 2015). The HTMT criterion is preferred over the traditional Fornell-Larcker criterion given evidence that the latter produces inflated assessments of discriminant validity in variance-based structural equation models. Where HTMT values approach or exceed 0.85, additional scrutiny of item cross-loadings is applied to confirm construct boundaries.

7.7 Analytical Procedure

Partial least squares structural equation modeling (PLS-SEM) is employed as the primary analytical method, implemented using SmartPLS software. PLS-SEM is selected over covariance-based structural equation modeling given the study's theoretical orientation toward prediction and explanation, its complex model structure incorporating higher-order constructs and interaction terms, and its comparative flexibility with respect to distributional assumptions (Hair *et al.*, 2019). Analysis proceeds in the two-stage sequence described above, with measurement model validation preceding structural model estimation.

At the structural stage, path coefficients, coefficients of determination (R-squared), effect sizes (f-squared), and predictive relevance (Q-squared via blindfolding, Stone-Geisser criterion) are reported. Serial mediation is tested using bias-corrected bootstrapping procedures with 5,000 re-samples to construct 95% confidence intervals for specific and total indirect effects. Moderation effects are examined by including the interaction term between each moderating variable and its corresponding predictor construct, with the change in R-squared attributable to the interaction term reported as the effect size. To address potential common method variance arising from the single-source, single-time-point design, procedural remedies are applied during survey design: attitudinal and behavioral items are separated within the instrument, reverse-coded items are included, and respondent anonymity is explicitly assured. Harman's single-factor test is applied as a post-hoc diagnostic check for common method bias.

8 CONCLUSION AND IMPLICATIONS

8.1 Summary and Theoretical Significance

This study addresses the persistent intention-behavior gap in green fashion consumption among Generation Z consumers in Bangladesh by advancing an integrated value-intention--planning-behavior framework that incorporates post-intention volitional mechanisms absent from conventional models. In responding to well-documented limitations of intention-centric approaches in sustainable consumption research, the study does not merely replicate existing theoretical configurations; it advances a framework that connects multidimensional consumer values to green purchase intention and subsequently to actual purchasing behavior through a structured series of mediating and moderating mechanisms. By positioning implementation intention as the critical behavioral planning mechanism and differentiated self-efficacy as the phase-specific regulatory variable, the research directly addresses the mechanistic and theoretical gaps identified in the literature and responds to longstanding calls to move beyond static, motivation-only accounts of ethical consumption toward more dynamic explanations of behavioral enactment.

8.2 Anticipated Theoretical Contributions and Extensions to the Literature

The proposed framework is positioned to make several distinct contributions to the existing theoretical literature on sustainable fashion consumption and the intention-behavior gap. These anticipated contributions are presented in relation to the specific theoretical claims they extend, qualify, or challenge.

First, the study challenges the prevailing assumption that purchase intention constitutes a sufficient proximal predictor of consumption behavior. The bulk of empirical research grounded in the Theory of Planned Behavior treats behavioral intention as the primary and often exclusive determinant of action, thereby treating the mechanism through which intentions are converted into behavior as theoretically transparent (Ajzen, 1991; Carrington *et al.*, 2010). The proposed framework argues, and is expected to demonstrate empirically, that motivational readiness and behavioral execution are separable psychological stages, each of which is sensitive to distinct enabling or inhibiting conditions. By demonstrating that implementation intention mediates the relationship between green purchase intention and actual green purchasing behavior, the study extends intention-based models by clarifying the volitional mechanism that bridges the motivation--action divide, a distinction that existing TPB-based applications in sustainable fashion research have largely elided (Carrington *et al.*, 2014; Tawde *et al.*, 2023).

Second, the differentiated treatment of self-efficacy as a phase-specific moderating mechanism advances a conceptual refinement absent from most existing sustainable fashion research. Prior studies have typically applied general self-efficacy as either a direct predictor or an undifferentiated moderator of intention-behavior linkages, without distinguishing the functional role of efficacy beliefs at distinct stages of behavioral enactment (Khare & Sadachar, 2017; Wiederhold & Martinez, 2018). By separately modeling action self-efficacy at the intention-to-planning transition and coping self-efficacy at the planning-to-behavior transition, the study provides a more granular account of why consumers who hold equivalent intentions and equivalent behavioral plans nonetheless produce divergent purchasing outcomes. This distinction extends Schwarzer's (2008)

phase-specific efficacy theory into the sustainable consumption domain, offering empirical evidence for its applicability in a constrained developing-economy context where situational barriers are particularly salient.

Third, the integration of the Theory of Consumption Values and the Theory of Planned Behavior within a single structural model extends both frameworks beyond their conventional application boundaries. TCV has been extensively applied to explain intention formation in sustainable consumption contexts but has rarely been extended to account for behavioral enactment or to engage with post-intention volitional processes (Tanrikulu, 2021; Sharma & Kushwah, 2025). TPB, conversely, has been applied primarily as an intention prediction model without adequately accounting for the multidimensional value substrate that shapes the attitudinal foundation of intention formation. The proposed integration addresses both limitations by establishing GCV as the value antecedent that activates the TPB mechanism and by connecting that mechanism to the post-intention enactment pathway. Consequently, the study is positioned to provide one of the more complete accounts of the full value-to-behavior pathway in sustainable fashion research.

Fourth, the study challenges the contextual portability assumption that is often implicit in sustainable consumption research, namely the assumption that findings generated in developed--economy contexts generalize directly to developing and producer economies. By empirically testing the model in Bangladesh, where structural constraints, collectivist cultural norms, and market access conditions differ substantially from those in Western consumer markets, the study anticipates findings that may diverge meaningfully from established theoretical predictions. The role of subjective norms within TPB may prove to be more pronounced than attitudinal determinants in the Bangladeshi context, consistent with the collectivist cultural dynamics elaborated in Section 4.3 (Hofstede, 2001; Al-Swidi *et al*., 2014). Similarly, coping self-efficacy may exert a stronger moderating effect on behavioral enactment in this context given the structural barriers, particularly price premiums and limited sustainable product availability, that characterize the Bangladeshi sustainable apparel market. Where such divergences from established findings emerge, they carry meaningful implications for the boundary conditions of existing theory and for the need to develop context-sensitive conceptual models.

8.3 Practical and Policy Implications

Beyond its theoretical contributions, the study is designed to generate actionable insights for practitioners and policymakers seeking to facilitate green fashion purchasing among young consumers in developing economies. For sustainable fashion brands and marketing practitioners, the framework suggests that campaigns targeting environmental awareness and attitude formation are necessary but insufficient conditions for behavioral change. If implementation intention mediates the intention-behavior pathway, then marketing interventions that facilitate concrete behavioral planning, including purchase commitment tools, targeted reminders triggered at relevant times and retail contexts, and decision scaffolding that reduces friction at the point of purchase, are likely to prove more effective than attitude-focused communication alone. Such approaches operationalize the implementation intention mechanism and are consistent with behavioral activation strategies demonstrated to be effective in analogous consumption domains (Gollwitzer, 1999).

For policymakers in Bangladesh and comparable developing economies, the study underscores the structural and relational dimensions of the intention-behavior gap. Interventions that enhance the affordability and accessibility of sustainable apparel, through certification programs, import duty structures that favor sustainable materials, or public procurement commitments that expand

sustainable market supply, directly address the situational barriers that undermine coping self--efficacy and thereby prevent consumers from enacting existing purchase intentions. Given the collectivist cultural dynamics identified in Section 4.3, policy efforts that engage peer networks, community organizations, and digitally prominent reference groups may be particularly effective in creating the social normative conditions under which sustainable fashion purchasing becomes the behaviorally and socially supported default among Generation Z consumers in Bangladesh.

8.4 Limitations and Directions for Future Research

Several limitations of the present study should be acknowledged. First, the cross-sectional research design restricts causal inference and limits the ability to observe how values, intentions, and behaviors evolve as sustainable fashion markets mature. Second, reliance on self-reported behavioral measures introduces potential response bias, particularly in socially desirable domains such as sustainable consumption. Third, the purposive urban sampling strategy may introduce selection effects that limit the representativeness of findings across the broader Generation Z population in Bangladesh, including less digitally connected or lower-income segments for whom structural barriers are even more severe. Fourth, while the study develops and tests the proposed framework in the Bangladeshi context, cross-cultural generalizability cannot be assumed without replication in comparable developing and producer economies.

These limitations suggest several productive directions for future research. Longitudinal designs would provide deeper insight into the temporal dynamics of the value-intention-behavior relationship and would permit stronger causal inference regarding the mechanisms identified in the proposed model. Comparative studies across multiple developing economies would support the systematic assessment of contextual moderators, including collectivist norms, structural market conditions, and generational cohort effects on the value-to-behavior pathway. Future research may also extend the framework by incorporating digital social influence mechanisms, including social media exposure, influencer-mediated consumption norms, and electronic word-of-mouth, which constitute particularly salient behavioral drivers for Generation Z consumers in digitally connected markets (Seyfi *et al.*, 2023; Chaturvedi *et al.*, 2020). Qualitative or mixed-methods research designs would further enrich understanding of the psychological processes and social dynamics through which implementation intentions are formed and executed in collectivist consumer contexts.

8.5 Concluding Statement

This study contributes to sustainable consumption research by proposing and empirically testing a theoretically integrated and context-sensitive account of green fashion purchasing behavior among Generation Z consumers in Bangladesh. By foregrounding the volitional mechanisms, specifically implementation intention and phase-specific self-efficacy, that either enable or constrain the translation of values into action, it underscores the importance of moving beyond awareness--driven and attitude-focused explanations toward interventions that actively support planning, self-regulatory confidence, and behavioral capability. Such an approach is essential for both advancing scholarly understanding of sustainable consumption enactment and for designing practical and policy interventions capable of fostering meaningful sustainable transformation within the global fashion industry.

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*. 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I., Fishbein, M. (1970). The prediction of behavior from attitudinal and normative variables. *Journal of Experimental Social Psychology*. 6(4), 466–487. https://doi.org/10.1016/0022-1031(70)90057-0

Al-Swidi, A., Mohammed Rafiul Huque, S., Haroon Hafeez, M., Noor Mohd Shariff, M. (2014). The role of subjective norms in theory of planned behavior in the context of organic food consumption. *British Food Journal*. 116(10), 1561–1580. https://doi.org/10.1108/BFJ-05-2013-0105

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*. 84(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Biswas, A., Roy, M. (2015). Green products: An exploratory study on the consumer behaviour in emerging economies of the East. *Journal of Cleaner Production*. 87, 463–468. https://doi.org/10.1016/j.jclepro.2014.09.075

Carrete, L., Castaño, R., Felix, R., Centeno, E., González, E. (2012). Green consumer behavior in an emerging economy: Confusion, credibility, and compatibility. *Journal of Consumer Marketing*. 29(7), 470–481. https://doi.org/10.1108/07363761211274983

Carrington, M. J., Neville, B. A., Whitwell, G. J. (2010). Why Ethical Consumers Don't Walk Their Talk: Towards a Framework for Understanding the Gap Between the Ethical Purchase Intentions and Actual Buying Behaviour of Ethically Minded Consumers. *Journal of Business Ethics*. 97(1), 139–158. https://doi.org/10.1007/s10551-010-0501-6

Carrington, M. J., Neville, B. A., Whitwell, G. J. (2014). Lost in translation: Exploring the ethical consumer intention–behavior gap. *Journal of Business Research*. 67(1), 2759–2767. https://doi.org/10.1016/j.jbusres.2012.09.022

Caruana, R., Carrington, M. J., Chatzidakis, A. (2016). "Beyond the Attitude-Behaviour Gap: Novel Perspectives in Consumer Ethics": Introduction to the Thematic Symposium. *Journal of Business Ethics*. 136(2), 215–218. https://doi.org/10.1007/s10551-014-2444-9

Chanda, R. C., Isa, S. M., Ahmed, T. (2024). Factors influencing customers' green purchasing intention: Evidence from developing country. *Journal of Science and Technology Policy Management*. 15(5), 1056–1084. https://doi.org/10.1108/JSTPM-03-2022-0044

Chaturvedi, P., Kulshreshtha, K., Tripathi, V. (2020). Investigating the determinants of behavioral intentions of generation Z for recycled clothing: An evidence from a developing economy. *Young Consumers*. 21(4), 403–417. https://doi.org/10.1108/YC-03-2020-1110

Cho, E., Gupta, S., Kim, Y. (2015). Style consumption: Its drivers and role in sustainable apparel consumption. *International Journal of Consumer Studies*. 39(6), 661–669. https://doi.org/10.1111/ijcs.12185

Dermody, J., Koenig-Lewis, N., Zhao, A. L., Hanmer-Lloyd, S. (2018). Appraising the influence of pro-environmental self-identity on sustainable consumption buying and curtailment in emerging markets: Evidence from China and Poland. *Journal of Business Research*. 86, 333–343. https://doi.org/10.1016/j.jbusres.2017.09.041

Deshmukh, S. M., Dhokpande, S. R., Sankhe, A., Khandekar, A. (2025). Effluent wastewater technologies for textile industry: A review. *Reviews in Inorganic Chemistry*. 45(1), 21–40. https://doi.org/10.1515/revic-2024-0046

Elgammal, I., Ghanem, M., Al-Modaf, O. (2024). Sustainable Purchasing Behaviors in Generation Z: The Role of Social Identity and Behavioral Intentions in the Saudi Context. *Sustainability*. 16(11), 4478. https://doi.org/10.3390/su16114478

Erickson, T. (2009, March 1). Generational Differences Between India and the U.S. *Harvard Business Review*. https://hbr.org/2009/02/global-generations-focus-on-in

Flanagan, A., Priyadarshini, A. (2021). A study of consumer behaviour towards food-waste in Ireland: Attitudes, quantities and global warming potentials. *Journal of Environmental Management*. 284, 112046. https://doi.org/10.1016/j.jenvman.2021.112046

Gazzola, P., Pavione, E., Pezzetti, R., Grechi, D. (2020). Trends in the Fashion Industry. The Perception of Sustainability and Circular Economy: A Gender/Generation Quantitative Approach. *Sustainability*. 12(7), 2809. https://doi.org/10.3390/su12072809

Gollwitzer, P. M. (1993). Goal Achievement: The Role of Intentions. *European Review of Social Psychology*. 4(1), 141–185. https://doi.org/10.1080/14792779343000059

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist.* 54(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Hair, J. F., Ringle, C. M., Sarstedt, M. (2019). *A primer on partial least squares structural equation modeling (PLS-SEM)*. 3rd ed. SAGE Publications.

Harris, F., Roby, H., Dibb, S. (2016). Sustainable clothing: Challenges, barriers and interventions for encouraging more sustainable consumer behaviour. *International Journal of Consumer Studies*. 40(3), 309–318. https://doi.org/10.1111/ijcs.12257

Hasan, M. M., Cai, L., Ji, X., Ocran, F. M. (2022). Eco-Friendly Clothing Market: A Study of Willingness to Purchase Organic Cotton Clothing in Bangladesh. *Sustainability*. 14(8), 4827. https://doi.org/10.3390/su14084827

Hasanbeigi, A. (2010). *Energy-Efficiency Improvement Opportunities for the Textile Industry*. https://escholarship.org/uc/item/6jw8s2gz

Henseler, J., Ringle, C. M., Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*. 43(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. SAGE Publications.

Islam, M. A., Van Staden, C. J. (2018). Social movement NGOs and the comprehensiveness of conflict mineral disclosures: Evidence from global companies. *Accounting, Organizations and Society*. 65, 1–19. https://doi.org/10.1016/j.aos.2017.11.002

Islam, M. M., Perry, P., Gill, S. (2021). Mapping environmentally sustainable practices in textiles, apparel and fashion industries: A systematic literature review. *Journal of Fashion Marketing and Management: An International Journal*. 25(2), 331–353. https://doi.org/10.1108/JFMM-07-2020-0130

Jaiswal, D., Kant, R. (2018). Green purchasing behaviour: A conceptual framework and empirical investigation of Indian consumers. *Journal of Retailing and Consumer Services*. 41, 60–69. https://doi.org/10.1016/j.jretconser.2017.11.008

Johnstone, M.-L., Tan, L. P. (2015). Exploring the Gap Between Consumers' Green Rhetoric and Purchasing Behaviour. *Journal of Business Ethics*. 132(2), 311–328. https://doi.org/10.1007/s10551-014-2316-3

Jung, H. J., Oh, K. W., Kim, H. M. (2021). Country Differences in Determinants of Behavioral Intention towards Sustainable Apparel Products. *Sustainability*. 13(2), 558. https://doi.org/10.3390/su13020558

Kang, J., Liu, C., Kim, S. (2013). Environmentally sustainable textile and apparel consumption: The role of consumer knowledge, perceived consumer effectiveness and perceived personal relevance. *International Journal of Consumer Studies*. 37(4), 442–452. https://doi.org/10.1111/ijcs.12013

Khare, A., Sadachar, A. (2017). Green apparel buying behaviour: A study on I ndian youth. *International Journal of Consumer Studies*. 41(5), 558–569. https://doi.org/10.1111/ijcs.12367

Lorente-Leyva, L. L., Alemany, M. M. E., Peluffo-Ordóñez, D. H. (2024). A conceptual framework for the operations planning of the textile supply chains: Insights for sustainable and smart planning in uncertain and dynamic contexts. *Computers & Industrial Engineering*. 187, 109824. https://doi.org/10.1016/j.cie.2023.109824

Lundblad, L., Davies, I. A. (2016). The values and motivations behind sustainable fashion consumption. *Journal of Consumer Behaviour*. 15(2), 149–162. https://doi.org/10.1002/cb.1559

Maichum, K., Parichatnon, S., Peng, K.-C. (2016). Application of the Extended Theory of Planned Behavior Model to Investigate Purchase Intention of Green Products among Thai Consumers. *Sustainability*. 8(10), 1077. https://doi.org/10.3390/su8101077

Mammadli, M. (2023). Factors Driving Sustainable Consumption in Azerbaijan: Comparison of Generation X, Generation Y and Generation Z. *Sustainability*. 15(20), 15159. https://doi.org/10.3390/su152015159

Mannheim, K. (1952). The sociological problem of generations. *Essays on the Sociology of Knowledge*. 306, 163-195.

Howe, N., Strauss, W. 1991. *Generations The History of America's Future, 1584 to 2069*. New York: William Morrow & Co.

Nguyen, H. V., Nguyen, N., Nguyen, B. K., Lobo, A., Vu, P. A. (2019). Organic Food Purchases in an Emerging Market: The Influence of Consumers' Personal Factors and Green Marketing Practices of Food Stores. *International Journal of Environmental Research and Public Health*. 16(6), 1037. https://doi.org/10.3390/ijerph16061037

Niinimäki, K. (2010). Eco-clothing, consumer identity and ideology. *Sustainable Development*. 18(3), 150–162. https://doi.org/10.1002/sd.455

Niinimäki, K., Peters, G., Dahlbo, H., Perry, P., Rissanen, T., Gwilt, A. (2020). The environmental price of fast fashion. *Nature Reviews Earth & Environment*. 1(4), 189–200. https://doi.org/10.1038/s43017-020-0039-9

Paul, J., Modi, A., Patel, J. (2016). Predicting green product consumption using theory of planned behavior and reasoned action. *Journal of Retailing and Consumer Services*. 29, 123–134. https://doi.org/10.1016/j.jretconser.2015.11.006

Schwarzer, R. (2008). Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. *Applied Psychology*. 57(1), 1–29. https://doi.org/10.1111/j.1464-0597.2007.00325.x

Seyfi, S., Sharifi-Tehrani, M., Hall, C. M., Vo-Thanh, T. (2025). Exploring the drivers of Gen Z tourists' buycott behaviour: A lifestyle politics perspective. *Journal of Sustainable Tourism*. 33(6), 1146–1164. https://doi.org/10.1080/09669582.2023.2166517

Sharma, C., Kushwah, S. (2025). Mapping the Theory of Consumption Values: A Systematic Review Using the TCCM Approach. *Journal of Consumer Behaviour*. 24(2), 562–610. https://doi.org/10.1002/cb.2434

Sheeran, P. (2002). Intention—Behavior Relations: A Conceptual and Empirical Review. *European Review of Social Psychology*. 12(1), 1–36. https://doi.org/10.1080/14792772143000003

Sheth, J. N., Newman, B. I., Gross, B. L. (1991). Why we buy what we buy: A theory of consumption values. *Journal of Business Research*. 22(2), 159–170. https://doi.org/10.1016/0148-2963(91)90050-8

Shirvanimoghaddam, K., Motamed, B., Ramakrishna, S., Naebe, M. (2020). Death by waste: Fashion and textile circular economy case. *Science of The Total Environment*. 718, 137317. https://doi.org/10.1016/j.scitotenv.2020.137317

Shirvanimoghaddam, K., Motamed, B., Ramakrishna, S., Naebe, M. (2020). Death by waste: Fashion and textile circular economy case. *Science of The Total Environment*. 718, 137317. https://doi.org/10.1016/j.scitotenv.2020.137317

Slepian, R. C., Vincent, A. C., Patterson, H., Furman, H. (2024). "Social media, wearables, telemedicine and digital health,"—A Gen Y and Z perspective. In: *Comprehensive Precision Medicine.* pp. 524–544. Elsevier. https://doi.org/10.1016/B978-0-12-824010-6.00072-1

Sobuj, Md., Khan, A. M., Habib, Md. A., Islam, Md. M. (2021). Factors influencing eco-friendly apparel purchase behavior of Bangladeshi young consumers: Case study. *Research Journal of Textile and Apparel*. 25(2), 139–157. https://doi.org/10.1108/RJTA-10-2019-0052

Song, S. (2020). *Exploring eco Fashion and Ethical Fashion Consumption: Comparing Consumers From Brazil, Portugal and China*. Master's thesis, Universidade de Coimbra, Portugal. https://www.proquest.com/openview/aae54ec5a04a8b10a1b1f099151172bd/1?cbl=2026366&diss=y&pq-origsite=gscholar

Srinivasan, V. (2012). Multi generations in the workforce: Building collaboration. *IIMB Management Review*. 24(1), 48–66. https://doi.org/10.1016/j.iimb.2012.01.004

Tandon, A., Sithipolvanichgul, J., Asmi, F., Anwar, M. A., Dhir, A. (2023). Drivers of green apparel consumption: Digging a little deeper into green apparel buying intentions. *Business Strategy and the Environment*. 32(6), 3997–4012. https://doi.org/10.1002/bse.3350

Tanrikulu, C. (2021). Theory of consumption values in consumer behaviour research: A review and future research agenda. *International Journal of Consumer Studies*. 45(6), 1176–1197. https://doi.org/10.1111/ijcs.12687

Tawde, S., Kamath, R., ShabbirHusain, R. V. (2023). 'Mind will not mind' – Decoding consumers' green intention-green purchase behavior gap via moderated mediation effects of implementation intentions and self-efficacy. *Journal of Cleaner Production*. 383, 135506. https://doi.org/10.1016/j.jclepro.2022.135506

Testa, F., Iovino, R., Iraldo, F. (2020). The circular economy and consumer behaviour: The mediating role of information seeking in buying circular packaging. *Business Strategy and the Environment*. 29(8), 3435–3448. https://doi.org/10.1002/bse.2587

Thangavel, P., Pathak, P., Chandra, B. (2021). Millennials and Generation Z: A generational cohort analysis of Indian consumers. *Benchmarking: An International Journal*. 28(7), 2157–2177. https://doi.org/10.1108/BIJ-01-2020-0050

Vermeir, I., Verbeke, W. (2006). Sustainable Food Consumption: Exploring the Consumer "Attitude – Behavioral Intention" Gap. *Journal of Agricultural and Environmental Ethics*. 19(2), 169–194. https://doi.org/10.1007/s10806-005-5485-3

Wiederhold, M., Martinez, L. F. (2018). Ethical consumer behaviour in Germany: The attitude-behaviour gap in the green apparel industry. *International Journal of Consumer Studies*. 42(4), 419–429. https://doi.org/10.1111/ijcs.12435

Witkowski, T. H., Reddy, S. (2010). Antecedents of Ethical Consumption Activities in Germany and the United States. *Australasian Marketing Journal*. 18(1), 8–14. https://doi.org/10.1016/j.ausmj.2009.10.011

Zerbini, C., Vergura, D. T., Latusi, S. (2019). A new model to predict consumers' willingness to buy fair-trade products. *Food Research International*. 122, 167–173. https://doi.org/10.1016/j.foodres.2019.04.008

Zhang, Y., Liu, C., Lyu, Y. (2023). Profiling Consumers: Examination of Chinese Gen Z Consumers' Sustainable Fashion Consumption. *Sustainability*. 15(11), 8447. https://doi.org/10.3390/su15118447

Zou, L. W., Chan, R. Y. K. (2019). Why and when do consumers perform green behaviors? An examination of regulatory focus and ethical ideology. *Journal of Business Research*. 94, 113–127. https://doi.org/10.1016/j.jbusres.2018.04.006

Contact information

Shovan Nasim Haider: e-mail: shaider@utb.cz

Miloslava Chovancova: e-mail: chovancova@utb.cz

ROBUSTNESS OF A FUZZY CONTROLLER VERSUS CLASSICAL PID CONTROL ON TYPICAL PLANTS

Marek Hebelka[1], Radim Farana[2]

[1]Department of Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

[2]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_009

ABSTRACT

This article deals with the robustness comparison of a Mamdani-type fuzzy controller and a classical PID controller on four typical controlled systems: proportional non-oscillatory, proportional oscillatory, proportional with transport delay, and integrating with inertia. The work responds to the lack of systematic comparison of both controller types under disturbances and parameter changes. Measurements were performed in MATLAB Simulink; PID was tuned automatically by the PID Tuner, the fuzzy controller manually by the Mamdani method with a simplified 3×3 rule base. Robustness was assessed for a position step of 10–40% and time-constant changes of 0–20%. Control quality was quantified by the IAE and ITAE summation criteria. For three of the four systems, the fuzzy controller was one to two orders of magnitude better in ITAE; for the system with transport delay, the PID was conversely two orders of magnitude better.

Keywords: fuzzy controller, PID, robustness, Mamdani, ITAE, transport delay
JEL Code: C88

1 INTRODUCTION

PID controllers have long been the standard in industrial automation. Their transfer properties are well known and tuning methods range from empirical (Ziegler-Nichols) to automated (PID Tuner in MATLAB Simulink). For linear systems with well-identifiable parameters, PID performs satisfactorily (Janošek and Farana, 2014). When the system contains nonlinearities, transport delay, or its parameters drift in time, the PID performance often drops. The work draws on agricultural applications, where systems with longer time constants occur (greenhouse temperature control, grain-silo humidity control) and the choice of a suitable controller is essential.

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1.1 Fuzzy Logic and the Mamdani FIS

Fuzzy logic, introduced by Zadeh (1965) and shaped into a controller by Mamdani and Assilian (1975), offers an alternative. Instead of binary values {0, 1} it works with the entire interval ⟨0,1⟩, and the central concept is the membership function, which assigns a single measurement degrees of membership in multiple fuzzy sets. A Mamdani-type Fuzzy Inference System (FIS) performs inference according to rules of the form:

(1)

$$\text{IF } \; e \text{ is } A_i \;\; \text{AND } \; \dot{e} \text{ is } B_j \;\; \text{THEN } \; u \text{ is } C_{ij},$$

where e is the control error, ė its derivative and u the actuation. The Mamdani FIS has a fixed input-output model: fuzzification → inference → defuzzification, with the centroid as a typical defuzzification method. An alternative to Mamdani is the Takagi-Sugeno-Kang (TSK) system (Takagi and Sugeno, 1993), which uses input functions instead of fuzzy sets in rule consequents; Mamdani was chosen in this work for its more intuitive design.

1.2 Robustness and Control-Quality Criteria

Robustness is the ability of a controller to maintain control quality under disturbances or system-parameter changes. The quality is typically quantified by summation criteria, particularly the integral of the absolute error and the time-weighted integral:

(2)

$$I_{\text{IAE}} = \int_0^\infty |e(t)|\, dt, \qquad I_{\text{ITAE}} = \int_0^\infty t \cdot |e(t)|\, dt,$$

ITAE penalises long-lasting errors and is often more relevant for practical evaluation. Studies comparing fuzzy and PID control usually focus on a single specific task (Modrlák, 2002) and benchmark against the best-fitting setup; a systematic comparison across multiple system types appears less frequently. Providing such a comparison is the aim of this work.

2 METHODOLOGY

Four representative controlled systems were chosen. A proportional non-oscillatory second-order system with transfer:

(3)

$$G_{s1}\left(s\right) = \frac{1}{(T_1 s+1)(T_2 s+1)}, \quad T_1 = 10, \left| T_2 = 5.\right.$$

A proportional oscillatory second-order system:

(4)

$$G_{s2}\left(s\right) = \frac{1}{T_0^2 s^2+2\xi T_0 s+1}, \quad T_0 = 5, \left| \, \xi = 0.8.\right.$$

A proportional system with transport delay:

$$G_{s3}(s) = \frac{e^{-T_d s}}{T_1 s+1}, \quad T_1 = 25, \left| T_d = 1. \right. \tag{5}$$

An integrating system with inertia:

$$G_{s4}(s) = \frac{1}{T_s(T_1 s+1)}, \quad T_s = 10, \left| T_1 = 10. \right. \tag{6}$$

Simulations ran in MATLAB Simulink. For each system a circuit was assembled in which PID and fuzzy control operated in parallel over the same plant so that the responses could be compared directly. The PID was tuned automatically by the PID Tuner with parameters *Response Time* 1.25 s and *Transient Behavior* 0.45; for the system with transport delay the *Response Time* had to be increased to 4.5 s, as lower values led to unstable response. The fuzzy controller was tuned manually by the Mamdani method with triangular membership functions for the error and its derivative (Figure 1) and a simplified 3×3 rule base (Figure 2). The tuning target was a maximum steady-state deviation of 0.003 from the setpoint as the reference criterion. Defuzzification used the centroid method. Robustness was tested in two independent scenarios: a position step of 10, 20, 30 and 40% of the input signal, and a change of time constants by 5, 10, 15 and 20% of the original value (modelling system ageing or an inaccurate model). For each scenario the overshoot, settling time and IAE and ITAE criteria were recorded.

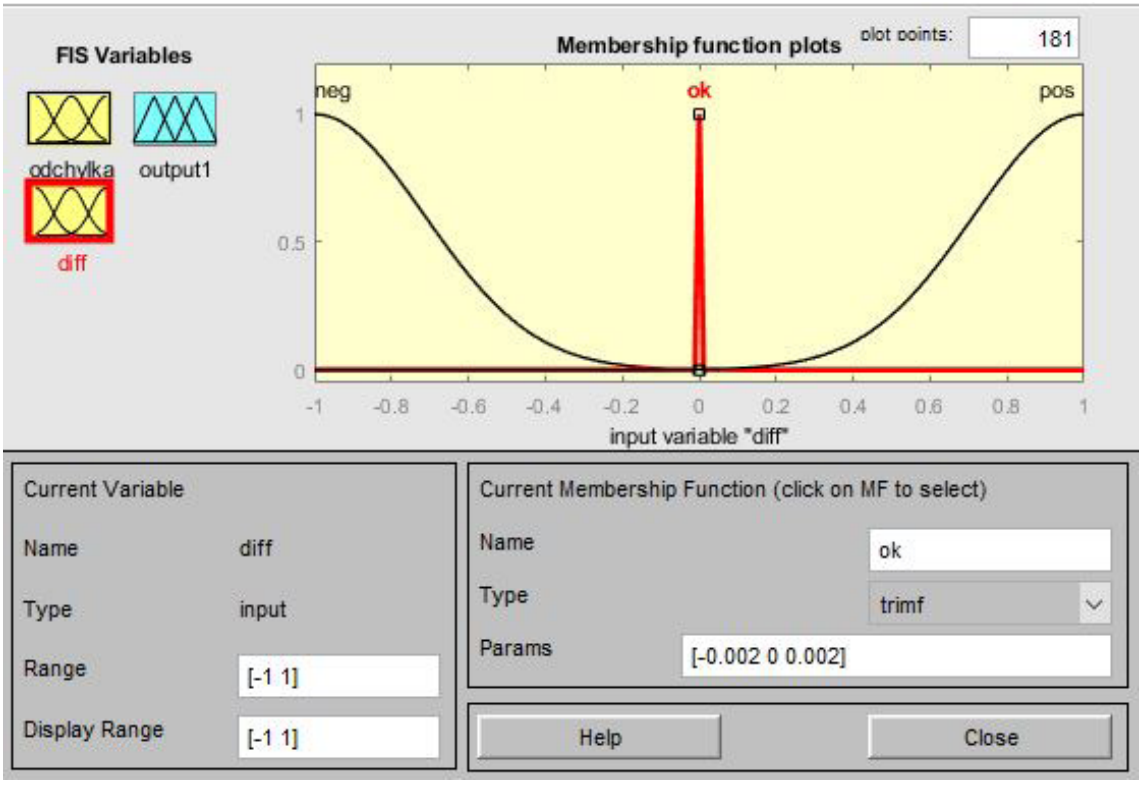


Fig. 1: Membership functions of the input "error" for the proportional non-oscillatory plant — three triangular sets *neg*, *ok*, *pos* over the range ⟨-5,5⟩; the narrow peak ok around zero ensures stabilisation in the steady state (adopted from Hebelka 2026)

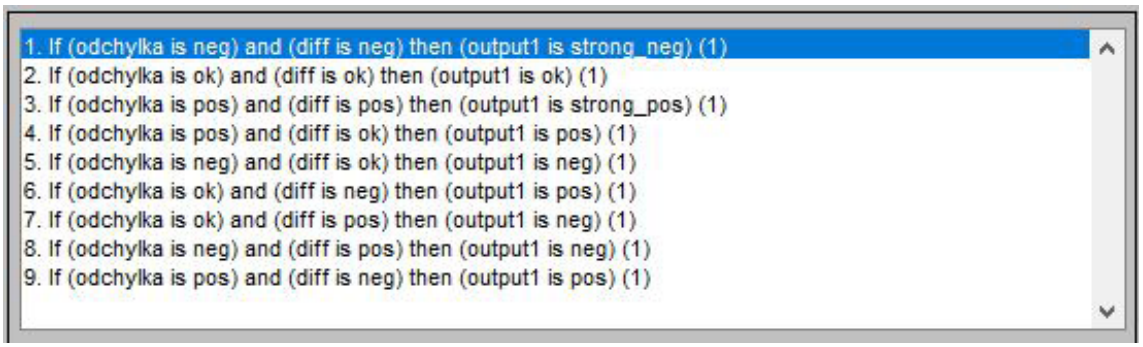


Fig. 2: Mamdani fuzzy controller rule base — nine IF (error × derivative) THEN action rules for the proportional non-oscillatory plant (adopted from Hebelka 2026).

3 RESULTS

The resulting PID parameters after tuning by the PID Tuner are summarised in Table 1.

Tab. 1 PID controller tuning by the PID Tuner

System	K_p	K_I	K_D
Proportional non-oscillatory	11.36	2.00	13.75
Proportional oscillatory	19.33	2.17	37.08
Proportional with transport delay	11.46	0.95	−3.28
Integrating with inertia	6.83	0.66	15.41

The negative derivative gain for the system with transport delay reflects that the PID Tuner chose a markedly less aggressive action to compensate the dead time.

Proportional non–oscillatory. PID overshoot grew roughly linearly with the error (0.080 → 0.317 for 10 → 40%), the fuzzy controller saturated around 0.16. Settling times were 4.6–15.7 s for PID and 0.4–0.82 s for fuzzy. ITAE for fuzzy was 1.30–2.15 versus 14.52–42.16 for PID. The fuzzy controller is more robust here.

Proportional oscillatory. PID overshoot 0.098–0.388 (linear with error), fuzzy 0.116–0.157 (asymptotically stable). Settling times PID 17.9–27.4 s, fuzzy 0.75–1.26 s. ITAE fuzzy 3.26–5.77 versus 37.65–98.55 for PID. The fuzzy controller is more robust.

Proportional with transport delay. The third test reversed the result. PID settling time 34.7–48.5 s, fuzzy 83.2–218.3 s — that is, 2.5–4.5× worse. IAE PID 19.1–24.3, fuzzy 245.7–385.5. ITAE PID 221–570, fuzzy 15 092–57 856, two orders of magnitude worse. PID is more robust here.

Integrating with inertia. PID overshoot 0.109–0.432, fuzzy 0.091–0.157. Settling times PID 10.2–12.2 s, fuzzy 0.44–0.85 s. ITAE PID 47.6–113.6, fuzzy 1.21–2.00 — two orders of magnitude better for fuzzy (Hebelka, 2026).

A comparison summary for all four systems is shown in Table 2; a typical time response is shown in Figure 3.

Tab. 2 Robustness summary — which controller wins per system

System	More robust controller	ITAE (PID / fuzzy)
Proportional non-oscillatory	Fuzzy	14–42 / 1.3–2.2
Proportional oscillatory	Fuzzy	38–99 / 3.3–5.8
Proportional with transport delay	**PID**	221–570 / 15 092–57 856
Integrating with inertia	Fuzzy	48–114 / 1.2–2.0

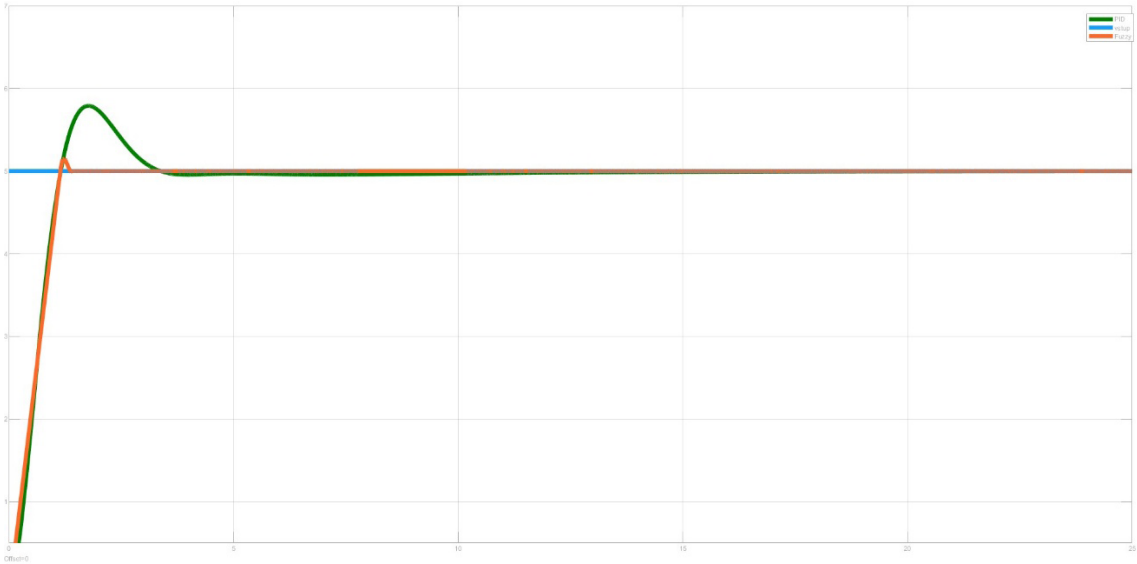


Fig. 3: Time response of PID (green) and fuzzy (orange) controllers for the proportional non-oscillatory system under a unit step; blue is the setpoint (adopted from Hebelka 2026).

4 DISCUSSION

The results confirm the general statement in the literature that fuzzy controllers usually outperform PID on nonlinear or poorly modelled tasks. The system with transport delay is an important exception. Dead time is tricky from a controller's standpoint: an action taken at time t shows up on the controlled variable only at t+T_d. If the controller reacts only to the instantaneous error and its derivative, it easily overshoots — keeping the action longer than necessary because it cannot yet see the effect. The PID copes with this via a negative K_D, which effectively brakes the action before the error appears. The Mamdani fuzzy controller with a simplified 3×3 rule base cannot achieve such fine behaviour; its rule base reacts to coarse categories, and damping the dead time would require substantially richer rules. A suitable path forward is a Smith predictor combined with fuzzy logic or a neuro-fuzzy controller.

A second point concerns the tuning method. PID was tuned automatically by the PID Tuner, which yields an essentially optimal result; the fuzzy controller was tuned by manual trial and error. This is fair in the sense that a manually tuned PID typically performs worse than a Tuner-tuned one — but it does not allow a fully symmetric comparison. Future work could employ optimisation methods for automated fuzzy tuning (for example a genetic algorithm to search the membership-function layout).

5 CONCLUSION

The work compared a Mamdani-type fuzzy controller and a classical PID controller on four typical controlled systems. For each system both controllers were tuned by the same methodology and tested under a 10–40% position step and a 0–20% change of time constants. Three findings deserve attention. For three of the four systems (proportional non-oscillatory, oscillatory, integrating with inertia) the fuzzy controller was more robust — ITAE typically 10–20× better than PID, settling time one order of magnitude shorter. For the system with transport delay PID was more robust — ITAE two orders of magnitude better (220–570 versus 15 000 – 58 000); a fuzzy controller built on the instantaneous error and its derivative cannot handle dead time without a substantially richer rule base. The choice of the quality metric materially affects interpretation — ITAE, which penalises late errors, better matches user perception than settling time or IAE alone. Future expansions may include automated fuzzy tuning by a genetic algorithm and a combination with predictive techniques for systems with transport delay.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

REFERENCES

HEBELKA, M. 2026. *Robustness assessment of a fuzzy controller versus classical PID control on typical controlled systems* [in Czech]. Master's thesis, head: Farana, R. Brno: Mendel University in Brno.

JANOŠEK, M., FARANA, R. 2014. *Fuzzy modelling and control* [in Czech]. Ostrava: University of Ostrava, 2014. 106 p. Electronic techning text.

MAMDANI, E. H., ASSILIAN, S. 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*. 7(1), 1-13. ISSN 0020-7373. https://doi.org/10.1016/S0020-7373(75)80002-2

MATHWORKS. 2020. PID Tuner App. *Mathworks*. https://www.mathworks.com/help/control/ref/pidtuner-app.html

MATHWORKS. 2023. Fuzzy Logic Toolbox. *Mathworks*. https://www.mathworks.com/help/fuzzy/

MODRLÁK, O. 2002. *Fuzzy control and regulation* [in Czech]. Study materials. Liberec: Technical University of Liberec.

TAKAGI, T., SUGENO, M. 1993. Fuzzy Identification of Systems and Its Applications to Modeling and Control. In: *Readings in Fuzzy Sets for Intelligent Systems*, pp. 387-403. https://doi.org/10.1016/b978-1-4832-1450-4.50045-6

ZADEH, L. A. 1965. Fuzzy sets. *Information and Control*. 8(3), 338-353. ISSN 0019-9958. https://doi.org/10.1016/S0019-9958(65)90241-X

Contact information

Marek Hebelka: e-mail: Marek1Hebelka@seznam.cz

Radim Farana: e-mail: radim.farana@mendelu.cz

PREDICTIVE ANALYTICS IN NEPALESE AGRICULTURE

Sandesh Kumar Kafle[1], František Dařena[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/10.11118/ecos2026_010

ABSTRACT

This study addresses climate vulnerability and the lack of data-driven tools in Nepalese agriculture by building predictive crop-yield models for 14 districts in the Koshi Province. Applying a data-science pipeline to historical agricultural records and NASA POWER agroclimatic data spanning 11 crops over 11 fiscal years (2011/12–2021/22), the research evaluates Random Forest and Gradient Boosting Regressors across five distinct weather-feature datasets. The results demonstrate that the Random Forest Regressor consistently outperforms Gradient Boosting, explaining 74% to 84% of yield variance. The highest predictive accuracy was achieved using engineered growing-season averages (Baseline $R^2 = 0.83$, Post-tuning $R^2 = 0.84$), confirming that crop-specific seasonal summaries are the most informative features. Temperature emerged as the strongest single predictor of crop yield ($R^2 = 0.82$), while standalone precipitation exhibited the weakest predictive power ($R^2 = 0.74$). These findings offer an empirical decision-support foundation for farmers and policymakers to quantify climate impacts and enhance sustainable agricultural planning in Nepal.

Keywords: predictive analytics, machine learning, crop yield prediction, Nepalese agriculture, weather impact
JEL Code: Q10, C54, C63

1 INTRODUCTION

Agriculture is the backbone of Nepal's economy, employing around 65% of the population and contributing approximately 27% of GDP. The country's diverse agroecological zones and rich biodiversity create significant potential for food and nutritional security; however, the sector faces persistent challenges, including unpredictable weather events, outdated irrigation infrastructure, and a lack of evidence-based decision-making tools.

Predictive analytics—using historical data to forecast future outcomes—offers a path toward more resilient and productive agriculture. Advances in machine learning (ML), the Internet of Things (IoT), and open data platforms now make it feasible to develop yield-prediction models

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

tailored to Nepal's conditions. Technology can thus assist in transforming the global food production system while mitigating its impact on the climate and environment (Lim, 2022).

Nepalese farmers lack the knowledge and tools to leverage past weather and yield data for decision-making. Existing crop-yield prediction models do not adequately capture the specific impact of weather variability on Nepalese agriculture. This study addresses the gap by building and evaluating predictive models using historical agroclimatic and agricultural data.

The objective is to identify problems in Nepal's agricultural sector that can be addressed through advanced analysis of historical data and to provide a decision-support foundation for farmers and policymakers. The study specifically pursues the following questions:

- *Can a predictive model capture the effect of weather factors (temperature, humidity, precipitation) on crop yields?*
- *Which weather parameter exerts the greatest influence on agricultural yields?*
- *Is there a significant relationship between growing-season weather conditions and crop yield?*
- *What actionable recommendations can improve the future potential of Nepalese agriculture?*

The study focuses on 14 districts of Koshi Province, Nepal, covering 11 crops (6 cereals, 3 cash crops, 2 pulses) over 11 fiscal years (2011/12–2021/22). Key limitations include sampling bias toward Koshi Province, data quality constraints, reliance on publicly available weather records, exclusion of soil and socioeconomic variables, and the difficulty of modelling unpredictable extreme weather events.

2 CURRENT STATE OF RESEARCH

2.1 Predictive Analytics in Agriculture

Predictive analytics has gained widespread adoption in agriculture globally, enabling more accurate yield forecasting, resource optimization, and risk management. Machine learning models—particularly ensemble methods such as Random Forest and Gradient Boosting—have demonstrated strong performance in non-linear regression tasks involving weather-crop relationships. Studies across East Africa (Aworka *et al.*, 2022), Europe (Paudel *et al.*, 2022), India (Sabu *et al.*, 2020), and Nepal (Fernández-Beltrán *et al.*, 2021) confirm the adaptability of these approaches across diverse agricultural contexts.

2.2 Weather and Crop Yield in Nepal

Nepal's agriculture is highly monsoon-dependent. Rice, the dominant cereal, requires timely and adequate rainfall; wheat is sensitive to heat and water stress during grain-filling. Climate studies confirm that increased temperature variability and erratic rainfall reduce yields for both crops, with effects varying by altitude and growing season. A 2016 CBS survey found that nearly 100% of households observed increased drought over the prior 25 years, and 97.7% noted a rise in disease incidence and irregular rainfall.

Climate change is projected to intensify these pressures through rising temperatures, altered monsoon patterns, and more frequent extreme events—flooding in 2017 alone damaged 10% of total paddy area (FAO, 2023).

2.3 Key Models and Evaluation Approaches

The literature documents several modelling frameworks relevant to Nepal: CNN-based models using Sentinel-2 satellite imagery for rice yield (Fernández-Beltrán *et al.*, 2021), ARIMA and LSTM for price/yield time-series forecasting, Random Forest and Support Vector Regression for multi--feature yield prediction (Ramu, K. & Priyadarsini, K., 2021); and interaction regression models that can outperform standard ML algorithms (Ansarifar *et al.*, 2021). Common evaluation metrics include R^2 (explanatory power), Mean Squared Error (MSE), and Mean Absolute Error (MAE), validated primarily via k-fold cross-validation (Oikonomidis *et al.*, 2022; Burdett & Wellen, 2022).

2.4 Research Gap

Despite a growing global body of work, predictive models specifically calibrated to Nepal's diverse agroecological conditions, fiscal-year data structure, and smallholder context remain scarce. This study addresses that gap by building district-level, multi-crop models from publicly available Nepalese data.

3 METHODOLOGY AND DATA

3.1 Research Approach and Design

The study combines quantitative modelling with qualitative contextualization. Quantitative analysis is central: numerical weather and yield data are used to train and evaluate regression-based ML models predicting yield. Qualitative insights from the literature inform feature selection, model interpretation, and recommendations. The workflow follows a standard data-science pipeline: data collection → preprocessing → feature engineering → model training → evaluation → tuning.

3.2 Data Sources

Two primary data sources were used: (1) annual crop yield records for 11 crops across 14 Koshi districts, obtained from the Ministry of Agriculture and Livestock Development (MOALD); and (2) monthly weather data—surface temperature (TS), 2-meter temperature (T2M), specific humidity (QV2M), relative humidity (RH2M), and corrected precipitation (PRECTOTCORR)—obtained from NASA POWER. The combined dataset spans 11 years, 14 districts, and 11 crops, and contains the following features: Year (Fiscal year), District (Name of the district), Crop (Name of the crop), Yield (Metric ton production over the area in Hectare), TS (Earth Skin Temperature/Surface Temperature (°C)), T2M (Temperature at 2 Meters (°C)), QV2M (Specific Humidity at 2 Meters (g/kg)), RH2M (Relative Humidity at 2 Meters (%)), PRECTOTCORR (Precipitation Corrected (mm/day)).

3.3 Data Preprocessing

After merging the two sources by fiscal year and district, the dataset was cleaned by removing duplicate records, replacing missing numerical values with column means, and dropping remaining null entries. All features were then normalised using the Yeo-Johnson power transformation (to approximate Gaussian distributions) and standardised with StandardScaler (zero mean, unit variance).

3.4 Feature Engineering

Raw monthly weather data produced a large feature space detrimental to model performance. Following Eck *et al.* (2020), crop-specific growing-season averages were calculated for each weather variable, reducing the feature set to five engineered variables: GSavgTS (Growing season average for Temperature Surface), GSavgT2M (Growing season average for Temperature at 2 meters), GSavgQV2M (Growing season average for Specific Humidity at 2 meters), GSavgRH2M (Growing season average for Relative Humidity at 2 meters), and GSavgPREC (Growing season average for Precipitation Corrected). Five distinct datasets were constructed to test the individual and combined predictive power of temperature, humidity, precipitation, and growing-season weather: Dataset 1 (Temperature + Humidity + Precipitation + Crop + Yield), Dataset 2 (Humidity + Crop + Yield), Dataset 3 (Temperature + Crop + Yield), Dataset 4 (Precipitation + Crop + Yield), Dataset 5 (Growing Season Averages + Crop + Yield).

3.5 Models

Two ensemble regression models were selected after initial experiments: Random Forest Regressor and Gradient Boosting Regressor (both from scikit-learn). Each model was trained on 75% of the data and tested on the remaining 25%. Key hyperparameters include n_estimators, max_depth, learning_rate (GBM only), min_samples_split, min_samples_leaf, and max_features.

3.6 Evaluation and Tuning

Model performance was measured using R^2, MSE, and MAE on the held-out test set. Generalisation was assessed with 5-fold cross-validation (KFold). Hyperparameters were optimised via two approaches: RandomizedSearchCV (faster, samples random combinations) and GridSearchCV (exhaustive but computationally expensive). Both tuning results are reported and compared against baseline models.

4 RESULTS AND DISCUSSION

4.1 Baseline Model Performance

Table 1 summarises the performance of the two base models across all five datasets. Random Forest consistently outperformed Gradient Boosting in R^2, MSE, and MAE. The highest baseline performance was achieved on Dataset 5 (Growing Season Averages) with Random Forest ($R^2 = 0.83$,

Tab. 1 Base Model Performance

Dataset	RF R^2	RF MSE	RF MAE	GB R^2	GB MSE	GB MAE
1: All weather	0.80	29.09	1.62	0.77	32.66	2.14
2: Humidity	0.79	29.83	1.70	0.79	30.93	2.11
3: Temperature	0.82	25.62	1.60	0.78	31.09	2.10
4: Precipitation	0.74	37.66	2.02	0.74	37.31	2.21
5: Growing Season Avg	0.83	25.03	1.76	0.81	27.80	2.03

MSE = 25.03) and Dataset 3 (Temperature only) also performing strongly (R^2 = 0.82, MSE = 25.62). Precipitation alone (Dataset 4) yielded the weakest results for both models (R^2 = 0.74).

4.2 Cross-Validation Results

5-fold cross-validation confirmed the robustness of both models. Average R^2 values ranged from 0.78–0.80 for both models across all datasets, with standard deviations of MSE between 8.99 and 11.09—indicating acceptable and consistent variance. The two models performed nearly identically in cross-validation, suggesting neither is prone to overfitting on this dataset.

4.3 Model Tuning Results

After hyperparameter optimization, performance improved modestly. Table 2 shows the post-tuning results. The best overall result was achieved by Random Forest on Dataset 5 using both RandomizedSearch and GridSearch (R^2 = 0.84, MSE = 23.24 and 23.01, respectively), a meaningful improvement over the baseline MSE of 25.03. RandomizedSearch generally yielded comparable or better results than GridSearch at far lower computational cost.

Tab. 2 Performance After Hyperparameter Tuning

Dataset	Model	Rand. R^2	Rand. MSE	Grid R^2	Grid MSE
1: All weather	Random Forest	0.81	28.03	0.80	28.15
1: All weather	Gradient Boosting	0.78	31.50	0.77	32.66
3: Temperature	Random Forest	0.83	25.06	0.83	24.60
5: Growing Season	Random Forest	0.84	23.24	0.84	23.01
5: Growing Season	Gradient Boosting	0.80	28.63	0.81	28.00

4.4 Key Findings

Several key findings can be identified after detailed analysis

- Temperature is the strongest single predictor of crop yield. Dataset 3 (temperature only) achieved a near-top R^2 = 0.82, outperforming humidity and precipitation individually.
- Growing-season averages are the most informative engineered feature. Dataset 5 achieved the best performance across both models and both tuning methods, confirming that crop-specific seasonal weather summaries are more predictive than raw monthly records.
- Precipitation has the weakest standalone predictive power. Dataset 4 consistently recorded the lowest R^2 (~0.74) and highest MSE, though it still contributes in combination with other variables.
- Random Forest Regressor is the preferred model. It outperformed Gradient Boosting on all five datasets in both baseline and tuned evaluations, with higher R^2 and lower MSE/MAE in most cases.
- Broad weather feature sets improve accuracy. Combining temperature, humidity, and precipitation (Dataset 1) produced better results than any single variable except growing-season averages, suggesting complementary predictive value.

4.5 Discussion and Recommendations

The findings confirm that weather—particularly temperature and growing-season conditions—is a significant driver of agricultural yield variability in Nepal's Koshi Province. The Random Forest model's robustness and its ability to capture non-linear feature interactions make it well-suited for this domain.

Recommendations for practitioners and policymakers include:

- Incorporate comprehensive weather variables in yield models, prioritising temperature and growing-season averages.
- Use RandomizedSearchCV for hyperparameter tuning as a cost-effective alternative to GridSearchCV.
- Extend the model with additional features (soil properties, fertiliser data, planting and harvest season averages) to improve precision.
- Establish feedback loops with farmers and policymakers to validate predictions and iteratively refine the model.
- Design seasonal climate-resilience policies that specifically account for temperature fluctuations during crop growing seasons.

5 CONCLUSIONS

This study demonstrates that machine learning—in particular, the Random Forest Regressor—can effectively model the relationship between weather conditions and crop yields in Nepal's Koshi Province, achieving R^2 values up to 0.84 after hyperparameter tuning. Across five datasets representing different combinations of temperature, humidity, and precipitation, the models consistently explained 74–84% of yield variance, confirming the significant impact of weather on agricultural productivity.

Among all weather variables examined, temperature and growing-season average weather emerge as the most influential predictors of crop yield. Precipitation, while important in combination, contributes less when used in isolation. These results align with the broader climate-agriculture literature and provide empirical grounding specific to Nepal's Koshi region.

The developed framework offers a practical decision-support foundation: it can predict yields based on historical data, quantify how specific weather patterns drive productivity, and guide resource allocation and risk management. With continuous data updates, integration of additional features, and collaboration between data scientists, agronomists, and policymakers, this approach can evolve into a robust tool for sustainable agricultural planning in Nepal.

Future research should broaden the geographic scope beyond Koshi Province, incorporate more crops and non-weather variables (e.g., soil chemistry, fertilizer application), and explore real-time data integration via IoT sensors and satellite imagery to further enhance predictive accuracy and practical utility.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ANSARIFAR, J., WANG, L., ARCHONTOULIS, S. V. 2021. An interaction regression model for crop yield prediction. *Scientific Reports*. 11(1).

AWORKA, R. *et al*. 2022. Agricultural decision system based on Advanced Machine Learning models for yield prediction: Case of East African countries. *Smart Agricultural Technology*. 2, 100048.

BURDETT, H., WELLEN, C. 2022. Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. *Precision Agriculture*. 23(5), 1553–1574.

ECK, M. A. *et al*. 2020. Influence of growing season temperature and precipitation anomalies on crop yield in the southeastern United States. *Agricultural and Forest Meteorology*. 291, 108053.

FAO. 2023. *Country Brief – Nepal*. Food and Agriculture Organization of the United Nations. https://www.fao.org/giews/countrybrief/country.jsp?code=NPL.

FERNÁNDEZ-BELTRÁN, R., BAIDAR, T., KANG, J., PLA, F. 2021. Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3D CNN: A Case Study in Nepal. *Remote Sensing*. 13(7), 1391.

KUMAR KAFLE, S. 2024. *Predictive Analytics in Nepalese Agriculture*. Diploma Thesis. Brno: Mendel University in Brno.

LIM, X. 2022. Agriculture Needs Technology for Resilient Food Production. *World Economic Forum*. [Accessed: August 2023] https://www.weforum.org/agenda/2022/03/how-technology-can-help-address-challenges-in-agriculture/

MINISTRY OF AGRICULTURE AND LIVESTOCK DEVELOPMENT (MOALD). [n.d.]. Welcome to Ministry of Agriculture and Livestock Development. *MOALD*. [Accessed: August 2023] https://moald.gov.np/welcome-to-ministry-of-agriculture-and-livestock-development/

NASA. 2024. *NASA POWER. Data Access Viewer.* [Accessed: March 2024] https://power.larc.nasa.gov/data-access-viewer/

OIKONOMIDIS, A., CATAL, C., KASSAHUN, A. 2022. Hybrid deep learning-based models for crop yield prediction. *Applied Artificial Intelligence*. 36(1).

PAUDEL, D. *et al*. 2022. Machine learning for regional crop yield forecasting in Europe. *Field Crops Research*. 276, 108377.

RAMU, K., PRIYADARSINI, K. 2021. A Review on Crop Yield Prediction Using Machine Learning Methods. In: *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*. pp. 1239-1245.

SABU, K. M., KUMAR, T. K. M. 2020. Predictive analytics in agriculture: Forecasting prices of arecanuts in Kerala. *Procedia Computer Science*. 171, 699–708.

Contact information

Sandesh Kumar Kafle: e-mail: 89995@node.mendelu.cz

František Dařena: e-mail: darena@mendelu.cz

VISUAL ATTENTION TO INFLUENCER MARKETING CONTENT IN DECORATIVE COSMETICS: EVIDENCE FROM AN EYE-TRACKING EXPERIMENT

Kateřina Klátilová[1], Stanislav Mokrý[1], Jana Turčínková[1], Michal Pšurný[1]

[1]Department of Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/10.11118/ecos2026_011

ABSTRACT

This study examines how selected elements of influencer marketing content affect consumer attention and perception in the context of decorative cosmetics. Primary data were collected through an eye-tracking experiment and follow-up in-depth interviews with 35 women aged 18 to 34. The experiment used visual and audiovisual stimuli based on social media content and included several A/B comparisons, such as user comments, number of hashtags, product arrangement, product photography style, make-up application, and the use of a well-known person's face. The results suggest that product arrangement and photographic style influenced visual attention, while celebrity presence and make-up application did not lead to statistically significant differences in observation time. Interviews showed that participants generally perceived influencer reviews positively, especially when they appeared authentic, critical, and based on real product use. Most participants also preferred influencers to disclose paid cooperation.

Keywords: influencer marketing; decorative cosmetics; eye-tracking; social media; Instagram; consumer attention; sponsored content
JEL Code: C91, D91, M31

1 INTRODUCTION

In January 2019, Relatable (2019) conducted a survey of 351 marketers from around the world and concluded that 94% of respondents considered influencer marketing to be an effective form of promotion. The survey also indicated that Instagram was the most popular platform for influencer campaigns. According to statistics available at that time, approximately 1 billion users were active



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

on Instagram. Interestingly, 68% of Instagram users were women. The platform was particularly popular among teenagers and people aged 18 to 29 (Aslam, 2019). For comparison, according to surveys from the last quarter of 2018, there were nearly 4.2 billion active Internet users worldwide, while 3.4 billion people were active on social networks (Kemp, 2018).

According to Hennessy (2018), Instagram is the platform to which the vast majority of marketing campaigns involving influencers are directed. Aslam (2019) reports that up to 78% of influencers prefer Instagram as a branding platform, and more than half of influencers also use the Stories feature for sponsored campaigns.

The Influencer Marketing Hub study (Conick, 2018) shows that the value of influencer marketing as an industry was USD 1.7 billion in 2016, USD 3 billion in 2017, USD 4.6 billion in 2018, and an estimated USD 6.5 billion in 2019. In a study conducted by CPC Strategy (2018) on a sample of 1,571 Americans aged 13 and older, consumers' perceptions of influencer recommendations were examined. When participants knew that influencers had been paid for cooperation, the largest group of respondents (44%) stated that it was difficult to answer without further information. They usually indicated that it depended on how well they knew and trusted the influencer. At the same time, the study showed that the most common reason why consumers considered influencers' recommendations credible was that the influencers used the products themselves. In contrast, participants reported that the number of followers played a less important role in influencer credibility.

From a global perspective, the cosmetics market has been growing steadily, with 2018 showing the largest increase, namely a 5.5% increase in sales compared with the previous year (Statista, 2019). A study by Allied Market Research (Rajput, 2016) expected this growth trend to continue due to factors such as the growth of global economies, lifestyle changes, increasing demand for sun protection cosmetics related to changing climatic conditions, and a growing preference for natural and organic cosmetics.

1.1 Influencer marketing

Newberry (2019) defines influencer marketing as a form of collaboration between a brand and an individual whose role is to promote a particular product, service, or campaign. In general, anyone with influence can be considered an influencer (Hennessy, 2018). According to Janouch (2014, p. 184), influencer marketing is based on identifying leaders of key communities on social networks. These leaders are typically able to present products and influence the opinions of others.

According to Conick (2018), influencer marketing differs from other forms of online advertising because it is less intrusive, more flexible, and seeks to build consumer trust. It offers an opportunity to reach a new generation of consumers who are accustomed to skipping, ignoring, or actively avoiding online advertising. In their study, Pitz, Köhler and Esch (2018) found that influencer campaigns affect brand perception, with a stronger effect observed for lesser-known brands than for strong brands.

Stokes (2017) examined the success of influencer marketing from a psychological perspective. In this respect, he refers to the mere exposure effect. In simplified terms, Zajonc's (2001) theory suggests that repeated exposure to certain content can increase preference for that content over time. Another relevant phenomenon is conformity, which refers to the adaptation of an individual's behaviour to surrounding groups, sometimes at the expense of suppressing one's own opinion. This tendency has been rooted in human social behaviour since ancient times, when group belonging was important for survival. Today, similar tendencies can be observed in the way users form communities on social networks, follow influencers, and allow their opinions to be influenced, either consciously or subconsciously.

The success of influencers may also be supported by the so-called halo effect, which refers to the formation of an overall opinion about a person based on one dominant impression that overshadows other characteristics. Social media users may therefore be inclined to believe, based on what they see online, that an influencer lives an admirable life and represents an attractive personality.

If a company is looking for a suitable influencer in its field, it is useful to observe several typical characteristics that influencers should possess (Řezníček and Procházka, 2014). Influencers regularly comment on issues and news in their field, have a broad community of people connected to them in some way, are considered influential by others in the field, and are able to influence the opinions of others.

A beauty influencer can be understood as a person known on social networks for expertise in skin care, hair care, or make-up. Such influencers gather audiences that trust their beauty advice and product recommendations. Collaboration between these influencers and brands usually aims to create sponsored content featuring the brand's product and to generate impact on social networks (Mediakix, 2018).

The aim of this paper is to present the results of research focused on the use of influencer marketing in the promotion of cosmetic products. Special emphasis is placed on identifying factors that influence the perception of selected forms of product presentation on social networks. Based on the findings, recommendations are formulated.

2 DATA AND METHODS

A combination of several techniques was used to collect primary data, namely eye-tracking research and in-depth interviews. The eye-tracking research was conducted in March 2019 in the ETLab eye-tracking laboratory at the Faculty of Business and Economics, Mendel University in Brno. SMI Experiment Center and SMI iView X software were used for stimulus presentation and data recording. A remote SMI RED 250 eye-tracker was used for the research, with the recording frequency set to 120 Hz. Individual stimuli were presented to participants on a 22-inch monitor with a 16:10 aspect ratio. Each experiment started with a five-point calibration followed by four--point validation. For calibration purposes, a calibration deviation of up to 0.6 degrees on the X and Y axes was accepted.

A total of 13 stimuli were investigated, of which four were audiovisual and nine were visual. This paper presents the results of six visual stimuli and two audiovisual stimuli. The research was designed in an A/B testing mode, meaning that some stimuli were modified. Participants were divided into two groups: one group was exposed to the original version of a stimulus, while the other group was exposed to its modified version. Before each stimulus, a fixation cross was displayed for 1,000 ms to unify the initial gaze position of all participants. Each stimulus was followed by one or two questions related to the stimulus. The duration of the individual experiment, including the in-depth interview, was approximately 45 minutes.

The design of the experiment was divided into four thematically different blocks in which stimuli of the same or similar character were grouped together. These blocks were: (i) video, (ii) post, (iii) post comparison, and (iv) news feed. Individual stimuli within each block were randomized and displayed in random order. The order of the blocks was always the same.

SMI BeGaze software was used to evaluate the eye-tracking data. Areas of interest (AOIs) were created in this software. Using these AOIs, the metrics needed to verify the assumptions and hypotheses were obtained. The main metrics tracked for each AOI were dwell time, expressed as the

percentage of time spent observing the selected AOI, and the sequence order of the selected AOI in relation to the other AOIs presented in the given stimulus. Detailed statistical data were generated for all stimuli and processed using Statistica 12 software. Student's t-test and the Mann–Whitney U test were used to analyse the data obtained from the eye-tracking experiment.

In-depth interviews followed immediately after each eye-tracking experiment. Their purpose was to identify participants' shopping habits and preferences in the area of decorative cosmetics, their attitudes towards social networks and influencers, and their perception of promotion through influencer marketing. The interviews also referred to selected stimuli displayed during the eye-tracking experiment in order to interpret the eye-tracking results in more detail and to obtain additional information that could not be captured by the eye-tracking device.

The research focused on women aged 18 to 34 years. A total of 35 participants aged 21 to 31 took part in the experiment. Snowball sampling was used. All participants confirmed that they purchased decorative cosmetics, although the frequency of purchase varied among them. Most often, participants stated that they purchased decorative cosmetics approximately once every two months (10 out of 35 participants) or once a month (9 out of 35 participants). When asked which social networks they used, all participants answered Facebook. The second most frequently used network was Instagram, which was mentioned by 32 participants.

3 RESULTS AND DISCUSSION

In the first stimulus, the impact of user comments on the perceived value of the product was examined. An Instagram post showing an Essence lip gloss together with user comments was used. The nature of the comments differed between the two groups. Group A was exposed to a version of the post containing positive or neutral comments about the product, such as "I am looking forward to it". Group B was exposed to a version containing negative comments, for example comments mentioning the unpleasant smell of the product or problems with colour stability.

The aim was to observe whether attention paid to the comments would differ between the two groups and whether the nature of the comments would be reflected in the perceived value of the product, expressed as the price that participants considered acceptable. The eye-tracking results showed that participants in Group A spent 20.4% of the total stimulus display time looking at the comment section. In Group B, the average dwell time on the comment section was higher, reaching 32.7% of the total stimulus display time.

The average acceptable price also differed between the groups. Participants in Group A stated an average price of CZK 133.47, corresponding to EUR 5.13 when using an exchange rate of approximately CZK 26 per EUR. Participants in Group B stated a lower average price of CZK 97.93, corresponding to EUR 3.76. Because the data did not meet the normality assumption, the Mann–Whitney U test was used to test the price difference. The resulting p-value of 0.0695 was higher than the significance level of $\alpha = 0.05$. Therefore, the difference in the values participants were willing to offer was not statistically significant. Although the p-value was close to the rejection region, the assumed effect of negative comments on the perceived value of the product was not statistically confirmed.

Another stimulus used in the experiment was a post depicting Urban Decay Troublemaker mascara. In version A, the post did not contain hashtags. In version B, a total of 34 hashtags were displayed in the post description, for example #czechgirls, #makeupaddict, #slave2beauty, #makeuplook, and #makeupartist. In this case, the aim was to determine whether the presence of a large number of hashtags would influence the perceived value of the promoted product.

Because the input data did not meet the normality assumption required for the parametric test, the Mann–Whitney U test was used. The resulting p-value of 0.036100 was lower than the selected significance level of $\alpha = 0.05$. It can therefore be stated that there was a statistically significant difference in perceived price between stimulus versions A and B. Based on these findings, the use of a high number of hashtags may have a negative effect on the credibility of the post and may consequently influence the perceived value of the presented product.

The third and fourth stimuli focused on attention paid to the person shown in the Instagram news feed. In the third stimulus, the aim was to determine whether there was a difference in observation time between a photograph showing an influencer without applied make-up and a photograph showing a face with applied make-up. Version A presented the face without applied make-up, while version B presented the same type of visual content with applied make-up, as shown in Figure 1.



Fig. 1: Attention to face without and with applied makeup

Student's t-test was used to test differences in observation time because the data met the normality assumption, as determined by the Shapiro–Wilk test. The resulting p-value was 0.284658. At the significance level of $\alpha = 0.05$, the results did not show that the use of make-up had a statistically significant effect on the observation time of the influencer's face.

The fourth stimulus was used to determine whether there was a difference in observation time between a photograph depicting an unknown person and a photograph depicting a celebrity. As in the previous stimulus, the Instagram news feed format was used. For Group A, a photograph of an unknown person was used. For Group B, the stimulus was modified so that a photograph of Angelina Jolie was presented together with the product.

In version A, participants spent 7.52% of the total stimulus observation time looking at the face. In version B, the observation time devoted to the celebrity's face was slightly higher, reaching 8.85% of the total observation time. The data were normally distributed, and Student's t-test was therefore used. The resulting p-value of 0.334962 indicated that there was no statistically significant difference between the mean observation times for the two stimulus versions at the significance level of $\alpha = 0.05$. Therefore, it can be stated that, in this Instagram news feed stimulus, the use of a celebrity face did not lead to significantly higher attention compared with the use of an unknown face.

Another set of stimuli was presented as pairs of images, from which participants chose the version that attracted them more. To avoid an order effect, the order of the images was reversed between the groups. For Group A, the first image was presented on the left side, while for Group B the same image was presented on the right side.

The content of these stimuli differed in product arrangement. One photograph showed cosmetic products neatly and systematically arranged, while the other showed products chaotically scattered, as if the influencer had taken them out of a handbag. The examined hypothesis concerned whether

there was a difference in observation time between a photograph depicting organized products and a photograph depicting unorganized products.

Based on the result of the Shapiro–Wilk test, which confirmed data normality, Student's t-test was used. The resulting p-value of 0.002697 indicated a statistically significant difference between the observation times of the two photographs at the significance level of $\alpha = 0.05$. The photograph with systematically arranged cosmetic products accounted for 40.86% of the average dwell time, while the photograph with the disordered arrangement accounted for 48.81% of the total observation time. These results suggest that photographs presenting products in an unorganized way may receive more visual attention than photographs showing products systematically arranged. However, participants' stated preferences differed from their visual attention patterns. A total of 23 out of 35 participants stated that they preferred the organized composition, in which the objects were systematically arranged.

Another pair of stimuli presented the same products in two different photographic styles. The aim was to compare attention to a detailed product photograph with attention to a still-life product photograph. As in the previous case, the analysis tested which image received more attention. Because the data met the normality assumption, Student's t-test was used. The resulting p-value of 0.003683 indicated a statistically significant difference in the mean observation times of the two stimuli at the significance level of $\alpha = 0.05$.

The average dwell time for the detailed product photograph was 41.04% of the total observation time, while the still-life product photograph accounted for 48.38% of the total observation time. It can therefore be stated that still-life product photography received more attention than detailed product photography. In addition, 28 participants stated that they considered the still-life version more attractive than the detailed version. Based on this finding, still-life photography appears to be more suitable for attracting attention than a detailed product photograph in this type of presentation.

Audiovisual presentations were also included in the experiment. These consisted of video presentations by influencers focused on fashion and beauty. The first video used in the experiment was a video tutorial in which the presenter demonstrated the use of three products: Catrice HD Liquid Coverage make-up, NYX HD Concealer corrector, and the NYX Ultimate colour palette. The length of the video tutorial was 63 seconds. The first product, Catrice HD Liquid Coverage make-up, was visible for 6,256.25 ms. The second product, NYX HD Concealer corrector, was visible for 2,377.38 ms. The third product, the NYX Ultimate colour palette, was visible for 1,292.96 ms.

For the purpose of A/B testing, the video presented to Group B was modified by removing the original audio, which contained the presenter's spoken commentary on the products. Music was used instead, and no speech was audible. The video presented to Group A was the original version without any modifications. The results for this stimulus are presented in Table 1.

Tab. 1 Attention to products in video tutorial

Product	version	dwell time [ms]	test	p-value
make-up HD liquid	A	4,005.90	t-test	0.466899
	B	4,251.44		
corrector HD	A	351.80	Mann-Whitney U test	0.105827
	B	544.53		
color palette	A	1,160.00	Mann-Whitney U test	0.228328
	B	1,246.59		

When looking at the mean values, there appeared to be some differences between the version with spoken commentary and the version without spoken commentary. For the first product, the version without spoken commentary had an average dwell time that was 245.54 ms longer. For the second product, the dwell time was 192.79 ms longer in the version without spoken commentary. The smallest difference was observed for the third product, where the average dwell time differed by 86.59 ms.

However, according to the statistical tests, there was no statistically significant difference in dwell time between versions A and B. Thus, the character of the audio background did not have a statistically significant effect on the observation time of the products presented in the tutorial video.

The second video used in the experiment focused on the perception of a male reviewer of decorative cosmetics. This video was not modified for A/B testing. Instead, it was presented in its original version to all participants in order to collect their opinions on this form of presentation of products primarily targeted at women. The protagonist of the video was Jeffree Star, an American YouTuber listed among influential beauty influencers. In the video, the use of Maybelline Age Rewind liquid concealer was demonstrated. The video was 54 seconds long, and the presented product was visible for 11,011 ms. The mean dwell time on the presented product was 3,643.14 ms.

During the interview that followed the video presentation, participants were asked how they perceived this type of product presentation. In total, 13 out of 35 participants stated that male reviewers of decorative cosmetics did not fit this product category, seemed unusual to them, or bothered them. Six participants expressed a neutral attitude. However, the largest group of participants, 16 out of 35, reacted positively to this type of presentation. Seven participants mentioned, based on their own experience, that male reviewers can often demonstrate make-up application better than some female reviewers.

In general, 31 out of 35 participants perceived reviews by decorative cosmetics influencers positively. These participants considered such product presentations to be a useful source of inspiration. Participants usually did not base their purchase decisions on only one review, but tried to collect more information or compare multiple reviews. During the interviews, participants also mentioned that when a reviewer is critical and attempts to be objective, it increases the credibility of the influencer. Participants also appreciated reviews that demonstrate products in real-life situations.

Four participants perceived influencer reviews negatively. They mentioned that there were too many such presentations and that they considered them a waste of time. Almost two-thirds of participants, namely 23 out of 35, stated that they had been consciously influenced by influencers when purchasing decorative cosmetics. Half of the participants, 17 out of 35, stated that they had purchased a decorative cosmetics product directly based on an influencer's recommendation.

According to the interviews, 31 out of 35 participants preferred influencers to disclose paid cooperation in their social media posts. Three participants stated that they did not care about this information, and one participant stated that influencers had the right not to disclose paid partnership in the post.

4 CONCLUSION

The eye-tracking research showed that, in visually oriented social media environments such as Instagram, consumers pay attention not only to visual content but also to accompanying text. The findings suggest that consumers may, in some cases, pay considerable attention to textual elements of a post. This implies that influencers should focus not only on the photograph or video itself, but also on the quality, originality, and credibility of the accompanying description.

The use of an excessive number of hashtags appeared to be inappropriate in the context of the examined stimuli. Based on the findings, it is recommended to use only a limited number of hashtags that are relevant to the post and its content. Although textual elements received a considerable degree of attention, the eye-tracking results also showed that participants usually looked at the photograph first. Therefore, if the visual aspect of the post does not attract consumers' attention immediately, there is a risk that they may skip the post without reading the accompanying text.

The eye-tracking results further showed that products may receive more attention when they are presented in a loosely scattered or unorganized arrangement. However, many participants stated that they preferred images in which products were systematically arranged. This suggests that visual attention and stated preference do not necessarily correspond. Similarly, still-life product photography received more attention than detailed product photography and was also evaluated as more attractive by most participants.

The results also showed that the use of a famous person's face did not lead to significantly higher visual attention than the use of an unknown person's face. Likewise, a face without applied make-up did not attract significantly more attention than a face with applied make-up. In the case of audiovisual stimuli, the presence or absence of spoken commentary did not have a statistically significant effect on the observation time of the presented products.

The follow-up interviews showed that most participants perceived influencer reviews of decorative cosmetics positively and considered them a useful source of inspiration. Participants appreciated reviews that appeared authentic, critical, objective, and based on real product use. A substantial proportion of participants also stated that they had been influenced by influencers when purchasing decorative cosmetics. Finally, the interviews indicated that most participants preferred influencers to disclose paid cooperation in their social media posts.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ASLAM, S. 2019. Instagram by the Numbers: Stats, Demographics & Fun Facts. *Omnicore*. https://www.omnicoreagency.com/instagram-statistics/

CONICK, H. 2018. How to win friends and influence millions: the rules of influencer marketing. *Marketing News*. (5)11, 36–45. ISSN 0025-3790.

CPC STRATEGY. 2018. *The 2018 influencer marketing report*. CPC Strategy. https://learn.cpcstrategy.com/rs/006-GWW-889/images/2018%20Influencer%20Marketing%20v3.pdf

HENNESSY, B. 2018. *Influencer: building your personal brand in the age of social media*. New York, NY: Citadel Press, Kensington Publishing. ISBN 978-0-8065-3885-3.

JANOUCH, V. 2014. *Internetový marketing*. 2nd Edition. Brno: computer press. ISBN 978-80-251-4311-7.

KEMP, S. 2018. The state of the internet in Q4 2018. *We Are Social*. https://wearesocial.com/blog/2018/10/the-state-of-the-internet-in-q4-2018

MEDIAKIX. 2018. A comprehensive guide to beauty influencer programs. Mediakix. Available from: http://mediakix.com/2018/01/beauty-influencer-programs-types-examples-guide/#gs.193n15

NEWBERRY, C. Influencer Marketing: How to Work With Social Media Influencers in 2019. *Hootsuite*. https://blog.hootsuite.com/influencer-marketing

PITZ, J., KÖHLER, I., ESCH, F.-R. 2018. The impact of influencer marketing on perception and experience of unknown, weak and strong brands. *Transfer: Werbeforschung & Praxis*. 64(4): 14–24.

RAJPUT, N. 2016. Cosmetics Market Report. *Allied Market Research*. https://www.alliedmarketresearch.com/cosmeticsmarket

RELATABLE. 2019. The 2019 State of Influencer Marketing Report. *Relatable*. https://www.relatable.me/the-state-of-influencer-marketing-2019

ŘEZNÍČEK, J., PROCHÁZKA, T. 2014. *Obsahový marketin*g. Brno: Computer Press. ISBN 978-80-251-4152-6

STATISTA. 2018. Annual growth of the global cosmetics market from 2004 to 2018*. *Statista*. https://www.statista.com/statistics/297070/growth-rate-of-the-global-cosmetics-market/

STOKES, Chris. 2017. Social Influencers: The Psychology Behind Great Influencer Marketing. *PMYB*. https://pmyb.co.uk/social-influencers-psychology/

ZAJONC, R. B. 2001. Mere Exposure: A Gateway to the Subliminal. *Current Directions in Psychological Science.* 10(6), 224-228. ISSN 0963-7214.

Contact information

Kateřina Klátilová: e-mail: katerinaklatilova@gmail.com
Stanislav Mokrý: e-mail: xmokry@mendelu.cz
Jana Turčínková: e-mail: jana.turcinkova@mendelu.cz
Michal Pšurný: e-mail: michal.psurny@mendelu.cz

A TOOL FOR AUTOMATICALLY TRANSCRIBING IN-TERVIEWS WITH SPEAKER RECOGNITION

Lukáš Kleveta[1], Jaromír Landa[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_012

ABSTRACT

The present paper presents the design and implementation of a software tool for automatically transcribing audio recordings, supplemented by speaker recognition (diarization). The tool integrates two open-source components: The implementation of OpenAI's Whisper AI model for speech-to-text conversion and the pyannote.audio library or, alternatively, NVIDIA NeMo - for speaker recognition. The architecture of the system is implemented as a linear pipeline with modules for audio preprocessing, diarization, transcription, and result merging. It is possible to modify key processing parameters (model, language, thresholds, clustering method) via a configuration file without modifying the code. The functionality was verified on two test sets. The first was a short English recording with a single speaker. In this case, a quantitative comparison of Whisper AI models was made using the Levenshtein metric. The second was a one-minute recording with two speakers. In this case, a comparison was made of twenty configurations in terms of transcription, diarization, and computational complexity. The medium and large-v2 models demonstrated the highest degree of mutual agreement among transcriptions (97.79%), while the lowest was exhibited by the tiny and large-v3 models (82.88%). With regard to overall balance, the configuration employing the large-v2 model and the NeMo diarization module with spectral clustering was identified as the optimal solution. The findings substantiate the practical viability of open-source instruments for automated transcription. However, they concurrently underscore the constraints associated with multilingual recordings, the computational intricacy, and the susceptibility of diarization to variations in audio quality.

Keywords: automatic speech transcription, speaker diarization, Whisper, pyannote.audio, NVIDIA NeMo, audio processing, open-source, Levenshtein distance, pipeline
JEL Code: C88, L86, O33

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

In recent years, the field of automatic speech processing has witnessed substantial advancements, largely attributable to the emergence of machine learning models. The capacity to transform spoken language into a structured text transcript, inclusive of speaker identification, engenders a plethora of pragmatic applications. These encompass the indexing of audiovisual archives, the analysis of conference recordings, the facilitation of journalistic and editorial endeavours, and the processing of qualitative data in research (Miró *et al.*, 2012; Park *et al.*, 2022).

An open-source model that has fundamentally pushed the boundaries of accessibility in this field is OpenAI's Whisper. The model is an encoder-decoder system with a Transformer architecture that has been trained on 680,000 hours of predominantly multilingual audio, a fact that has enabled it to deliver robust performance across a range of accents and background noise (Radford *et al.*, 2023). Whisper AI is distributed as open-source, allowing for free deployment in both academic and commercial projects without the need to adapt the model to specific input.

Speech transcription naturally involves the task of speaker identification, known as diarization. Diarization provides an answer to the question of the speaker's identity and temporal segmentation of the recording based on the speakers' identities (Tranter and Reynolds, 2006). Current approaches encompass classical clustering algorithms based on i-vectors or x-vectors, as well as end-to-end neural models (Park *et al.*, 2022). An open-source tool for this task is the Python library pyannote.audio, which combines neural segmentation with agglomerative clustering of speaker vectors (Bredin *et al.*, 2020).

Notwithstanding advances in individual sub-tasks, the successful deployment of integrated transcription and diarization systems remains non-trivial. The following persistent challenges have been identified: short speech segments, overlapping speech, noise, and sensitivity to sampling rate and audio compression (Miró *et al.*, 2012). The practical implementation of this technology is further complicated by differences in speaker characteristics due to factors such as age, gender, and accent (Church *et al.*, 2017). Furthermore, the synchronisation of the outputs of two independent models – one for transcription and one for diarisation – generates a distinct class of errors related to the temporal alignment of segments.

This paper summarises the results of a bachelor's thesis whose objective was to design, implement, and test a functional open-source tool that combines these components into a coherent pipeline. The emphasis was placed on three aspects: The system under discussion features the following characteristics:

- Modular architecture, which allows individual processing steps to be swapped without affecting the rest of the system.
- Configurability of key transcription and diarization parameters via an external YAML file.
- Robustness when processing longer recordings using overlapping segmentation.

The impetus for this endeavour is rooted in the mounting demand for automated tools to facilitate the transcription of textual data derived from qualitative research methods. The transcription of an hour-long interview, a task typically undertaken by a qualified transcriber, traditionally necessitates several hours of labour. The implementation of an automated solution has the potential to markedly reduce the time investment required, albeit at the expense of necessitating human review of the generated transcription. The reliability of such automation is contingent upon a combination of the transcription model, the diarization model, and the processing parameter settings.

The proposed paper will address the following research questions. The first question is to what extent the transcriptions generated by different Whisper AI model sizes differ for the same recording. The second question pertains to the identification of the most effective combinations of configuration parameters (Whisper AI model size, temperature parameter value, choice of diarization module, and clustering method) that yield optimal results in terms of transcription quality, diarization accuracy, and computational complexity. The objective of this research is not to attain state-of-the-art results on standard benchmark data, but rather to ascertain the viability of constructing a practically usable tool from freely available components and to delineate its vulnerabilities.

The remainder of the paper is structured as follows. The Methodology and Data chapter provides a comprehensive overview of the pipeline architecture, the technologies employed, audio preprocessing, the principles of diarization and transcription, and the evaluation metrics. The Results chapter presents a quantitative comparison of Whisper AI models using the Levenshtein metric and a comparative test of twenty configurations on a longer recording. The final section summarises the main findings, the limitations of the chosen approach, and directions for further research.

2 METHODOLOGY AND DATA

The proposed system is to be implemented as a linear pipeline consisting of four main modules. The initial module, designated as AudioProcessor, is responsible for the preliminary processing of the input audio. The second module (Diarizer) performs speaker recognition. The third module, designated as Transcriber, facilitates the conversion of speech into written form. The fourth module (ResultsMerger) functions to merge the diarization and transcription outputs, thereby generating the final structured output in the selected format (TXT, JSON, SRT). The communication infrastructure is characterized by the implementation of clearly defined data interfaces, facilitating seamless interaction between modules. A comprehensive logging system is in place to ensure meticulous documentation of each phase, thus facilitating diagnostic processes. The architecture diagram is shown in Figure 1.

Audio preprocessing can be delineated in three distinct stages. The initial procedure entails the implementation of noise reduction through the utilization of the spectral subtraction method, with the ‘prop_decrease’ parameter set to 0.75. This parameter serves to specify the proportion of the identified noise that is to be subtracted from the signal. It has been demonstrated that an increase in value results in enhanced intelligibility; however, it is important to note that overly aggressive reduction can lead to the distortion of speech segments. The subsequent stage involves amplitude normalization along the channel dimension of the tensor, thereby ensuring equilibrium of the volume. The third step in the process is the conversion of the audio file to a standardised WAV format, with a sampling rate of 16 kHz and a single channel (mono). This is a prerequisite for the Whisper AI model. For recordings exceeding a specified size, a segmentation technique with configurable overlap is implemented. This approach serves to mitigate the risk of context loss at boundaries, thereby ensuring the integrity of the recorded data.

The diarization module has been designed to allow for extensibility. The abstract class, designated as BaseDiarizer, establishes an interface that is subsequently implemented by specific modules. Two such implementations are available within the system: the first is labelled `PyAnnoteDiarizer`, and employs the `pyannote.audio` library with the pre-trained model `pyannote/speaker-diarization-3.1`; the second is labelled `NemoDiarizer`, and is built on top of the NVIDIA NeMo library.

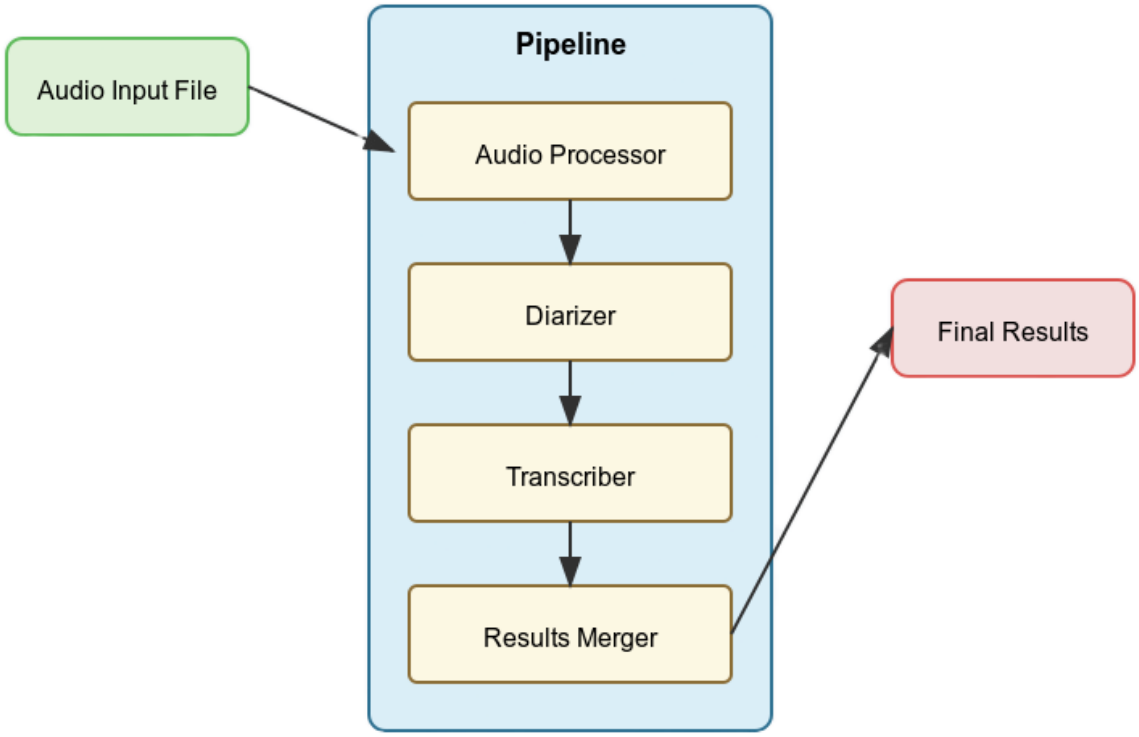


Fig. 1: Diagram of a linear pipeline for audio processing with diarization and transcription (Audio Input → AudioProcessor → Diarizer → Transcriber → ResultsMerger → Final Results) (Kleveta, 2025)

The module is selected through the use of a configuration file and the DiarizationFactory class. The present version of the PyAnnote pipeline (Bredin *et al.*, 2020) operates in three phases: local segmentation using five-second sliding windows, speaker vector extraction, and global agglomerative clustering (Bredin, 2023). NeMo employs a similar methodology, utilizing its own models for voice activity detection (vad_multilingual_marblenet), embedding extraction (titanet_large), and spectral clustering. The output in both cases is a list of time segments with assigned speaker identifiers.

The transcription module utilizes Whisper AI models of varying sizes, ranging from the compact model (39 million parameters) to the large-v3 model (1.55 billion parameters), in addition to an optimized turbo variant. The key parameters that must be considered include the task selection (transcribe or translate), temperature (which affects output variability), compression_ratio_threshold (for detecting excessive repetitiveness), and no_speech_threshold (which sets the threshold for detecting silence). By default, the temperature is set to 0, thereby ensuring the output is deterministic. In the absence of user input regarding the desired language, the system will automatically detect the language of the audio. In order to utilize GPU acceleration via CUDA technology, the system automatically switches to the hardware variant if available.

The evaluation of transcription quality was conducted through the utilization of two complementary metrics. The standard metric employed is the Word Error Rate (WER), defined as the ratio of the sum of word substitutions, insertions, and deletions to the number of words in the reference transcription (Ali and Renals, 2018). For languages with complex morphology or ambiguous word boundaries, the Character Error Rate (CER), which operates at the character level, is

more suitable (Thennal *et al.*, 2024). In the comparison performed, the Levenshtein distance was additionally utilized. This is defined as the minimum number of insertions, deletions, and character substitutions required to transform one string into another (Haldar and Mukhopadhyay, 2011). Computationally, it is implemented using dynamic programming on a matrix of dimensions (m+1)×(n+1). This metric enables a direct comparison of transcription pairs without the necessity for a reference "true" transcription. The Diarization Error Rate (DER) is a metric employed for the evaluation of diarization processes. It calculates the time proportions of false detections, omitted speech, and speaker identity mix-ups (Bredin, 2017). Supplementary metrics, including purity and coverage, facilitate the discernment of errors arising from excessive or insufficient segmentation.

The test data consists of two recordings. The initial recording is a concise, approximately 10-second recording in English featuring a sole speaker. It is characterized by clarity and absence of substantial noise, and is employed for the purpose of comparison of Whisper AI models through the utilization of the Levenshtein metric. The second recording is approximately one minute in duration and features two speakers, with the first speaker dominating for approximately 90% of the time. The audio is characterized by clarity and fluidity in the enunciation of American English, with subtle indications of Canadian influences. A total of twenty distinct configuration files were subjected to rigorous testing on this recording, with a particular emphasis on alterations to transcription parameters (temperature, thresholds), the selection of diarization module, embedding model, and clustering method. The limitations of the test set are discussed in the concluding chapter.

3 RESULTS

The initial experiment focused on comparing six multilingual Whisper AI models (tiny, base, small, medium, large-v2, large-v3) using a 10-second English test recording. For each pair of models, the Levenshtein distance between their transcriptions and the corresponding percentage match was calculated.

The highest mutual match was observed between the medium and large-v2 pair (97.79%, Levenshtein distance equal to 3). The transcriptions differed only in the ending of the last sentence. In contrast, the lowest match was observed between the tiny and large-v3 models (82.88%, distance 25). This outcome merits further consideration. It is a commonly held belief that a larger model should always provide a more accurate transcription. However, a qualitative analysis reveals a more complex picture.

In the test recording, the tiny model terminated the transcription at the word "world", which corresponded to the actual end of the utterance. In contrast, the large-v3 model incorporated the sentence "had been destroyed," which is not present in the original recording. This phenomenon is referred to as a hallucination, where the model generates coherent but contextually unsupported text. Conversely, smaller models tend to exhibit a distinct category of errors, including the generation of nonsensical word forms (e.g., "und" in the transcription of the tiny model) and punctuation errors.

This analysis suggests the following partial conclusion: for brief, linguistically well-defined recordings, increasing the model size does not guarantee an improvement in the quality of the output. Whilst larger models have been shown to excel in the identification of punctuation, sentence structure and complex vocabulary, there is also an increased risk of hallucinations with increasing capacity. With regard to consistency, the medium and large-v2 pair demonstrated the greatest stability.

The second experiment focused on evaluating the impact of configuration parameters on a more complex test recording featuring two speakers. Twenty configuration variants were prepared, which sequentially varied the size of the Whisper AI model, the value of the temperature parameter, the diarization module used (pyannote vs. NeMo), the embedding model (titanet_large vs. resnet), and the clustering method (spectral, sc, ahc). The evaluation was based on four criteria: content consistency of the transcription, diarization quality, output stability, and computational complexity. Table 1 provides a synopsis of the five optimal configurations.

The most balanced results were achieved by the Config_14 configuration, which was the only one to achieve the maximum score in both the transcription and diarization categories, albeit at the cost of a slightly lower performance score. The most accurate transcription was achieved by the Config_11 configuration, which combined the large-v2 model with a low temperature parameter value and optimized thresholds for compression_ratio_threshold and logprob_threshold. Although the Config_19 configuration showed comparable diarization quality, it failed to detect the secondary speaker, and its performance score was the lowest among the five configurations compared.

Three observations emerge from the analysis. First, lower values of the temperature parameter (typically 0 to 0.2) lead to more stable and contextually consistent transcriptions, while higher values increase output variability and, in some cases, generate less accurate formulations. Second, the NeMo diarization module achieved higher quality in the tests than the baseline pyannote.audio, particularly in distinguishing between two speakers on a recording with an unbalanced speaking share. Third, the choice of clustering method significantly affects the diarization output: spectral clustering (sc) and agglomerative hierarchical clustering (ahc) outperformed simpler approaches.

Drawing upon the findings of individual case studies, the following proposed configuration was put forward: the Whisper AI large-v2 model, the transcribe task, a temperature set to 0.2, the NeMo diarization module with the titanet_large embedding model, spectral clustering, and VAD parameters set to threshold = 0.45 and window_length_in_sec = 0.15. This configuration combines the strengths of successful variants from the test set; however, it cannot be claimed that this is the universally best setting for any type of recording. The generalisability of the findings is limited by two factors. Firstly, the size of the test set is small. Secondly, the predominance of the English language in the test data is also a factor (see Discussion and Conclusions section).

Tab. 1 Comparison of the best five configurations out of twenty tested. Values on a scale of 1–5; the "Second speaker" column indicates whether the presence of a second speaker was correctly detected. (Kleveta, 2025)

Configuration	Transcription	Diarization	Performance	Number of speakers	Second speaker	Time (s)
Config_14	5.0	5.0	4.0	2	Yes	12.5
Config_11	4.6	4.9	4.5	2	Yes	11.2
Config_03	4.8	4.7	4.5	2	Yes	10.3
Config_19	4.0	4.7	2.5	2	No	13.7
Config_20	4.4	4.5	3.8	2	Yes	10.9

4 DISCUSSION AND CONCLUSIONS

The findings of the experiments demonstrate that by integrating the open-source components Whisper AI and pyannote.audio, or NVIDIA NeMo, it is feasible to construct a functional tool for automatic speech transcription with speaker recognition, yielding a practical output quality for concise, intelligible recordings of spoken English. The modular pipeline architecture ensures that individual components can be exchanged without affecting other parts of the system, and the external YAML configuration allows for parameter tuning without modifying the source code.

Concurrently, the experiments conducted revealed several significant limitations. Of these, the size of the test set itself is of the greatest importance. The initial experiment is founded upon a solitary 10-second recording, whilst the subsequent experiment is predicated upon a 60-second recording. Consequently, any conclusions that are drawn about the relative quality of the models and configurations must be generalised with the understanding that they may not be universally applicable. In order to perform a more robust comparison, it would be necessary to use a more extensive and linguistically diverse test set, ideally an established benchmark with a standardized reference. In this sense, the present study is more of an exploratory feasibility study than a rigorous evaluation.

A secondary constraint pertains to the system's reliance on two distinct models: one for transcription and the other for diarization. This approach is not without risk of temporal mismatches between the outputs of the two modules, which can lead to inaccurate assignment of text segments to individual speakers, particularly near speaker-switch boundaries. This problem is well-documented in the academic literature and has motivated the development of integrated approaches such as end-to-end neural diarization (Park *et al.*, 2022) or post-processing diarization using language models (Wang *et al.*, 2024). However, the implementation of such approaches would result in a substantial increase in the requirements for training data and computational power.

The third limitation pertains to the behaviour of Whisper AI models when confronted with multilingual content. In the course of the testing process, it was repeatedly observed that the model consistently identified the language of the entire recording as English, even in cases where the recording contained segments in other languages. The model then proceeded to translate or completely mis-transcribe these segments. In certain instances, the phenomenon of "stuck" behaviour manifested, wherein the subsequent segments of transcription became disassociated from the underlying content following the occurrence of a problematic segment. This issue persisted even when the transcribe task was explicitly set. In multilingual contexts, which are prevalent in the European environment, this behaviour poses a significant obstacle to practical deployment. Current developments in the field of massively multilingual ASR are paving the way for models covering hundreds of languages (Pratap *et al.*, 2024). However, a separate task beyond the scope of this paper is the systematic comparison of these models with Whisper AI models on recordings containing Central European languages.

The fourth limitation pertains to computational complexity. It has been demonstrated that larger Whisper AI models (large-v2, large-v3) require GPU acceleration when combined with NeMo diarization in order to process data within a reasonable timeframe. In a purely CPU-based environment, processing time increases to such an extent that practical application to longer recordings is rendered unfeasible. The necessity to divide extended recordings into overlapping segments gives rise to a further source of error, namely the potential loss of contextual information at segment boundaries, which may consequently lead to inconsistencies in the identification of the same speaker across adjacent segments.

From the perspective of practical application, these results have several implications. The utilisation of automated transcription with diarization facilitates the reduction of the time required for the preparation of text transcripts in qualitative research, journalism, podcast production, and audiovisual archive management. However, none of the tested systems provides an output that could be published directly without human review. Errors in punctuation, capitalisation of proper names, and hallucinations in larger models require final editing. The potential for real-world time savings is therefore contingent on the availability of an editor proficient in the target language and the subject matter of the recording.

It is possible to identify a number of potential avenues for the tool's future development. The initial approach involves the implementation of an end-to-end model that integrates diarization and transcription, with the objective of resolving the issue of temporal alignment (Park *et al.*, 2022). The second is the implementation of a visual user interface that will enable interactive correction of the transcript and re-assignment of segments to speakers. The third direction is deployment in a cloud environment with flexible allocation of GPU resources, which would lower the barrier to use for users without their own powerful hardware. The fourth direction is systematic evaluation on a larger and more linguistically diverse dataset, which would allow for more reliable conclusions about the relative performance of the configurations.

In conclusion, it can be stated that the proposed pipeline fulfils the objective of this bachelor's thesis—namely, to verify whether it is possible to build a practical tool for automatic transcription of recordings with speaker recognition using freely available open-source components. The experiments conducted confirm that the choice of the Whisper AI model and the configuration of the diarization module have a significant impact on output quality. The most balanced results on the test data were achieved by a configuration combining the Whisper AI large-v2 model with the NeMo diarization module and spectral clustering. The identified limitations – chiefly the modest proportions of the test set, challenges pertaining to multilingual content, computational intricacy, and errors at segment boundaries – delineate the agenda for future research. In view of the rapid advances being made in the field of speech models, it is reasonable to assume that some of the aforementioned limitations will be addressed in the near future by subsequent generations of open-source models.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

Several AI tools were used in the preparation of this article. They assisted with summarizing source material, translating passages between languages, and identifying typographical and stylistic issues in the text. All output produced by these tools was reviewed and verified. Summaries were checked against the original sources, translations were assessed for accuracy, and suggested corrections were evaluated before being accepted. The tools served as an aid to the writing process; the final content reflects the authors' judgment and remains the authors' responsibility.

REFERENCES

ALI, A., RENALS, S. 2018. Word error rate estimation for speech recognition: e-WER. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne: Association for Computational Linguistics, pp. 20–24.

BREDIN, H. 2017. pyannote.metrics: a toolkit for reproducible evaluation, diagnostic, and error analysis of speaker diarization systems. In: *Proceedings of Interspeech 2017*. Stockholm: ISCA. https://doi.org/10.21437/interspeech.2017-411

BREDIN, H. 2023. pyannote.audio 2.1 speaker diarization pipeline: principle, benchmark, and recipe. In: *Proceedings of Interspeech 2023*. Dublin: ISCA.

BREDIN, H. *et al*. 2020. pyannote.audio: neural building blocks for speaker diarization. In: *ICASSP 2020 – IEEE International Conference on Acoustics, Speech and Signal Processing*. Barcelona: IEEE. https://doi.org/10.1109/ICASSP40776.2020.9052974

CHURCH, K. *et al*. 2017. Speaker diarization: a perspective on challenges and opportunities from theory to practice. In: *ICASSP 2017 – IEEE International Conference on Acoustics, Speech and Signal Processing*. New Orleans: IEEE. https://doi.org/10.1109/ICASSP.2017.7953098

HALDAR, R., MUKHOPADHYAY, D. 2011. Levenshtein distance technique in dictionary lookup methods: an improved approach. arXiv preprint. *arXiv*:1101.1232. https://doi.org/10.48550/arXiv.1101.1232

KLEVETA, L. 2025. *Nástroj pro automatické přepisy rozhovorů s rozpoznáním řečníků*. Bachelor's thesis. Brno: Mendel University in Brno, Faculty of Business and Economics.

MIRÓ, X. A. *et al*. 2012. Speaker diarization: a review of recent research. *IEEE Transactions on Audio, Speech, and Language Processing*. 20(2), 356–370. https://doi.org/10.1109/TASL.2011.2125954

PARK, T. J., KANDA, N., DIMITRIADIS, D. 2022. A review of speaker diarization: recent advances with deep learning. *Computer Speech & Language*. 72, 101317. https://doi.org/10.1016/j.csl.2021.101317

PRATAP, V., TJANDRA, A., SHI, B. *et al*. 2024. Scaling speech technology to 1,000+ languages. *Journal of Machine Learning Research*. 25.

RADFORD, A., KIM, J. W., XU, T., BROCKMAN, G., MCLEAVEY, C., SUTSKEVER, I. 2023. Robust speech recognition via large-scale weak supervision. In: *Proceedings of the 40*th *International Conference on Machine Learning (ICML 2023)*. PMLR, vol. 202.

THENNAL, D. K., JAMES, J., GOPINATH, D., ASHRAF, M. 2024. Advocating character error rate for multilingual ASR evaluation. arXiv preprint. *arXiv*:2410.07400. https://doi.org/10.48550/arXiv.2410.07400

TRANTER, S. E., REYNOLDS, D. A. 2006. An overview of automatic speaker diarization systems. *IEEE Transactions on Audio, Speech, and Language Processing*. 14(5), 1557–1565. https://doi.org/10.1109/TASL.2006.878256

WANG, Q., HUANG, Y., ZHAO, G., CLARK, E., XIA, W., LIAO, H. 2024. DiarizationLM: speaker diarization post-processing with large language models. arXiv preprint *arXiv*:2401.03506. https://doi.org/10.48550/arXiv.2401.03506

Contact information

Lukáš Kleveta: e-mail: xkleveta@node.mendelu.cz

Jaromír Landa: e-mail: jaromir.landa@mendelu.cz

INTEGRATION OF CHRONOS HIGH-SPEED CAMERAS INTO THE MERCURYRT DIGITAL IMAGE CORRELATION SYSTEM

Martin Kmenta[1]

[1]Department of Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/10.11118/ecos2026_013

ABSTRACT

This article deals with the design and implementation of an integration module that connects Chronos 1.4 and 2.1-HD high-speed cameras with the MercuryRT digital im-age correlation (DIC) system. The work responds to the previously missing integration that required manual transfer of recordings and separate setup of both systems. The proposed solution uses the camera REST API for control and an RTSP stream decoded by an external FFmpeg process for live preview. A custom integer-based YUV420p-to-RGB conversion algorithm in C# was implemented; it handles a 10 FPS preview in real time with typical latency around 300 ms. Functionality and stability were verified by a four-hour stress test and trigger measurements using an LED flash; the recording completed without frame loss.

Keywords: high-speed imaging, digital image correlation, REST API, RTSP, FFmpeg
JEL Code: C88

1 INTRODUCTION

Digital image correlation (DIC) is a contactless method for surface deformation measurement that obtains full-field deformation maps from a pair of images captured before and after loading, replacing point-based sensors. Sutton *et al.* (2009) summarized the mathematical foundations of the method; later studies applied it to crack monitoring in wood (Tippner *et al.* 2019), in sheet metal materials (Kořeček *et al.* 2022) and to volume change measurement using 2D DIC (Kotlánová *et al.* 2023). Combining DIC with a high frame rate enables analysis of fast events — from component vibration (Beberniss and Ehrhardt 2016) to dynamic failure. The cost barrier of high-speed imaging was lowered by the low-cost CMOS camera design of Duke *et al.* (2018, 2019), which forms the basis of the Chronos series from the Canadian company Kron Technologies.

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

The DIC laboratory at the Department of Informatics, Mendel University, uses the MercuryRT software by Mercury MS s.r.o. Version 3.1.7 supports a range of industrial cameras (Basler, FLIR, Photon), but no integration module existed for Chronos. Without it, the user had to switch between the camera interface and the application and transfer frames manually via a network share. The aim of this work was to remove this gap and enable camera control directly from the environment where the results are evaluated.

1.1 Chronos Communication Interface

The Chronos camera contains an embedded Linux computer and exposes a REST API version 0.7.0 via the /control path. Fielding (2000) described the architecture of HTTP-based web services in his dissertation; this is the conceptual basis of REST used by the camera API. For preview the camera provides an RTSP stream defined in RFC 2326 (Schulzrinne *et al.* 1998) in H.264 format (Richardson 2011). The stream is encoded in the YUV420p pixel format, which separates the luminance component Y from the chrominance components U, V and stores colour at half resolution; this saves bandwidth but requires conversion to RGB for most computer vision tasks.

1.2 MercuryRT Modularity

MercuryRT is built around the abstract CameraBase class, which defines the minimum interface for any connected camera (IsConnected, OpenCamera(), CloseCamera(), NewFrame). High-speed devices add the IHighSpeedCamera interface with methods for sequential recording

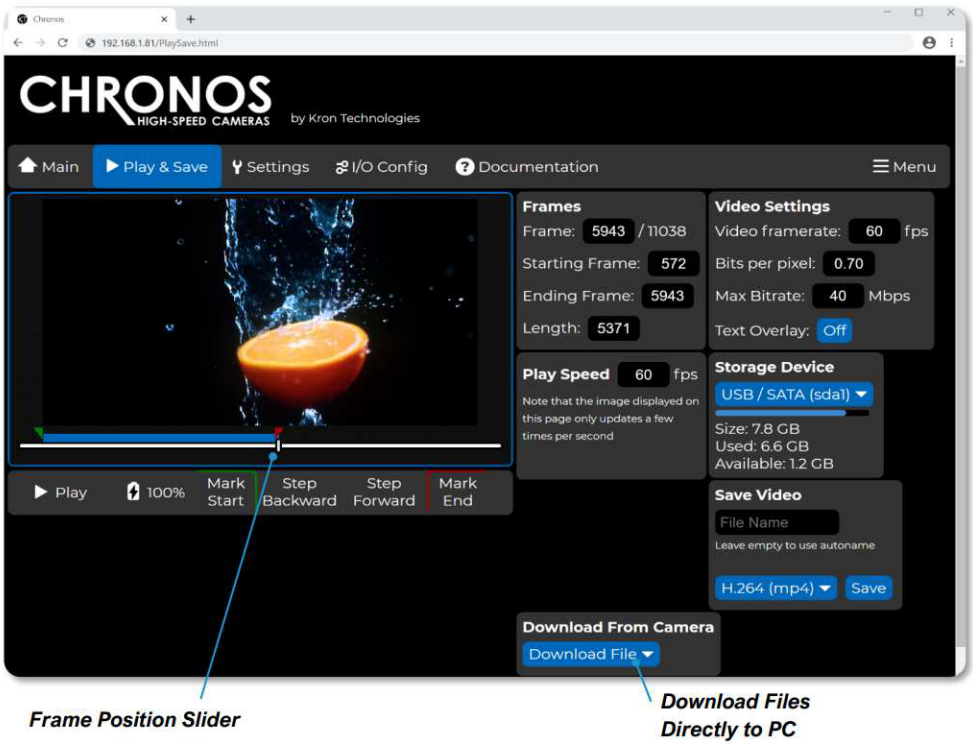


Fig. 1: Chronos web user interface — the Play & Save tab for selecting the frame range, output format and recording download; the left panel shows the live preview of a high-speed scene

(ReadyCameraRecording(), TriggerCameraRecording(), RetrieveHsFrame()). A class for a new camera must implement both interfaces — this automatically integrates the new type into the existing pipeline (probes, storage, correlation analysis) without touching the application core.

1.3 Integration Module Architecture

Three cooperating classes were designed for Chronos. ChronosCamera inherits from CameraBase and implements IHighSpeedCamera; its constructor accepts the IP address and verifies camera availability by querying the model and serial number. ChronosCameraAPI wraps all REST calls (getRequest(), postRequest()) and a Server-Sent Events (SSE) listener for asynchronous notifications about recording completion. ChronosCameraGrabber runs in its own thread, launches an external FFmpeg process, reads raw YUV420p data from its stdout and converts them to RGB frames. The control channel (REST) is thus separated from the data channel (RTSP) — each runs independently and a failure in one does not block the other.

The recording sequence differs from cameras that stream every frame in real time. Chronos records into internal RAM and only allows download after the sequence ends: (i) the user sets parameters, (ii) the startRecording call starts a circular buffer, (iii) after the trigger the camera completes the recording once post-trigger frames are filled, (iv) startFilesave converts data from RAM to a file (MP4/H.264, DNG, TIFF) and (v) the MercuryRT Capture Tool offers HTTP download and immediate import into the project.

2 REAL-TIME YUV420P TO RGB CONVERSION

The YUV420p pixel format stores one byte of luminance Y per pixel while the U and V chrominance components are shared by four neighbouring pixels (4:2:0 subsampling); the total data volume is 1.5 bytes per pixel. For a resolution of 1280 × 1024 at 60 FPS this means about 120 MB/s — naive iteration with floating-point arithmetic proved insufficient. The implemented algorithm went through three optimization steps: (i) conversion coefficients (1.164, 1.596, 2.017) were stored as constants, (ii) recomputed to integers by scaling with a factor of 2^8=256 (multiplication by 1.164 was replaced by multiplication by 298 with a subsequent right shift by 8 bits), (iii) the loop was placed in an unsafe block with pointer arithmetic. The resulting conversion is:

$$\begin{aligned} Y' &= 298 \cdot (Y - 16) + 128 \\ R &= (Y' + 409 \cdot (V - 128)) \gg 8 \\ G &= (Y' - 100 \cdot (U - 128) - 208 \cdot (V - 128)) \gg 8 \\ B &= (Y' + 516 \cdot (U - 128)) \gg 8 \end{aligned} \tag{1}$$

Results are clamped to the 0-255 range. The grayscale version for DIC modules is computed as a weighted sum $0.299R+0.587G+0.114B$ for compatibility with the other cameras supported by MercuryRT.

External RTSP decoding is handled by FFmpeg launched with parameters `-r 10 -f image2pipe -pix_fmt yuv420p -vcodec rawvideo -`. Limiting the rate to 10 FPS prevents overloading the decoder; higher frame rates produced increased latency and artifacts. The first SSE reading variant using a long-running `HttpWebRequest` proved unstable — the robust implementation waits in a loop for event: and data: lines and re-establishes the connection if interrupted.

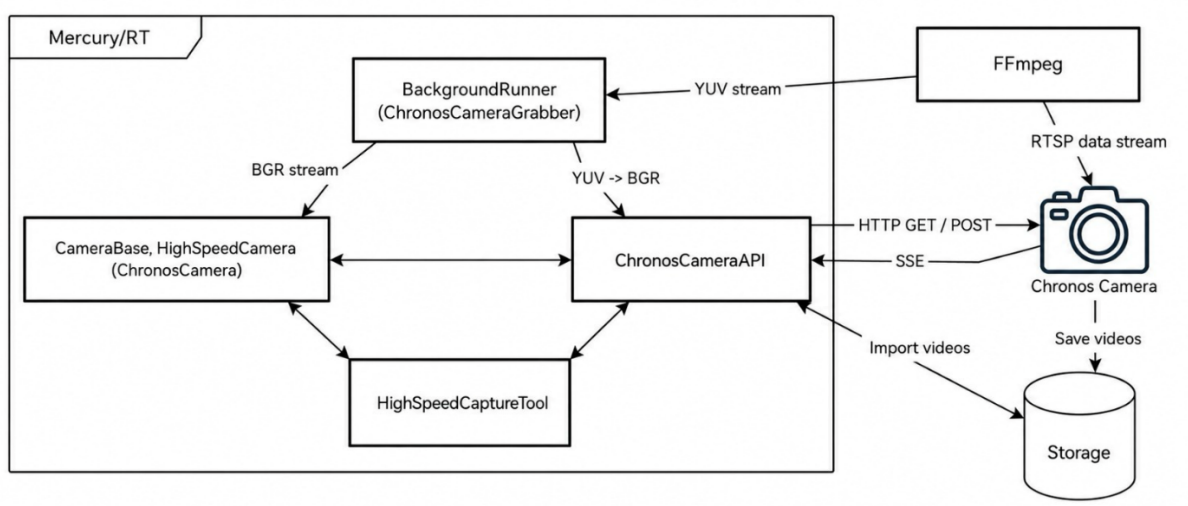


Fig. 2: Data flow between the Chronos camera, the FFmpeg decoder and the MercuryRT modules — RTSP stream for YUV frames and HTTP/SSE for REST API control and video download

3 RESULTS

Live preview latency was measured by two methods: a digital stopwatch displayed on a monitor that the camera filmed, with parallel capture by a reference camera. The time difference was approximately 300 ms at a 10 FPS preview. Attempting to raise the rate to 30 FPS led to latency on the order of seconds and frame skipping caused by FFmpeg decoder saturation — the CPU could not keep up with 30 FPS conversion, as a counter test showed: at 10 FPS the system processed 595 frames out of an expected 600 in 60 s, while at 30 FPS the losses reached tens of percent. The frame rate was therefore fixed at 10 FPS for stability (Kmenta, 2026).

Recording reliability was verified by a trigger → save → import test. The camera was set to 1000 FPS at 640 × 480 resolution with a 5-second pre-trigger; a brief LED flash was generated in the field of view. The recording captured the flash exactly in the expected frame, confirming both the pre-trigger buffer function and correct synchronization. To verify memory stability the preview ran for 4 hours with short recordings every 30 minutes. Application memory usage remained stable with a minimal negligible increase matching a comparative run with a Basler camera — the small increases stem from the MercuryRT core, not from the Chronos module. Tab. 1 summarizes the results.

Tab. 1 Summary of integration module testing results

Metric	Value	Note
Preview latency (10 FPS)	~300 ms	Including decoding and YUV-to-RGB
Throughput (10 FPS)	595/600 frames in 60 s	0.8% loss
Latency at 30 FPS	seconds	Not recommended
4-hour stability	no memory growth	Matches Basler camera
Trigger detection	no frame loss	LED flash test

4 DISCUSSION

Separating REST and RTSP into distinct classes proved valuable in practice — control communication remained stable even during debugging of the data channel and vice versa. This approach is suitable for similar integrations of high-speed cameras that expose both interfaces simultaneously but lack an official C# SDK. The main limitation is decoding performance: even after three rounds of optimization the C# code reliably handles YUV-to-RGB conversion only up to 10 FPS on a typical development machine. For 30 FPS it would be necessary to parallelize the loop across multiple threads, use SIMD vector instructions, or rely on the specialized libyuv library at the cost of a native DLL dependency. For the purposes of a live preview — which serves only for setting up the experiment, not for the recording itself — current performance is sufficient.

5 CONCLUSIONS

The work demonstrated that commercially available Chronos high-speed cameras can be fully integrated into the existing MercuryRT DIC system using standard protocols (REST + RTSP) and an external FFmpeg decoder, without waiting for an official SDK from the manufacturer. The key was an architecture separating the control channel from the data channel and a custom integer-based YUV-to-RGB conversion in C# with an unsafe block. Three results confirmed usability: a stable preview with 300 ms latency, lossless pre-trigger recording verified by the LED test, and memory stability over a four-hour run. Future expansions may include GPU acceleration of the YUV-to-RGB conversion (CUDA, OpenCL) to raise the preview rate towards the native 60 FPS of the RTSP stream, and structured logging of REST calls for post-mortem analysis.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

REFERENCES

BEBERNISS, T. J., EHRHARDT, D. A. 2016. High-speed 3D digital image correlation vibration measurement: Recent advancements and noted limitations. *Mechanical Systems and Signal Processing*. 86, 35-48. ISSN 0888-3270. https://doi.org/10.1016/j.ymssp.2016.04.014

DUKE, D. J., KNAST, T., EDGINGTON-MITCHELL, D. 2018. Low Cost High Speed Digital Cameras for Experimental Fluid Mechanics. In: *Proceedings of the 21st Australasian Fluid Mechanics Conference*. Adelaide, Australia.

DUKE, D. J., KNAST, T., THETHY, B., GISLER, L., EDGINGTON-MITCHELL, D. 2019. A low-cost high-speed CMOS camera for scientific imaging. *Measurement Science and Technology*. 30(7), 075403. ISSN 1361-6501. https://doi.org/10.1088/1361-6501/ab1832.

FIELDING, R. T. 2000. *Architectural Styles and the Design of Network-based Software Architectures*. PhD Dissertation. University of California, Irvine. https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm

KMENTA, M. 2026. *Integration of Chronos high-speed cameras into the MercuryRT system* [in Czech]. Master's thesis, head: Farana, R. Brno: Mendel University in Brno.

KOŘEČEK, D., SOLFRONK, P., SOBOTKA, J. 2022. Analysis of the Dual-phase Steel DP500 Stress-strain Characteristics During the Plane Shear Test. *Manufacturing Technology*. 22(1), 34-38. ISSN 1213-2489. https://doi.org/10.21062/mft.2022.015.

KOTLÁNOVÁ, B., JAVOŘÍK, J., ŽALUDEK, M., OGUNLEYE, R. O. 2023. Volume Change Measurements Using 2D DIC System. *Manufacturing Technology*. 23(3), 307-312. ISSN 1213-2489.https://doi.org/10.21062/mft.2023.044.

RICHARDSON, I. E. G. 2011. *The H.264 Advanced Video Compression Standard*. 2nd ed. Chichester: John Wiley & Sons. ISBN 978-0-470-51692-8.

SCHULZRINNE, H., RAO, A., LANPHIER, R. 1998. *Real Time Streaming Protocol (RTSP)*. RFC 2326. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/rfc2326

SUTTON, M. A., ORTEU, J.-J., SCHREIER, H. W. 2009. *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts, Theory and Applications*. New York: Springer. ISBN 978-0-387-78746-6.

TIPPNER, J., PRAUS, L., BRABEC, M., SEBERA, V., VOJÁČKOVÁ, B., MILCH, J. 2019. Using 3D digital image correlation in identification of defects of trees subjected to bending. *Urban Forestry & Urban Greening*. 46, 126513. ISSN 1618-8667. https://doi.org/10.1016/j.ufug.2019.126513.

Contact information

Martin Kmenta: e-mail: martin-kmenta@seznam.cz

WEB APPLICATION FOR VISUALIZATION AND COMPARISON OF MINIMUM-REDUNDANCY CODES

Ivan Kolesnychenko[1], Radim Farana[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/10.11118/ecos2026_014

ABSTRACT

This article presents the results of the development and testing of a specialized web application designed for teaching and comparing lossless data compression algorithms based on minimum redundancy codes. The work focuses on two fundamental pillars of statistical coding: the Shannon-Fano (Shannon 1948) and Huffman algorithms (Huffman 1952). The main contribution of the created solution is an interactive environment that allows users to monitor calculation steps and the creation of tree structures in real-time on identical inputs, which supports an intuitive understanding of the differences in the efficiency of both methods. The technical solution is built on modern Vue 3 and D3.js technologies. Validation took place in the form of user testing, which proved the high didactic potential of the application and its benefit for students of technical fields.

Keywords: lossless compression, Shannon-Fano coding, Huffman coding, entropy, algorithm visualization, web application
JEL Code: C63

1 INTRODUCTION

Data compression is a fundamental process in the field of modern computer science and telecommunications (Cover and Thomas 2006). With the constantly increasing volume of transmitted data, effective coding methods are becoming key to optimizing storage space and network throughput. Despite their practical ubiquity, the teaching of basic principles of lossless compression at universities is often limited to static theoretical interpretation, which may not sufficiently reflect the dynamics of the algorithms themselves. The motivation for this work was the need to create an interactive and visually attractive tool that would allow students to experiment with input data and observe how changes in the probability of symbol occurrence are reflected in the structure of the resulting code. The presented text describes the design of such an application, its technological implementation, and an evaluation of its contribution to the educational process (Kolesnychenko 2026).



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

2 CURRENT STATE AND ANALYSIS

2.1 Theoretical Foundations

Lossless compression is based on the principle of eliminating redundancy while maintaining information integrity. A key metric in this context is source entropy, which determines the minimum average code length per symbol required for error-free reconstruction (Farana 2014).

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{1}$$

Where p(x) represents the probability of a symbol's occurrence in the source alphabet. Shannon-Fano and Huffman coding then represent prefix codes that attempt to approach this theoretical limit by assigning shorter code words to symbols with higher probability (Salomon 2007). A minimum-redundancy code is better the closer its average codeword length is to the entropy.

2.2 Existing Solutions

During the analysis, a number of available visualization tools were examined. It was found that although general platforms exist (e.g., Algorithm Visualizer), they often lack a specific focus on information theory or do not allow a direct comparison of two methods over the same text. Educational resources such as the Purdue University tutorial provide quality theory but lack interactivity. These findings led to the definition of requirements for a new solution that would combine these aspects (Mishra 2023).

3 APPLICATION DESIGN

The goal of the design was to create an environment that minimizes the user's cognitive load and maximizes the didactic effect. The application was designed as a single-page application (SPA) with an emphasis on clarity and responsiveness.

3.1 Functional Requirements

Lossless compression is based on the principle of eliminating redundancy while maintaining information integrity. A key metric in this context is source entropy, which determines the minimum average code length per symbol required for error-free reconstruction.

- **Input Data:** Option to enter a custom string or select from predefined test sets (e.g., homogeneous distribution vs. extremely unbalanced frequencies).
- **Visualization:** Generation of binary trees using the D3.js library, which dynamically react to the algorithm's stepping.
- **Comparison:** Parallel display of Shannon-Fano and Huffman algorithm progress side by side.
- **Metrics:** Immediate calculation of entropy, average code length, and efficiency for both methods.

4 APPLICATION DESCRIPTION AND IMPLEMENTATION

The application was developed using the modern JavaScript ecosystem. The core consists of the Vue 3 framework using the Composition API, which allows for efficient state management and reactivity. The entire code is written in TypeScript, which significantly facilitated the implementation of complex algorithm logic and reduced the risk of errors when manipulating tree structures.

4.1 Visualization Logic

A key element is the generation of intermediate algorithm steps. Each step is saved as a "snapshot" of the current state of the data structures. For Huffman coding, the phases of joining two nodes with the lowest probability are visualized, while for the Shannon-Fano algorithm, the application displays the process of dividing the symbol list into two parts with the closest possible sum of probabilities, see Fig. 1.

4.2 Graphical Interface

The D3.js library was used for tree rendering, enabling smooth transitions and node animations. Thanks to this, the user "sees" how the nodes group or divide over time. The interface is localized into Czech and English using the vue-i18n module, which expands the potential target group.

5 TESTING RESULTS

User testing was conducted to verify the functionality and benefits of the application. It was attended by 15 respondents from the ranks of students (both IT and non-IT oriented) and the general public. The methodology included solving specific tasks in the application, followed by a structured interview with ratings of individual aspects on a scale of 1–10.
From the results, it is clear that the application was rated very highly in all monitored categories. The most significant success was recorded in the "suitability for teaching" category, where it reached an average value of 9.27.

Tab. 1 Average rating of selected categories

Respondent Group	Understanding of Principles	Suitability for Teaching	Overall Rating
IT students	9.4	9.6	8.8
Non-IT students	9.2	8.0	9.6
Non-students	9.0	8.0	8.6
Total average	8.47	9.27	9.00

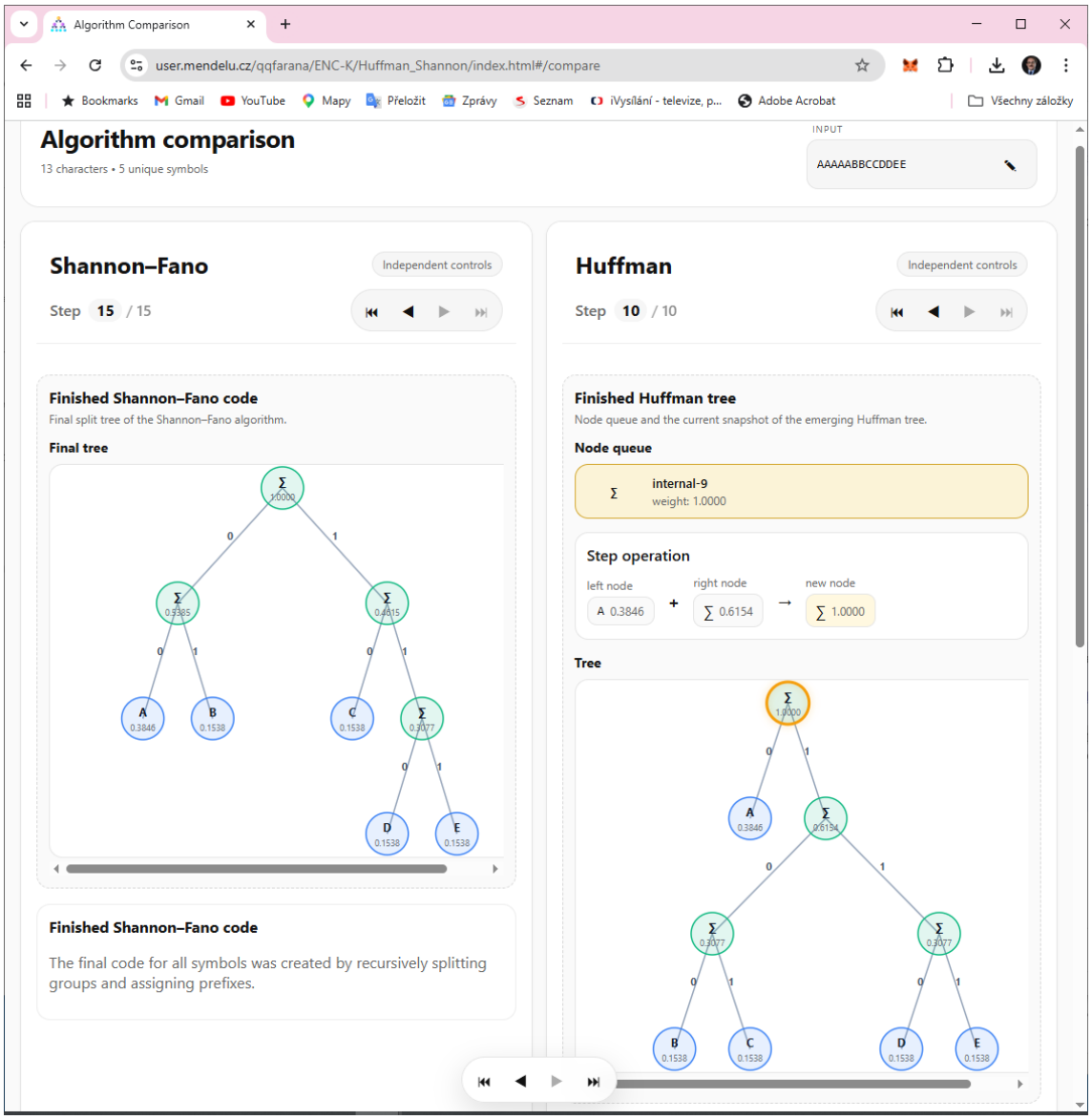


Fig. 1: Algorithm comparison

6 EVALUATION

The feedback analysis confirmed that the visual representation of the algorithm's progress is much more understandable for students than studying pseudocode or static diagrams. Users particularly appreciated the ability to step back and compare the results of both methods on identical data. Higher ability to interpret efficiency results was observed in the IT student group, while the lay public primarily appreciated the aesthetic and technical side of the application. The created solution outperforms analyzed competition mainly in the areas of interactivity and didactic clarity. The ability to step through the algorithm both forwards and backwards, along with comments explaining each step, contributed most to the correct understanding of both algorithms. This allowed the user to go back and forth through the problematic part of the algorithm several times.

By comparing the results of both algorithms, the user will also verify the validity of the empirical finding that the Shannon-Fano algorithm generates codes of worse average length, at most the same as the Huffman code.

7 CONCLUSION

The project successfully demonstrated that modern web technologies can serve as an effective platform for teaching complex algorithmic concepts. The web application for lossless compression visualization represents a functional tool that helps students overcome the barrier of theoretical complexity and gain intuitive insight into the issues of information theory. Future development of the application could move toward integrating other coding methods (e.g., LZW, arithmetic coding) or expanding the theoretical part with interactive quizzes to further consolidate the knowledge gained through visualization. The created tool is ready for deployment in teaching as supplementary material for subjects focused on data processing and information theory.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

REFERENCES

COVER, T. M., THOMAS, J. A. 2006. *Elements of Information Theory*. 2nd ed. Hoboken: Wiley-Interscience. ISBN 978-0-471-24195-9

FARANA, R. 2014. *Data Coding and Compression* [in Czech]. Ostrava: VŠB-TU Ostrava. ISBN 978-80-248-3564-8

HUFFMAN, D. A. 1952. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*. 40(9), 1098–1101, ISSN 0096-8390. https://doi.org/10.1109/JRPROC.1952.273898

KOLESNYCHENKO, I. 2026. *Web Application for Visualization and Comparison of Lossless Compression Algorithms* [in Czech]. Bachelor's thesis, head: Farana, R. Brno: Mendel University in Brno.

MISHRA, A. 2023. Shannon-Fano Algorithm for Data Compression. *Scaler*. [Accessed: 2026-04-12]. https://www.scaler.com/topics/shannon-fano-algorithm-for-data-compression/

SALOMON, D. 2007. *Data Compression: The Complete Reference*. 4th ed. London: Springer. ISBN 978-1-84628-602-5

SHANNON, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal*. 27(3), 379–423, ISSN 0005-8580. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Contact information

Ivan Kolesnychenko: e-mail:ikolesny@seznam.cz

Radim Farana: radim.farana@mendelu.cz

ONLINE PLATFORMS AS TOOLS FOR ENVIRONMENTAL CONSUMER EDUCATION

Lívia Kolibárová[1], Irena Baláková[1]

[1]Department of Marketing and Trade, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/10.11118/ecos2026_015

ABSTRACT

In the digital era, online platforms are becoming a significant channel for environmental consumer education. This paper aims to identify consumers' preferred options for environmental education through online platforms, evaluate their actual utilization, and describe the motivations and barriers to engagement. The empirical section is based on the quantitative questionnaire survey primary data. To statistically verify relations, Spearman's rank correlation coefficient and Pearson's χ^2-test were applied; furthermore, factor and cluster analyses were conducted. The results indicate that 92% of respondents utilize online platforms to obtain environmental information, with social media, and online articles being predominant. A weak positive correlation ($\rho = 0{,}29$) was found between the perceived availability of eco-friendly options in the place of residence and declared pro-environmental behavior, whereas a moderate positive correlation ($\rho = 0{,}53$) was demonstrated between the subjective level of environmental education and pro-environmental behavior. The preferences regarding platform types differ statistically significantly based on the respondents' economic activity. Factor analysis identified three dimensions of attitudes toward online environmental education. The cluster analysis identified three distinct groups of respondents—passive online users, highly active and university-educated learners utilizing a wide spectrum of digital tools and disengaged non-users. The formulated recommendations are primarily directed at educational institutions and non-profit organizations.

Keywords: environmental education, online platforms, social media, sustainability, consumer behavior, digital literacy, cluster analysis, ecological literacy
JEL Code: Q56, I21, M31

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Global environmental issues rank among the most pressing challenges of the contemporary era, and their resolution demands not only systemic measures at the state level but also responsible behavior on the part of individual consumers (Guerra *et al.*, 2024). A key prerequisite for pro-environmental behavior is an adequate level of environmental literacy, which can be developed through education in both formal and informal settings (Kollmuss & Agyeman, 2002). In the digital era, the significance of online platforms is expanding as a tool capable of making such education accessible to broad groups of consumers (Hassan *et al.*, 2024; Liu & Zhang, 2024).

Online platforms offer flexibility, visual appeal, and opportunities for interactive engagement, enabling them to reach target groups that traditional educational paths no longer influence (Xiaomeng, 2024). Social media, video platforms, podcasts, blogs, specialized non-profit websites, and online courses represent diverse formats with varying depths of content and degrees of credibility (Hajdukiewicz & Pera, 2020; Pabian & Pabian, 2023). Alongside these opportunities, the digital environment introduces risks in the form of inaccurate or misleading information, high demands on media literacy, and the phenomenon of greenwashing, which undermines the credibility of sustainability communication (Yang *et al.*, 2020). The role of information sources is further complemented by social media influencers, whose credibility and attractiveness affect consumers' intentions to act in a more environmentally friendly manner (Awang *et al.*, 2021).

Empirical studies confirm that a higher level of environmental awareness is associated with a greater propensity for pro-environmental behavior (Mkumbachi, 2020; Saptura, 2022). However, the effectiveness of education also depends on consumers' perceived capacity to act and their confidence in the impact of their own behavior (Graca & Khare, 2023). Concurrently, barriers preventing engagement have been repeatedly identified: lack of time, low trust in online sources, and the perception that individual behavior does not exert a sufficient impact (Chung *et al.*, 2020).

Despite the extensive literature on the effectiveness of digital environmental education, the question remains open as to how consumers within the Central European context actually choose among types of online platforms, what motivations and barriers guide their decision-making, and whether segments with distinct educational preferences can be identified to be targeted by specific communication strategies. Mapping these patterns is a fundamental prerequisite for effective environmental awareness.

This paper aims to identify consumers' preferred options for environmental education through online platforms and to evaluate their actual utilization. The specific sub-objectives are: (1) to identify consumer motivations for environmental education and their preferences regarding online platforms; (2) to determine the barriers preventing participation in educational activities; and (3) to segment consumers based on their educational preferences.

2 METHODOLOGY AND DATA

To achieve the established objective, a mixed-methods research design was adopted, combining secondary and primary data in the form of a quantitative questionnaire survey. Secondary sources (DataReportal, AMI Digital, documents from international environmental organizations, and academic publications) were utilized to contextualize the study and operationalize the research variables.

The quantitative survey was administered through an online questionnaire using the Microsoft Forms application between February 22 and March 10, 2025. Distribution was carried out via social

media (Facebook, Instagram), email, and through environmentally oriented non-profit organizations and educational institutions. The questionnaire featured a branching structure: respondents who declared at least a partial interest in environmental issues in the opening filter questions completed an extended version consisting of 17 items (5 of which were demographic/identifying items); respondents expressing no interest were redirected to a shortened version containing 5–6 items. Closed-ended questions with scale, single-choice, and multiple-choice options were employed. In total, 276 complete questionnaires were obtained.

The demographic profile of the respondents reflected the nature of online data collection (Table 1): females accounted for 66.66%, the 18–25 age group represented 60.87%, university-educated respondents constituted 53.26%, students made up 55.79%, and residents of major cities with over 100,000 inhabitants comprised 51.81%. Consequently, the sample primarily captures a young, digitally literate population, a factor that must be taken into account when interpreting the results.

Tab. 1 Demographic profile of the questionnaire survey respondents (source: own questionnaire, n = 276)

	Absolute frequency	Relative frequency
Gender		
Female	184	66.66%
Male	89	32.25%
Other	3	1.09%
Age		
18–25	168	60.87%
26–35	48	17.39%
36–45	18	6.52%
46–55	23	8.33%
56–65	14	5.07%
66+	5	1.81%
Place of residence		
100 000 and more inhabitants	143	51.81%
50 000-100 000 inhabitants	20	7.25%
10 000-50 000 inhabitants	25	9.06%
5 000-10 000 inhabitants	17	6.16%
1 000-5 000 inhabitants	53	19.20%
Up to 1 000 inhabitants	18	6.52%
Highest level of education		
Primary education	1	0.72%
Secondary education	128	46.38%
Higher education / University	147	53.26%
Economic status		
Employee	82	29.71%
Self-employed	21	7.61%
Unemployed	2	0.72%
Student	154	55.79%
Retired	8	2.90%
Maternity / Parental leave	5	1.81%
Other	4	1.45%

To test relationships, Spearman's rank correlation coefficient was employed (for ordinal variables) and Pearson's χ^2-test of independence was utilized (for categorical data); the significance level was set at $\alpha = 0{,}05$. Prior to the analysis, variable coding was performed, with higher values representing a stronger pro-environmental attitude. Variables regarding the utilization of individual online platforms were converted into binary form.

Two hypotheses were tested. H1 examined the relationship between the perceived availability of eco-friendly options in the place of residence and the respondents' declared pro-environmental behavior. H2 tested the relationship between economic activity (student / employee / other) and the utilization of specific types of online platforms for environmental education. Additionally, the correlation between the subjectively perceived level of environmental education and the degree of pro-environmental behavior was evaluated.

To explore motivational and barrier factors more deeply, an exploratory factor analysis was conducted using Kaiser's criterion (factors with an eigenvalue greater than 1 were included, and variables with a factor loading greater than 0.7 were interpreted). For respondent segmentation, a k-means cluster analysis was applied prior to the k-means cluster analysis, the optimal number of clusters was assessed using the elbow method. Based on the identified elbow point in the within-cluster sum of squares curve, a three-cluster solution was selected. The variables included in the analysis were pro-environmental behavior, level of environmental education, use of online platforms, and the type of most frequently utilized platform. Data processing was performed using Statistica and Microsoft Excel software.

3 RESULTS

Online platforms represent the predominant information channel for the respondents, with 92% declaring that they utilize them to obtain environmental information. In a direct comparison of the effectiveness of online platforms versus traditional media, respondents rated online platforms as more effective (mean score of 1.94 on a five-point scale, where 1 = online platforms, 5 = traditional

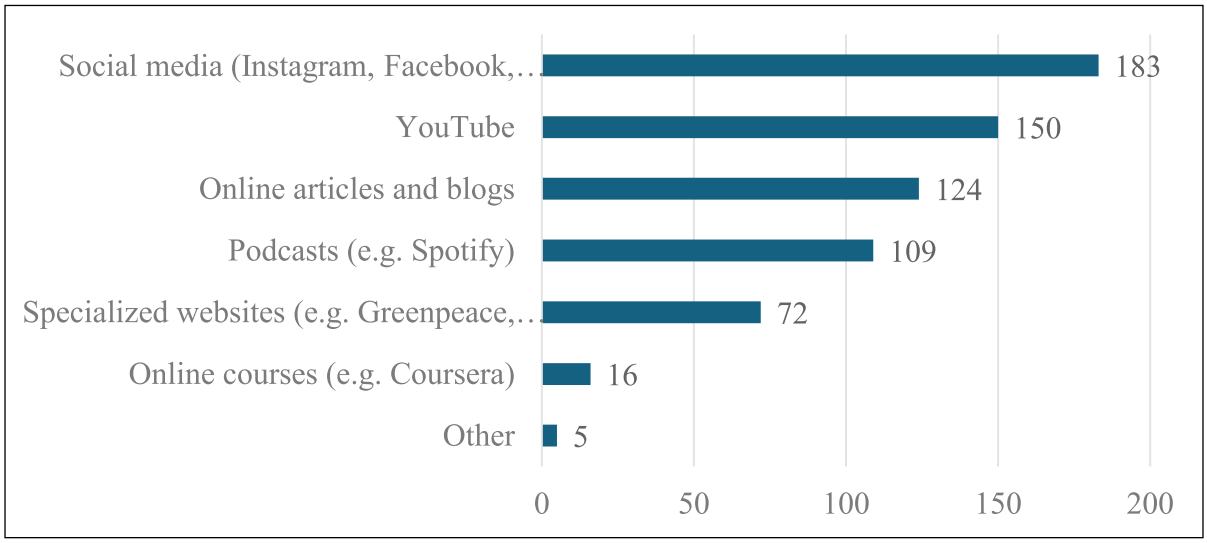


Fig. 1: Most frequently utilized types of online platforms for environmental education (source: own questionnaire, n = 276)

media). Respondents primarily turned to social media, YouTube, and online articles for environmental information (indicated via a multiple-choice question – Figure 1). Conversely, online courses and specialized websites received the lowest engagement.

The subsequent question focused on the credibility of information sources. Specifically, respondents indicated the highest level of trust in scientists and experts, who received 227 choices, followed by non-profit organizations with 153 choices. Conversely, sources such as news media, eco-influencers, and family or friends achieved substantially lower recognition, while outright skepticism toward online environmental information remained entirely marginal.

To verify H1, Spearman's rank correlation coefficient was applied between the perceived availability of eco-friendly options in the place of residence and the degree of pro-environmental behavior. The result ($\rho = 0.29$) indicates a weak, statistically significant positive correlation, on the basis of which the null hypothesis was rejected. Respondents who perceive a higher availability of eco-friendly options thus also declare a higher degree of pro-environmental behavior. An additional analysis revealed a stronger relationship between the subjectively perceived level of environmental education and pro-environmental behavior ($\rho = 0.53$), aligning with the assumption that education serves as an important predictor of pro-environmental behavior.

Hypothesis H2 was tested using a χ^2-test of independence for each type of online platform individually (Table 2). A statistically significant relationship between the respondents' economic activity and platform utilization was demonstrated for social media ($p = 0.0014$), YouTube ($p = 0.0073$), and online courses ($p = 0.031$). Regarding social media and YouTube, students exhibit higher utilization rates, whereas for online courses, the opposite is true, with employees showing higher engagement. For online articles and blogs, podcasts, and specialized websites, no statistically significant relationship was proven, indicating their cross-cutting acceptance across different economic groups.

The factor analysis identified three factors with an eigenvalue greater than 1, which collectively explain 63.0% of the total data variance. Factor 1 (Intrinsic Motivation; 31.4% of variance) loads heavily on a desire to contribute to environmental protection (0.837), interest in the topic (0.830), and social responsibility (0.769), while showing a strong negative correlation with an apathetic attitude (−0.563). Factor 2 (Information Barriers; 18.4% of variance) highlights a clear connection between digital literacy and trust. It links a lack of knowledge on where to find quality resources (0.749) with a general distrust of online information (0.796). Factor 3 (Intrinsic Passivity; 13.0% variance) captures internal behavioral barriers, dominated by laziness or a lack of energy (0.865) and further constrained by time limitations (0.530).

Among the strongest motivations for online environmental education are the perception of social responsibility (mean score of 4.09 on a five-point scale) and the effort to know how to personally

Tab. 2 Summary of hypothesis testing regarding the relationship between economic activity and the utilization of individual online platform types (source: own questionnaire processed in Statistica software)

Online platform	p-value	Decision	Conclusion
Social networks	0.0014	H0 rejected	Statistically significant
YouTube	0.0073	H0 rejected	Statistically significant
Online articles and blogs	0.558	H0 not rejected	Statistically non-significant
Podcasts	0.408	H0 not rejected	Statistically non-significant
Specialized websites	0.161	H0 not rejected	Statistically non-significant
Online courses	0.031	H0 rejected	Statistically significant

Tab. 3 Consumer segments (source: own questionnaire processed in Statistica software)

Questions from the questionnaire	Segment 1	Segment 2	Segment 3
Do you do something for environmental protection in your daily life?	2.909	3.919	2.786
Do you consider yourself to be a sufficiently environmentally educated person?	3.082	3.661	2.405
Do you use online platforms to obtain information?	1.000	0.992	0.024
What online platforms do you use for education in the field of ecology?			
Social media	0.846	0.726	0.000
Online articles and blogs	0.473	0.581	0.000
YouTube	0.627	0.653	0.000
Podcasts	0.400	0.524	0.000
Specialized websites	0.118	0.476	0.000
Online courses	0.045	0.089	0.000
ewobox	0.000	0.016	0.000
EPALE	0.009	0.000	0.000
Newsletter	0.000	0.008	0.000

contribute to environmental protection (3.64). The most significant barriers identified are a lack of time (3.15) and a lack of intrinsic motivation or energy (3.05). Consequently, consumers generally recognize the need for education but encounter organizational and motivational obstacles.

The k-means cluster analysis segmented the respondents into three distinct profiles based on their behaviors and platform preferences (Table 3). Segment 1 (n = 110) is characterized by a passive approach to environmental topics. Although all members actively utilize online platforms, they exhibit only moderate pro-environmental behavior and self-perceived education. Social media constitutes their dominant information channel. Demographically, this group is predominantly composed of students and urban residents aged 18–25. Segment 2 (n = 124) represents the largest and most proactive segment, where respondents display the highest rates of both pro-environmental behavior and self-perceived education. Unlike the other groups, they utilize a wide spectrum of digital platforms, frequently engaging with specialized websites. This segment includes both students and employed professionals, with university education being predominant. Segment 3 (n = 42) clusters respondents with the lowest interest in ecology. They manifest almost no utilization of online platforms for environmental information and exhibit the lowest rates of pro-environmental behavior and self-assessed knowledge. Within this segment, secondary education with a school-leaving exam prevails.

Insights from the minority segment of unengaged respondents (21 individuals who declared no interest in environmental issues) revealed that the main reasons provided were a general disinterest in the topic and a lack of time; according to them, the most effective motivational stimulus would be clear, practical benefits of environmental education for everyday life.

4 DISCUSSION AND CONCLUSIONS

The presented study confirmed that online platforms serve as the predominant information channel for consumers to obtain environmental information; the declared utilization rate of 92% aligns with the position of Slovakia as a highly digitalized country with a 91.4% internet penetration rate (DataReportal, 2024). These findings are consistent with the literature, which asserts that digital tools represent an effective pathway for raising environmental awareness across various age groups (Hassan *et al.*, 2024; Liu & Zhang, 2024; Xiaomeng, 2024).

The moderate positive correlation between subjectively perceived environmental education and pro-environmental behavior ($\rho = 0{,}53$) supports the assumption that knowledge translates into daily decisions. Similar findings were reported by Mkumbachi (2020) among university students and Saptura (2022) within a broader consumer context. The weaker, yet statistically significant correlation between the availability of eco-friendly options in the place of residence and behavior ($\rho = 0{,}29$) suggests that knowledge and attitude are stronger predictors than infrastructural accessibility; therefore, infrastructure improvement and educational activities should complement each other.

A key practical finding is that the choice of an online platform is related to economic activity. Students prefer visually interactive social media and YouTube, whereas employees seek more professionally oriented online courses. This conclusion is aligned with the findings of Chung *et al.* (2020), according to whom students utilize social media as their primary information channel. For educational content creators, this implies a necessity to segment both the format and the channel according to the target audience: short videos and social media campaigns for younger audiences, and courses and professional texts for working professionals.

The factor analysis demonstrated that the effectiveness of online environmental education depends not only on content availability but also on individual values, the digital literacy required to identify high-quality information, and intrinsic initiative. This three-dimensional perspective corresponds to the theoretical model of motivations and barriers to pro-environmental behavior proposed by Kollmuss and Agyeman (2002) and the empirical results of Graca and Khare (2023), which suggest that education operates through the consumer's perceived self-efficacy.

Segmentation via cluster analysis uncovered three distinctly different target groups whose needs require differentiated communication approaches. Passive social media users require concise, visually appealing content with an emphasis on practical applicability. Active, educated consumers represent not only recipients but also potential multipliers suitable for more advanced professional content and community activities. Respondents with low interest require a completely distinct strategy: engagement through emotionally powerful content highlighting specific personal benefits, or a combination of online and offline formats.

Two priority recommendations for practice emerge from the results. For educational institutions, it is advisable to integrate environmental content in the form of short interactive videos, online quizzes, and webinars with experts, and to conduct communication through channels that students utilize (social media, YouTube). Non-profit organizations should more intensively occupy formats with the highest perceived credibility among respondents—given that NGOs ranked second in credibility after scientists and experts—while simultaneously making professional content accessible through podcasts and short visual videos, i.e., formats with cross-cutting reach.

The study is subject to several limitations. The primary limitation concerns sample representativeness: the dominance of students, younger age groups, and residents of major cities reflects the distribution method of the questionnaire via online channels, leading to an underrepresentation of older and less digitally literate individuals. Furthermore, the subjective evaluation of one's own

environmental education may not accurately correspond to objectively measurable knowledge levels. Future research should focus on capturing older age groups, utilizing gamification and virtual reality in environmental education, and conducting longitudinal tracking of transitions between the identified segments.

In conclusion, online platforms possess high potential to support environmental literacy and pro-environmental consumer behavior; however, their effectiveness is conditional upon a sensitive choice of channel and format regarding the characteristics of the target segment, content credibility, and the enhancement of consumers' digital literacy.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, co-financed by the Operational Programme Research, Development and Education.

REFERENCES

AWANG, H., *et al*. 2021. The effect of social media influencer attributes towards pro-environmental intention. *ResearchGate*. https://www.researchgate.net/publication/350564491

CHUNG, N., JEONG, C., KIM, S. 2020. Applying social media to environmental education: Is it more impactful than traditional media? *Information Development*. 48(4), 255–266. https://doi.org/10.1108/IDD-04-2020-0047

DATAREPORTAL. 2024. Digital 2024 Global Overview Report. *Datareportal*. https://datareportal.com/reports/digital-2024-global-overview-report

GRACA, S. S., KHARE, V. 2023. Educating Global Green Consumers: The Role of Online Education and Brand Communication in Promoting Green Buying Behavior. *ResearchGate*. https://www.researchgate.net/publication/373760801

GUERRA, K. L. *et al*. 2024. Importance of Environmental Education in the Context of Natural Sustainability. *ResearchGate*. https://www.researchgate.net/publication/381014984

HAJDUKIEWICZ, A., PERA, B. 2020. Education for Sustainable Development—The Case of Massive Open Online Courses. *Sustainability*. 12(20), 8542. https://doi.org/10.3390/su12208542

HASSAN, M. H. *et al*. 2024. Environmental Education: A Systematic Review on the Use of Digital Tools for Fostering Sustainability Awareness. *Sustainability*. 16(9), 3733. https://doi.org/10.3390/su16093733

KOLLMUSS, A., AGYEMAN, J. 2002. Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*. 8(3), 239–260. https://doi.org/10.1080/13504620220145401

LIU, J., ZHANG, X. 2024. Enhancing Environmental Awareness through Digital Tools in Environmental Education in China. *ResearchGate*. https://www.researchgate.net/publication/379323194

MKUMBACHI, R. 2020. Environmental awareness and pro-environmental behavior: A case of university students in Malang city. *ResearchGate*. https://www.researchgate.net/publication/342698852

PABIAN, A., PABIAN, B. 2023. Role of Social Media in Managing Knowledge of the Young Generation in the Sustainability Area. *Sustainability*. 15(7), 6008. https://doi.org/10.3390/su15076008

SAPTURA, N. 2022. The correlation of environmental education, environmental knowledge, environmental involvement and waste management behavior. *ResearchGate*. https://www.researchgate.net/publication/366381403

XIAOMENG, X. 2024. Enhancing Environmental Education through Digital Tools. *International Journal of New Developments in Education*. 6(9). https://doi.org/10.25236/IJNDE.2024.060932

YANG, Z. *et al*. 2020. Greenwashing behaviours: Causes, taxonomy and consequences based on a systematic literature review. *ResearchGate*. https://www.researchgate.net/publication/345844645

Contact information

Irena Baláková: e-mail: irena.balakova@mendelu.cz

CZECHUP: A WEB SYSTEM FOR PREPARING FOREIGNERS FOR THE CZECH LANGUAGE EXAM WITH ADAPTIVE VOCABULARY AND AI-GENERATED EXERCISES

Polina Komárková[1]

[1]Department of Economics, Faculty of Business and Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/10.11118/ecos2026_016

ABSTRACT

This article deals with the design and implementation of the CzechUP web system that helps foreigners prepare for the mandatory Czech language exam required for permanent residence (level A2) or citizenship (level B1). The work responds to the lack of suitable tools for Russian- and Ukrainian-speaking users, for whom the most popular application Duolingo is available only via English. The system integrates a database of 4,372 words parsed from three official CEFR textbooks, the Majka morphological analyzer (up to 347 forms for a single adjective), exercise generation via the Gemini API, and translation via Seznam Slovník and DeepL. The backend is built in ASP.NET Core over PostgreSQL, the frontend in Next.js with localization into three languages. User testing with five respondents confirmed the usefulness of the grammatical-form filter as a unique feature and estimated the error rate of AI-generated exercises at approximately 10%.

Keywords: Czech language teaching, Czech for foreigners, CEFR, web application, morphological analysis, Gemini API
JEL Code: I21

1 INTRODUCTION

A foreigner seeking permanent residence in the Czech Republic must pass a Czech language exam at level A2; citizenship requires level B1 together with a test of civic knowledge (Národní pedagogický institut 2025). The exam lasts over three hours with a 60% pass threshold in each section — reading, writing, listening, speaking. According to the Czech Statistical Office (2025), the number of foreigners in the country grew by more than 50% after 2022, with Russian- and Ukrainian-speaking residents making up the dominant share. For them, a paradox opens up: the most widespread learning application Duolingo teaches Czech only via English. Users without strong English face a double language barrier before they can start studying. A survey of 78 respondents

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

from the Brno community (Ukraine 38.5%, Russia 35.9%, Kazakhstan 15.4%) showed that 91% plan to obtain permanent residence or citizenship, but 37% hold no Czech language certificate. The aim of this work was to design and build a functional system that removes three repeatedly mentioned deficiencies of existing tools: the lack of a multilingual interface, missing vocabulary adaptability, and limited coverage of grammatical forms.

1.1 Language Level and Existing Tools

The Common European Framework of Reference for Languages (CEFR 2002) defines six levels from A1 to C2. For Czech, official descriptions exist for levels A1, A2, and B2 (Línek and Vlasáková 2005; Bidlas *et al.* 2005; Adamovičová *et al.* 2005), updated by Cvejnová (2016) and complemented by a description of A1–B2 (Nováková *et al.* 2022). Above B2 the available literature drops sharply. The classical printed source remains the textbook by Holá (2016) for levels A1 to B1. Duolingo is the most used Czech self-study application in the surveyed sample, but users repeatedly point out its limitations: Czech only via English, often unnatural sentences, and missing systematic grammar. Quizlet and Anki provide flashcards but no structured CEFR exercises. Molendová (2024) reached an analogous conclusion for English taught with digital tools: what matters is not the digitization itself, but how well the tool reflects the individual progress of the student.

1.2 Czech Morphology and Content Generation

Czech has rich inflection — a single adjective can have more than 300 forms. The Internet Language Reference of the Czech Language Institute offers no public API and its coverage of adjective forms is incomplete. The Majka morphological analyzer (Šmerk 2007), developed at the Faculty of Informatics of Masaryk University, covers approximately 90% of the CZES corpus and generates up to 347 forms for the word “hezký” with tags in a compact format (K1gFnSc7 = noun, feminine, singular, instrumental). The exercise typology of the system is based on the work of Wu *et al.* (2007), who described seven exercise types for a digital library of language learning exercises.

2 SYSTEM ARCHITECTURE

From the foreigner survey and the analysis of existing applications, five groups of functional requirements were derived: account management (registration, JWT authentication), word learning (filtering by CEFR, topic, and tags, manual additions), grammar rules with user notes, exercises with automatic evaluation, and offline study via PDF export.

The backend is implemented in ASP.NET Core over PostgreSQL with Entity Framework Core (Code First) following the four-layer Clean Architecture model (Martin 2018). Three projects separate the entity layer (CzechUp.EF with 19 entities), the service layer (CzechUp.Services with DTOs and business logic), and the controller layer (CzechUp.WebApi). The frontend is built in Next.js with the App Router and TypeScript; localization is handled by next-intl for three languages (Czech, Russian, English), indexed in URL prefixes `/cs/, /ru/, /en/`. Styling uses Tailwind CSS, client-side PDF export uses jsPDF.

The database contains 25 tables and 19 entities in third normal form. The tables are split into two groups: `General*` for static data (official vocabulary, default topics, grammar rules) and `User*` for dynamic data of a specific user (personal dictionary, exercise results, notes). For a newly

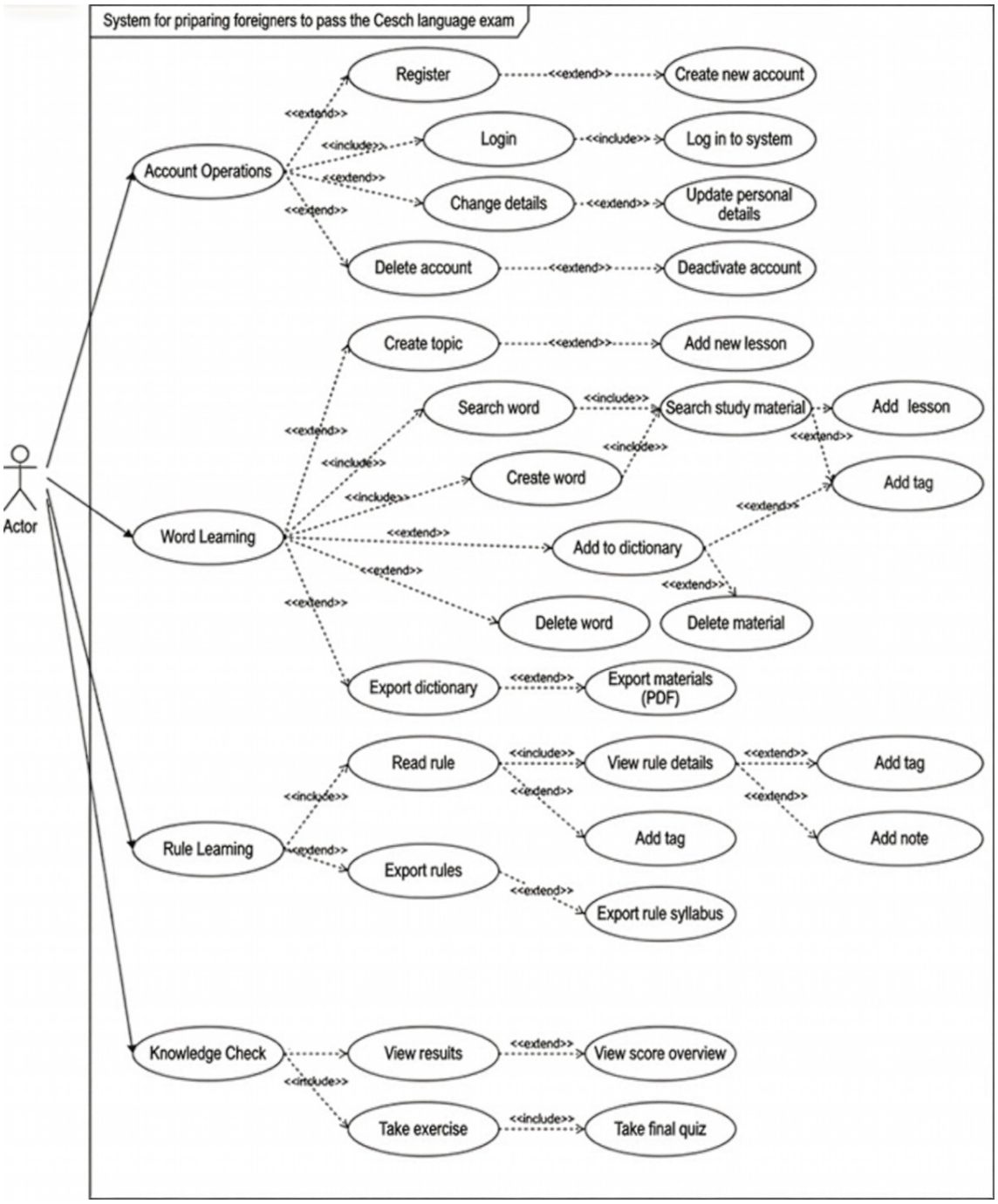


Fig. 1: USE CASE diagram of the CzechUP system (adopted from Tselishcheva 2025).

registered user, data from `General*` tables are copied into `User*` tables, so the user can edit the default content without affecting others. This split requires more storage but dramatically simplifies user data isolation.

For authentication, the system uses JSON Web Tokens with a 1-hour validity, and passwords are hashed with PBKDF2 with a random salt. The [Authorize] attribute protects most endpoints; data isolation works by every CRUD endpoint operating only on records whose UserGuid matches the identifier in the token.

3 IMPLEMENTATION AND RESULTS

To populate the default database, a console application WordHandler was written that parses three different formats of the official CEFR Czech reference textbooks. The merging into the shared GeneralOriginalWords table is summarized in Tab. 1.

Tab. 1 Number of words from three CEFR reference sources after merging

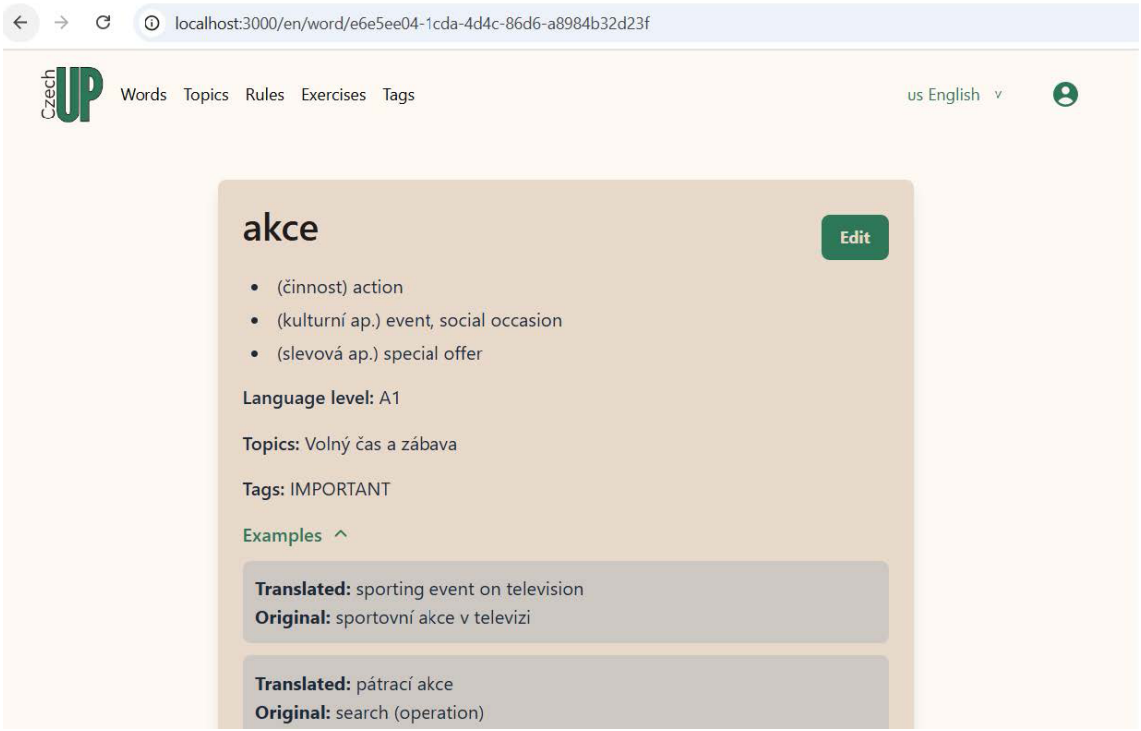
Level	Source 1	Source 2	Source 3	Total
A1	1094	123	156	1373
A2	1213	88	110	1411
B1	1128	0	0	1128
B2	353	0	107	460
Σ	**3788**	**211**	**373**	**4372**

For automatic translation, two sources were combined: the Seznam Slovník API as the primary source for individual words (Czech plus 10 languages) and the DeepL API as a fallback for missing words and for whole sentences. Testing showed that Seznam is more accurate for individual words while DeepL is stronger on sentences; for single words DeepL has approximately 10% inaccuracy — typically picking one of several possible translations without context. For multi-word expressions, Majka supplies the morphological forms; five typical structures were identified (verb + noun, adjective + noun, etc.) with rules for which words to inflect.

Exercise generation uses the Gemini API from Google (2025), available in a free tier sufficient for occasional generation. Three exercises were planned per word per level — more than 13,000 exercises in total, which manual authoring could not produce. Five exercise types were implemented: completing a word form in a sentence, sentence construction from words, completing key words in a text (all via AI), choosing the correct translation from 4–8 options, and writing the Czech translation (without AI, drawn from the database). Finding a prompt that returns consistent JSON required more than 30 iterations — the models occasionally changed the output structure, added comments outside the JSON, or used different field names.

The Word Forms section is a unique feature of the application. For nouns the filter offers two criteria (number, case), for adjectives five (case, number, negation, gender, degree of comparison), and for verbs six. For “přijatelný” in dative, plural, second degree, the application returns “nepřijatelnějším / přijatelnějším”. For “kočka” Majka generates 13 forms, for “hezký” up to 347.

The application was tested by five users: three Russian-speaking foreigners at levels A1, B2, and C1 (Russian localization), one English-speaking foreigner with no Czech, and one native Czech speaker with C1 English (who evaluated translation quality). Positive feedback highlighted the breadth of functionality, translation quality, and especially the grammatical-form filter, which is missing from other tools and serves as a key aid for advanced foreigners when writing in Czech. All respondents stated they would use the application after release. Negative feedback comprised three points: three of five users noted that the UI does not feel modern enough and requires a graphic redesign; two suggested an instructional video, since some functions are not immediately obvious; and the native speaker estimated that approximately 1 in 10 generated exercises contains a minor error (Tselishcheva, 2025).



Tab. 2 Detail of the word "akce" with translations, examples, and the Word Forms section (adopted from Tselishcheva 2025)

For regression stability, three groups of xUnit tests were written: external API tests (URL validity, data structure, translation quality), endpoint tests with mock objects via the Moq library, and integration tests with an in-memory database via `Microsoft.EntityFrameworkCore.InMemory`.

4 DISCUSSION

The feature that users named the most valuable — the grammatical-form filter — emerged from a technical decision that originally had a different purpose. When it became clear that the Internet Language Reference offered no API while Majka did, the complete list of forms became a by-product. This is a general lesson: some of the most valuable features are not designed during planning, but emerge from solving technical problems with other goals. For AI-generated content the second lesson holds: the ~10% error rate means that AI alone is not enough as a final source of content for language teaching — it requires either human review (expensive), a second LLM as a validator, or a user error-reporting mechanism. The main limitations of the system remain the missing audio for pronunciation and listening, suboptimal UI design, and the status of external APIs, which complicates public deployment without commercial agreements.

5 CONCLUSIONS

The work demonstrated that commonly available APIs and open-source tools can be used to build a prototype language learning application that addresses a concrete pain point of foreigners in the Czech Republic — the Czech language exam for permanent residence and citizenship. Multi-source integration works, but at the cost of more complex tooling — merging three official CEFR textbooks (4,372 words) required three different parsers, and the combination of Seznam Slovník + DeepL + Majka was necessary because no single API handles both Czech translation and morphology. AI-generated content for language teaching needs a validation layer — the Gemini API accelerated exercise preparation, but ~10% of outputs contain errors a student would treat as authoritative. A feature that does not seem important during planning (a filter over 347 adjective forms) can be what users see as the deciding factor. Future expansions may include adding audio recordings via TTS, a validation layer for AI-generated exercises, an administrator interface for bulk corrections, and extending the localization to Ukrainian and Vietnamese.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

REFERENCES

Adamovičová, A. *et al*. (2005). *Czech as a foreign language. Level B2 according to the Common European Framework of Reference for Languages [in Czech]*. Praha: Tauris.

Bidlas, V., Confortiová, H., Turzíková, M. (2005). *Czech as a foreign language. Level A2 according to the Common European Framework of Reference for Languages* [in Czech]. Praha: Tauris.

CEFR. (2002). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Czech edition)*. Olomouc: Univerzita Palackého. ISBN 80-244-0404-4.

Cvejnová, J. (2016). *Reference Description of Czech for the Language Exam for Permanent Residence in the Czech Republic — Levels A1, A2* [in Czech]. Praha: Národní ústav pro vzdělávání. ISBN 978-80-7481-164-7.

Czech Statistical Office. (2025). Number of foreigners, demographic events [in Czech]. *CSÚ*. https://csu.gov.cz/pocet-cizincu-demograficke-udalosti

Google. (2025). *Google AI for Developers (Gemini API)*. https://ai.google.dev/

Holá, L. (2016). *Česky krok za krokem 1 = Czech step by step 1: A1–A2*. 1st ed. Praha: Akropolis. ISBN 978-80-7470-129-0.

Línek, J., Vlasáková, K. (2005). *Czech as a foreign language. Level A1 according to the Common European Framework of Reference for Languages* [in Czech]. Olomouc: Univerzita Palackého.

MARTIN, R. C. (2018). *Clean Architecture: A Craftsman's Guide to Software Structure and Design*. Boston: Prentice Hall. ISBN 978-0-13-449416-6.

Molendová, K. (2024). *Vocabulary Teaching in English with Digital Technologies* [in Czech]. Master's thesis. Plzeň: University of West Bohemia.

National Pedagogical Institute of The Czech Republic (2025). Czech language exam for permanent residence and citizenship [in Czech]. *npi*. https://cestina-pro-cizince.cz/

Nováková, R. *et al*. (2022). *Descriptions of Reference Levels A1–B2 for Czech Sign Language* [in Czech]. Praha: Národní pedagogický institut České republiky. ISBN 978-80-7578-102-4.

Šmerk, P. (2007). Fast Morphological Analysis of Czech. In: Sojka, P., Horák, A. (eds.) *Proceedings of Third Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2009*. Brno: Masaryk University, pp. 13–16. ISBN 978-80-210-5048-8.

Tselishcheva, P. (2025). CzechUP: a web system for preparing foreigners for the Czech language exam [in Czech]. Brno: Mendel University in Brno. Master's thesis, head: FARANA, R.

Wu, S., Witten, I. H., Edwards, A., Nichols, D., Aquino, R. (2007). A digital library of language learning exercises. *International Journal of Emerging Technologies in Learning*. 2(1).

Contact information

Polina Komárková: e-mail: polia.tselisheva@mail.ru

GENERATION Z AND AI-DRIVEN SHOPPING: COMPARING ONLINE AND IN-STORE CONSUMER EXPERIENCES

Benjamin Křištof[1], Jana Dannhoferová[2]

[1]Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

[2]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_017

ABSTRACT

This paper examines the methods and intensity of Artificial Intelligence (AI) tool utilization by Generation Z within online and offline shopping environments. The empirical investigation employs a mixed-methods research design, combining a quantitative questionnaire survey ($n = 114$) and six semi-structured interviews conducted in April and May 2026 among respondents aged 18–27 in the Czech Republic. The objective was to identify dominant AI tools, key motivators, and barriers to their adoption, and to evaluate consumers' willingness to share personal data in both environments. Online interaction with AI is more conscious, focusing primarily on generative assistants (ChatGPT), visual search, and recommendation systems. In stores, self-checkout counters and interactive kiosks dominate, though their utilization is often unreflected by consumers. Statistical testing of five hypotheses revealed a higher willingness to share personal data in the online environment ($p = 0.018$). Surprisingly, however, it also demonstrated statistically significantly lower privacy concerns in the offline environment compared to online ($p < 0.001$), which contradicts the assumption derived from the literature. No gender differences in the frequency of AI usage were detected. Based on these findings, a "phygital" strategy was formulated, linking a unified customer data identity, the smartphone as a mediator between environments, and the principles of Explainable AI (XAI).

Keywords: Generation Z; artificial intelligence (AI); online shopping; in-store shopping; phygital retail; mixed-methods research; consumer behavior; privacy paradox
JEL Code: M31, O33, L81

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Artificial intelligence is progressively becoming an integral part of everyday consumer shopping behavior. Recommendation algorithms, visual search, conversational agents, and autonomous checkouts are reshaping how individuals shop in both digital and physical environments (Kotler *et al.*, 2021; Knihová, 2024). For Generation Z, having grown up within a fully interconnected digital ecosystem, these tools represent a natural baseline of daily life. However, an open question remains as to whether their utilization is truly conscious, intentional, and comparable across both shopping environments.

Generation Z, defined in this study by the birth years 1999–2008, is characterized in the literature as the first cohort fully socialized within the digital environment and a bibliometric analysis of current research on Gen Z consumer behavior (Halibas *et al.*, 2025) confirms a rapid increase in academic interest in this cohort. Their shopping behavior combines a high level of information literacy, price-conscious ethical consumption, and a demand for authenticity and immediate access to information (Grigoreva *et al.*, 2021; Djafarova & Foots, 2022). Regarding technological innovations, a strong orientation toward functional utility is evident: AI is adopted when it demonstrably saves time and reduces information overload, rather than for its novelty alone.

Extant literature dedicated to AI adoption in the shopping process exhibits several substantial gaps. Studies typically focus either on e-commerce or, in isolation, on specific technologies within in-store retail, whereas a concurrent comparison of both environments through the lens of a single generational cohort remains scarce. The Technology Acceptance Model (TAM) framework, as proposed by Pavlou (2003), and its extension in the UTAUT2 model (Venkatesh *et al.*, 2012), provide robust theoretical foundations; however, their simultaneous application to online and offline contexts within a single study has not yet been systematically conducted in the Czech environment.

A second significant gap is the insufficient grounding of available studies within the Czech and Central European reality. Foreign research originates predominantly from Western European or Asian contexts, and its transferability to the Czech Generation Z population has not been adequately verified. Concurrently, available data for the Czech market present a paradoxical picture: the Czech Generation Z combines high adoption rates of AI tools with a strong preference for in-store shopping, where they place an unexpectedly high emphasis on experience, thereby creating specific demands on in-store marketing (Klánová, 2025).

The third gap lies in the inadequate empirical verification of the so-called "privacy paradox"—the discrepancy between declared privacy concerns and the actual willingness to share personal data in exchange for personalization (McKee *et al.*, 2024). The literature assumes that Generation Z is more tolerant of online tracking, while the physical presence of technology in in-store shopping may evoke a stronger sense of surveillance. This assumption has not yet been empirically tested in the Czech context.

The primary objective of this paper is to compare the methods and intensity of AI tool utilization by Generation Z during shopping in online environments and in-stores, to identify the motivators and barriers to this utilization—including the willingness to share personal data for personalization purposes—and to formulate recommendations for integrating digital technologies with physical retail spaces. The objective is operationalized into four research questions:

- ***(RQ1)*** *Which specific AI tools does Generation Z actively utilize?*
- ***(RQ2)*** *What are the main motivators and barriers to AI adoption in both environments?*
- ***(RQ3)*** *How frequently does Generation Z utilize AI tools during shopping, and do gender differences exist?*

- ***(RQ4)*** *To what extent is Generation Z willing to share personal data in exchange for a personalized shopping experience, and what are the differences between the environments?*

The paper is structured as follows: the second section describes the mixed-methods methodology and the statistical testing procedure; the third section presents the main empirical findings, including the results of testing five hypotheses; the concluding section contextualizes the findings within the broader literature, formulates "phygital" recommendations for retail practice, and outlines the limitations of the study.

2 METHODOLOGY AND DATA

The study was conducted using a mixed-methods design that combines a quantitative questionnaire survey with qualitative data collection via semi-structured interviews. The choice of a mixed--methods approach was necessitated by the nature of the formulated hypotheses—which require statistical testing of differences between groups and environments—and the simultaneous need for an in-depth understanding of the mechanisms underlying the statistical data. The target population consists of members of Generation Z residing in the Czech Republic; the definition is based on the birth years 1999–2008, corresponding to an age range of 18–27 years at the time of data collection.

The quantitative survey was administered via a structured online questionnaire using Google Forms. The instrument comprised 21 questions divided into five thematic blocks: sociodemographic data, AI utilization in the online environment, AI experiences in in-store shopping, motivators and barriers, and the willingness to share personal data. Scaled items were measured using a five--point Likert scale (1 = lowest degree of agreement or frequency, 5 = highest). Prior to deployment, the questionnaire was piloted with five respondents from the target group. Data collection took place in April and May 2026.

The sampling strategy combined purposive sampling via distribution within student groups at Mendel University in Brno and the snowball sampling method. The total number of valid responses reached n = 114. The sample consisted of 56.1% women and 43.9% men; the age distribution reflected the distribution channel—53.5% of respondents fell into the 21–23 age group, 27.2% into the 18–20 group, and 19.3% into the 24–27 group. In terms of education, university students accounted for 63.2% of the sample.

The qualitative component comprised six semi-structured interviews with respondents aged 20–27, with each of the three age cohorts represented by one male and one female. The interviews were guided by a protocol structured into five thematic blocks (shopping habits, AI in everyday life, AI in online shopping, AI in in-stores shopping, and motivators for future utilization). The average interview duration was approximately 15 minutes. The recordings were transcribed, anonymized using codes R1–R6, and processed using thematic analysis, wherein recurring motifs were synthesized into categories corresponding to the research questions.

The theoretical framework of the research is based on the Technology Acceptance Model (TAM) as presented by Pavlou (2003), which extends Davis's original formulation by incorporating perceived trust and perceived risk within the e-commerce context. Trust in an e-vendor has long been conceptualized in the literature through three dimensions: competence, benevolence, and integrity (McKnight *et al.*, 2002). An extended framework is further provided by the UTAUT2 model (Venkatesh *et al.*, 2012), which incorporates hedonic motivation, price value, and habit, in addition to performance expectancy and effort expectancy. On this basis, five hypotheses were formulated:

- ***H1:*** *Female members of Generation Z utilize AI tools in the online shopping environment more frequently than male members.*
- ***H2:*** *Male members of Generation Z exhibit a higher level of trust in AI tools within the offline environment than female members.*
- ***H3:*** *The intensity of AI tool utilization by Generation Z is statistically significantly higher in the online environment than in the offline environment.*
- ***H4:*** *Generation Z perceives barriers to AI utilization—particularly privacy concerns—to be more pronounced in the offline environment than in the online environment.*
- ***H5:*** *Generation Z's willingness to share personal data for personalization purposes is higher in the online environment than in the offline environment.*

Statistical data processing was performed using Statistica software. First, the Shapiro-Wilk test for normality was conducted for all variables. Due to deviations from a normal distribution, non-parametric alternatives were employed instead of the originally planned independent samples and paired t-tests: the Mann-Whitney U test was used for hypotheses concerning independent groups (H1, H2), and the Wilcoxon signed-rank test was utilized for paired comparisons between environments (H3, H4, H5). For the Wilcoxon test, respondents with a zero difference between the two ratings were excluded from the analysis. The level of statistical significance was set at \alpha = 0.05.

Ethical considerations of the research were strictly respected in compliance with the standards of the Faculty of Business and Economics at Mendel University in Brno. Participation in the questionnaire survey was voluntary and anonymous; consent to participate was confirmed by submitting the form. Participants in the qualitative component provided informed consent for the recording and processing of the interviews, and the audio recordings were deleted following transcription.

3 RESULTS

This section presents the primary findings from both empirical components of the study. The results are organized according to the shopping environment and conclude with a summary of the testing of the five hypotheses, complemented by additional insights from the interviews.

3.1 AI Tool Utilization in the Online Environment

In the online environment, the conscious utilization of generative assistants and visual tools dominates. The most prevalent AI tool was ChatGPT (or an alternative AI assistant), which was utilized by 63.2% of respondents within the last three months. The second most frequently reported tool was visual search via photography (38.6%), closely followed by recommendation systems such as “similar products” (37.7%). Furthermore, AI-generated review summaries were utilized by 27.2% of respondents, and e-shop chatbots by 16.7%. Augmented reality (AR) and virtual try-on technologies remain on the periphery, with a share of 5.3%. A total of 15.8% of the sample conducted their online shopping without utilizing any AI tools. The distribution of these utilized tools is illustrated in Figure 1.

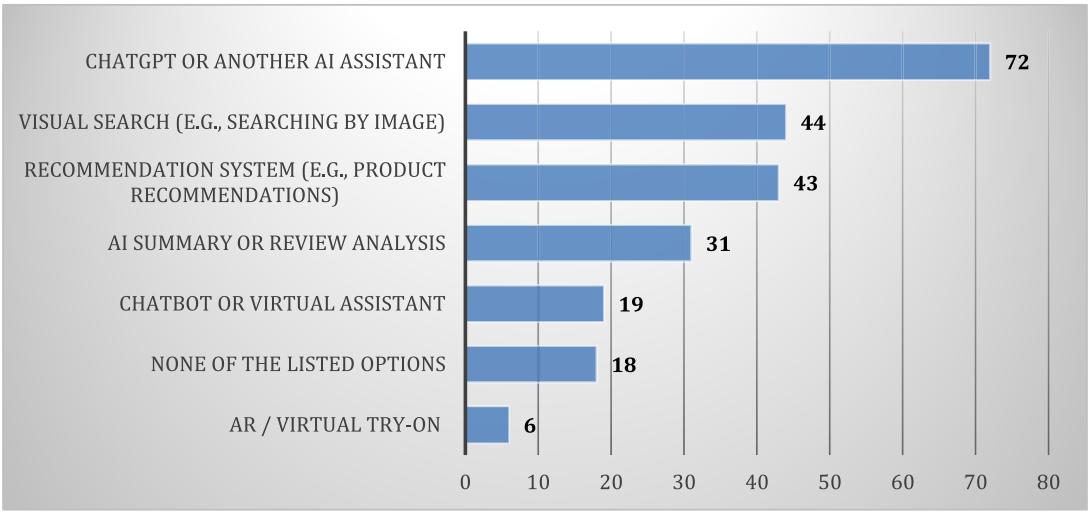


Fig. 1: AI tool utilization in online shopping within the last three months (n = 114)

The mean frequency of AI utilization in online shopping reached M = 2.54 (Mdn = 2.0; SD = 1.03) on a five-point scale, which corresponds to occasional usage. Conversely, the perceived benefit was notably more positive—the mean agreement score with the statement that AI tools facilitate online shopping was M = 3.41 (Mdn = 3.0; SD = 1.11).

3.2 AI Tool Utilization in the Offline Environment

In in-store shopping, experience with AI technologies was paradoxically high, albeit heavily concentrated. Self-checkout counters were utilized by 92.1% of respondents, interactive kiosks or digital screens by 66.7%, and automated loyalty-card discounts by 57.0%. Advanced elements, such as personalized in-store displays or voice assistants, were noted by only a minor fraction of the sample. Merely 3.5% of respondents reported no experience with AI elements within physical stores. The structure of these experiences is summarized in Figure 2.

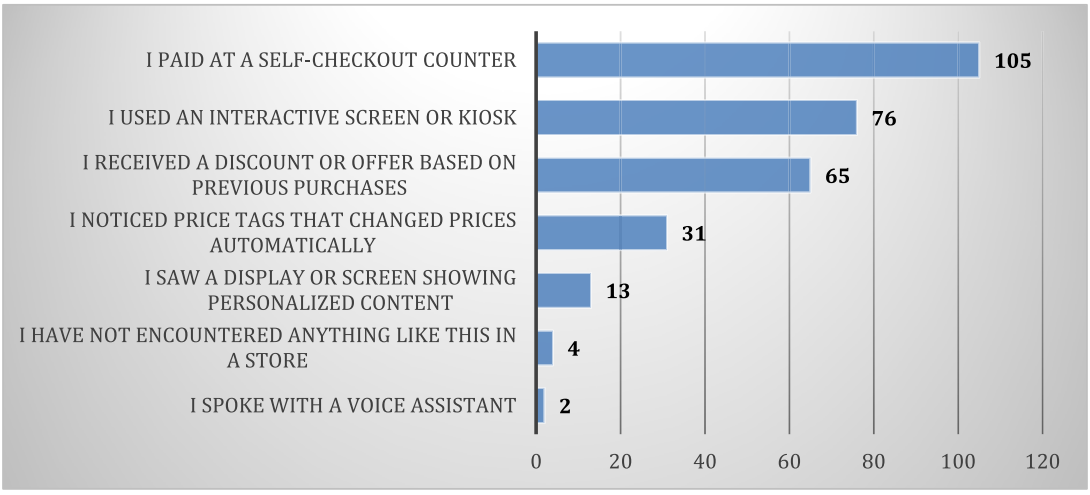


Fig. 2: Respondents' experience with AI elements in in-store shopping (n = 114)

The frequency of conscious AI utilization in the offline environment (M = 2.37; Mdn = 2.0; SD = 1.08) was slightly lower than in the online environment; the perceived benefit was also lower (M = 2.97; Mdn = 3.0; SD = 1.10). This discrepancy likely reflects the fact that numerous offline AI technologies—such as electronic shelf labels (ESLs) or predictive analytics systems—operate in the background, meaning consumers may not register their presence at all.

3.3 Motivators and Barriers to AI Adoption

The primary motivator for utilizing AI during shopping was speed and time savings (66.7%), followed by convenience and comfort (46.5%), personalized recommendations (40.4%), and discounts or promotional offers (38.6%). Hedonic motivation, manifested as curiosity toward new technologies, was cited by only 19.3% of respondents, indicating a highly utilitarian orientation.

On the side of barriers, 35.1% of respondents reported no impediments whatsoever. The most frequently cited barriers were distrust in the accuracy of recommendations (33.3%) and privacy concerns (29.8%). A preference for personal contact with sales staff plays a role for 21.9% of respondents, while technical complexity or a lack of familiarity with the tools deters 16.7% of the sample.

3.4 Hypothesis Testing

The Shapiro-Wilk test demonstrated significant deviations from a normal distribution across all tested variables; consequently, non-parametric alternatives were employed. The summarized results are presented in Table 1.

Tab. 1 Results of hypothesis testing for H1–H5.

Hypothesis	Statistical Test	Key Statistical Values	Result
H1: *Females utilize AI online more frequently*	Mann-Whitney	U = 1339.5; z = 1.485; p = .137	Not Supported
H2: *Males exhibit higher trust in offline AI*	Mann-Whitney U	U = 1598.5; z = 0.006; p = .995	Not Supported
H3: *AI usage intensity is higher online*	Wilcoxon signed-rank	T = 1121.0; z = 1.436; p = .151	Not Supported
H4: *Barriers are more pronounced offline*	Wilcoxon signed-rank	T = 84.0; z = 3.912; p < .001	Rejected (Inverse)
H5: *Willingness to share data is higher online*	Wilcoxon signed-rank	T = 137.5; z = 2.365; p = .018	Supported

Out of the five formulated hypotheses, only H5 was statistically supported—Generation Z's willingness to share personal data is statistically significantly higher in the online environment (M = 2.87) than in the offline environment (M = 2.69). Gender-based differences in the frequency of AI utilization were not confirmed in either the online or offline environments. Conversely, hypothesis H4 was rejected in the inverse direction of what had been originally hypothesized: privacy concerns are statistically highly significantly lower in the offline environment (M = 2.29) than in the online environment (M = 2.53). This counter-intuitive finding runs contrary to both initial assumptions derived from the literature review and represents the primary empirical contribution of this study.

3.5 Supplementary Insights from Qualitative Interviews

The qualitative analysis revealed several underlying mechanisms that expand upon the quantitative findings. First, the intensity of conscious AI utilization during online shopping is lower than the questionnaire values might suggest; respondents perceive a wide range of interactions with recommendation systems and algorithmic feeds as standard platform features rather than instances of artificial intelligence.

Second, passive acceptance of AI dominates the offline environment; respondents utilize self-checkout counters and loyalty applications without primarily identifying this interaction as an experience with artificial intelligence.

Third, the privacy paradox exhibits a more nuanced structure than captured by the questionnaire; respondents distinguish between preference data in the online environment (perceived as less sensitive) and behavioral or location data collected by loyalty programs in the offline environment. Within this dynamic, the smartphone serves as a critical mediator between both environments, as respondents frequently use it in-store for price comparison, photographing product codes, and retrieving reviews.

4 DISCUSSION

The findings of this study are fundamentally consistent with the framework of Marketing 5.0 advanced by Kotler *et al.* (2021), which defines AI as a technologically mediated yet human-centric tool within the customer journey. The dominance of ChatGPT and visual search in the online environment confirms that Generation Z adopts generative AI as a natural pre-purchase research tool rather than a technological novelty—a trend that aligns with Kingsnorth's (2022) conclusions regarding the integration of AI into the digital shopper journey. Conversely, the low level of conscious AI utilization in the offline environment resonates with the observations of Grewal and Roggeveen (2020), who point out that in-store shopping lags behind digital channels in implementing customer-facing technologies.

The validation of hypothesis H5—indicating a higher willingness to share personal data in the online environment—is consistent with the theoretical framing of the privacy paradox by McKee *et al.* (2024). While Generation Z declares privacy apprehensions, they willingly disclose data in practice in exchange for personalization, discounts, or time savings. The qualitative interviews not only illustrate this paradox but also deconstruct its underlying architecture: respondents clearly differentiate between data types and their respective contexts. Preference data in the online environment (such as purchase history or clothing size) are perceived as substantially less sensitive than the behavioral and location data collected by loyalty programs within in-store shopping.

The most significant empirical contribution of this study, however, resides in the unexpected outcome of hypothesis H4. The assumption derived from the literature review—that the physicality and tangibility of data collection in in-store shopping (consumer movement, loyalty cards, computer vision) would trigger a stronger sense of surveillance than invisible cookie-based tracking—was empirically rejected. On the contrary, privacy concerns are statistically highly significantly lower in the offline environment ($p < .001$). A plausible explanation is provided by the qualitative interviews: Generation Z exhibits heightened awareness regarding the mechanisms of online tracking, which contrasts with their less reflexive perception of in-store technologies, where many AI elements operate subtly in the background. Paradoxically, Generation Z perceives the

online environment as riskier to personal privacy, despite factually disclosing more data within it. This asymmetry underscores Kranzberg's tech thesis, which posits that technology is neither good nor bad, nor is it neutral relative to its social environment (Pitt *et al.*, 2023); within the available literature, particularly in the Czech and Central European context, this phenomenon has not yet been systematically explored.

Furthermore, these results contrast with the assumptions of the UTAUT2 model (Venkatesh *et al.*, 2012) regarding the prominent role of hedonic motivation. In this study, utilitarian motivation—time savings and convenience—overwhelmingly dominates, whereas curiosity and interest in novel technologies were cited by a mere 19.3% of respondents. Consequently, Generation Z perceives AI not as a source of entertainment, but as a functional utility whose value is measured by efficiency rather than novelty. This finding likely relates to the fact that AI no longer represents an innovation for this cohort, but rather a structural standard whose absence would be perceived negatively.

4.1 Practical Recommendations: Phygital Strategies

Three interconnected recommendations for retail practice emerge from the empirical findings. First, the creation of a **unified consumer data identity** that integrates the online user profile with the physical store's loyalty program. The lower privacy concerns documented in the offline environment (H4) provide strategic latitude for a more ambitious data convergence; however, this integration is strictly conditional upon a clearly communicated value proposition (discounts, personalization, or time savings).

Second, utilizing the **smartphone as a mediator**: the mobile application should synchronize the shopping cart in real time, navigate the consumer through the physical store layout, and deliver AI-generated review summaries or product visualizations via QR codes. Informants in the qualitative phase explicitly highlighted virtual try-ons within Augmented Reality (AR) as a key catalyst for future AI adoption.

Third, the implementation of **Explainable AI (XAI)** principles: the primary barrier identified in the online environment was distrust in recommendation accuracy (33.3%). Recommender systems must transparently articulate why a specific product is being displayed to the consumer, thereby eliminating the perception of the algorithm as an opaque "black box" (Knihová, 2024). In this context, trust in AI is not a static state, but a dynamic process cultivated through demonstrable system reliability and transparency (Glikson & Woolley, 2020).

4.2 Limitations and Future Research

This study encounters several methodological limitations that constrain the generalizability of its findings. First, the sample was obtained through a combination of purposive sampling and the snowball sampling method, primarily recruited via student communities at Mendel University in Brno. Consequently, higher education students constitute 63.2% of the sample matrix, meaning its generalizability to the entire population of Generation Z within the Czech Republic remains restricted.

Second, the cross-sectional research design captures consumer behaviors and attitudes specifically during April and May 2026. Given the exponential evolutionary trajectory of artificial intelligence tools, these findings may represent a transient snapshot with limited long-term validity.

Finally, the qualitative phase—consisting of six respondents—is inherently exploratory and does not achieve full thematic saturation.

5 CONCLUSION

This paper has demonstrated that Generation Z utilizes AI tools across both shopping environments with varying degrees of conscious awareness and intentionality. The online sphere is dominated by generative assistants and visual search, whereas self-checkout counters and loyalty applications prevail in in-store shopping, with their utilization often occurring passively. Statistical testing confirmed a higher willingness to share personal data in the online environment and, surprisingly, lower privacy concerns in the offline environment. This finding paves the way for retailers to implement "phygital" strategies grounded in a unified data identity, the smartphone as a mediator, and transparent algorithmic operations. Future research should expand the sample beyond student communities, employ a longitudinal design to track the evolution of consumer attitudes, and conduct a cross-generational comparative analysis.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from Operational Programme Research, Development and Education.

REFERENCES

DJAFAROVA, E., FOOTS, S. 2022. Exploring ethical consumption of generation Z: theory of planned behaviour. *Young Consumers*. 23(3), 413–431. https://doi.org/10.1108/YC-10-2021-1405

GLIKSON, E., WOOLLEY, A. W. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*. 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

GREWAL, D., ROGGEVEEN, A. L. 2020. Understanding Retail Experiences and Customer Journey Management. *Journal of Retailing*. 96(1), 3–8. https://doi.org/10.1016/j.jretai.2020.02.002

GRIGOREVA, E. A., GARIFOVA, L. F., POLOVKINA, E. A. 2021. Consumer Behavior in the Information Economy: Generation Z. *International Journal of Financial Research*. 12(2), 164–171. https://doi.org/10.5430/ijfr.v12n2p164

HALIBAS, A. *et al*. 2025. Unveiling the future of responsible, sustainable, and ethical consumption: a bibliometric study on Gen Z and young consumers. *Young Consumers*. 26(7), 142–171. https://doi.org/10.1108/YC-11-2024-2327

KINGSNORTH, S. 2022. *Digital marketing strategy: an integrated approach to online marketing*. 3rd edition. London: Kogan Page. ISBN 978-1-3986-0599-2.

KLÁNOVÁ, E. 2025. Generace Z: On-line je samozřejmost, ale zážitek hledá v prodejně. *Retail News*. 11(10), 16–18. ISSN 2336-7083.

KNIHOVÁ, L. 2024. *AI marketing playbook: jak ChatGPT a umělá inteligence mění svět marketingu*. Praha: Grada. ISBN 978-80-271-5226-1.

KOTLER, P., KARTAJAYA, H., SETIAWAN, I., 2021. *Marketing 5.0: technology for humanity*. Hoboken (NJ): Wiley. ISBN 978-1-119-66851-0.

McKEE, K. M., DAHL, A. J., PELTIER, J. W., 2024. Gen Z's personalization paradoxes: a privacy calculus examination of digital personalization and brand behaviors. *Journal of Consumer Behaviour*. 23(2), 405–422. https://doi.org/10.1002/cb.2199

McKNIGHT, D. H., CHOUDHURY, V., KACMAR, C. 2002. Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*. 13(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

PAVLOU, P. A. 2003. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce*. 7(3), 101–134. https://doi.org/10.1080/10864415.2003.11044275

PITT, C. *et al*. 2023. Artificial Intelligence, Marketing, and the History of Technology: Kranzberg's Laws as a Conceptual Lens. *Australasian Marketing Journal*. 31(1), 81–89. https://doi.org/10.1177/18393349211044175

VENKATESH, V., THONG, J. Y. L., XU, X. 2012. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*. 36(1), 157–178. https://doi.org/10.2307/41410412

Contact information

Benjamin Křištof: e-mail: benjamin.kristof@mendelu.cz

Jana Dannhoferová: e-mail: jana.dannhoferova@mendelu.cz, phone: +420 545 132 243

SERVICE MANAGEMENT FOR SMART BUILDING ADMINISTRATION USING VIRTUAL IOT SIMULATION AND PREDICTIVE ANALYTICS

David Kubart[1], Zuzana Špendel[1]

[1]Department of Economics, Faculty of Business and Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_018

ABSTRACT

This article addresses the design and implementation of a virtual service management model for smart building management within an enterprise environment. The work innovatively links the ITIL 4 methodology with a containerized IoT infrastructure, edge-cloud architecture, and predictive analytics. The objective of the research is to demonstrate that even a fully software-defined smart building environment can be systematically managed as a complex service with a clearly defined Service Level Agreement (SLA) and processes for effective incident and event management. To validate the proposed approach, a fully portable laboratory environment was developed using Docker containers, MQTT asynchronous communication, an InfluxDB telemetry data store, Grafana visualization, and Python modules for data generation and advanced analytics. Building behavior is dynamically configured via configuration files, allowing a single codebase to simulate various types of facilities with different operational priorities and SLA requirements. The proposed model was thoroughly verified through experimental scenarios, including simulations of network communication outages, MQTT message delays, sensor data errors, and complete physical disasters. The results demonstrate the practical usability and benefits of ITIL 4 principles (ITIL Foundation 2019) for digitalized building management, increasing data reliability, and automating operations.

Keywords: ITIL 4, Smart Building, Internet of Things (IoT), IT Service Management (ITSM, Predictive Analytics, MQTT, Service Level Agreement (SLA)
JEL Code: C88



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

In recent years, the digitalization of enterprise buildings has significantly transformed how organizations ensure operational reliability, security, energy efficiency, and user comfort. The development of the Internet of Things (IoT) and continuous telemetry collection allows building status monitoring at a much higher granularity than before, using the acquired data not only for operational oversight but also for qualified operations management (Cirani 2019). Within an enterprise environment, however, the mere availability of raw data is often insufficient. What matters critically is whether this data is correctly interpreted in relation to the quality of the service provided, operational risks, and the strategic goals of the organization. In practice, smart building management is often understood merely as a technical task encompassing data collection, status visualization, and basic automation interventions. While functional at the subsystem level, such a concept by itself is inadequate for comprehensive and systematic service management. There is a clear need to view smart building management as an integrated corporate service with established responsibilities, quality parameters, response rules for anomalies, and continuous improvement mechanisms. The ITIL 4 methodology represents a dominant and exceptionally suitable framework for this approach, as it allows purely technical events and operational states to be translated into the structured language of service management.

The primary goal of this work is to design and experimentally validate a service management model for smart building management using the principles of ITIL 4 to their full extent. The core of the research does not lie in the physical technology itself, but in the design and validation of the process and service approach. The technological layer acts here as a means through which the proposed model can be realized in a laboratory, simulated, and critically assessed for its real benefits and limits.

2 CURRENT STATE

Traditional Building Management Systems (BMS) are frequently proprietary, isolated, and lack the flexibility to integrate with modern cloud platforms and advanced analytics. While the rise of the Internet of Things opens possibilities for distributed management and detailed monitoring, it creates a fundamental problem with interpreting the enormous amount of data generated. A smart building is a typical example of a cyber-physical system where technical monitoring must be closely tied to operational interpretation. A single measured value (e.g., an elevated temperature) gains true meaning only in relation to the wider context, the purpose of the building, SLA limits, and the actual impact on end-users.

2.1 The ITIL 4 Methodological Framework in the Context of IoT

The ITIL 4 framework defines service management as a set of specialized organizational capabilities for generating value for customers in the form of services. For the smart building domain, the following practices form the core of our proposal:

- Monitoring and Event Management: Continuous monitoring of infrastructure states, filtering sensor noise, and capturing significant events that may indicate an emerging problem.
- Incident Management: Rapid detection, classification, logging, and effective resolution of operational deviations to restore standard service operations as quickly as possible (MTTR) and minimize negative impact (Adhikari *et all* 2024).

- Service Level Management (SLM): Defining clear, measurable parameters of service quality (monitoring availability, environmental stability) and relentlessly evaluating them against contractual obligations.
- Problem Management: Identifying and analyzing the deeper root causes of recurring incidents and finding preventive solutions to prevent their further occurrence.

2.2 The Role of Predictive Analytics

Time series are the natural data model of the Internet of Things. Predictive analytics (IBM 2026) applied to this data allows transforming reactive management into proactive and preventive management. Early prediction of physical variables (e.g., predicting that a critical temperature threshold in a server room will be breached within the next few tens of minutes) gives the system or operator time to react before an SLA breach and an incident actually occur. At the same time, however, it is necessary to view the predictive layer as a decision support tool rather than a 100% replacement for operational judgment, due to its inherent dependence on the quality of input data and the potential for false alarms (Géron 2022).

3 APPLICATION AND SYSTEM ARCHITECTURE DESIGN

The proposed system is conceived as a distributed edge-cloud architecture that reflects modern trends in IoT and edge computing. The underlying idea is to divide responsibility between a local autonomous unit (Edge layer) located directly in the facility and a central coordination and analytics platform (Cloud layer).

3.1 Distributed Architecture and Communication Model

The system consists of three main structural layers:

- Building Edge Layer: Runs directly within the simulated facility. It includes virtual IoT sensors generating telemetry (temperature, humidity, device status), a local MQTT communication bus, and a key component—the Edge Controller. This controller possesses local decision--making logic, a local Telegraf/InfluxDB database, and is capable of fully autonomous building management (e.g., HVAC control) even in the event of a total loss of cloud connectivity.
- VPN Layer: Provides a secure and isolated connection between the distributed edge locations and the central cloud via a virtual private network, closely simulating a real enterprise deployment.
- Cloud Layer: Represents the central brain of the system. It receives aggregated data, performs advanced data quality validation, runs predictive models, maintains a global service incident log, evaluates SLAs, and provides a web-based dispatching interface for operators. The communication model is strictly built on the MQTT protocol (Publish/Subscribe). The designed topic model is hierarchical (format: building/floor/room/sensor_type) and strictly distinguishes between local operations (frequent telemetry with a low QoS 0 level) and cloud communication (critical status messages, heartbeats, incidents, and reverse control commands with QoS 1 and the Retain flag).

3.2 Data and Simulation Model

The simulation model implements a real physical runtime loop that models thermal inertia, ambient environment influences, and the behavior of ventilation and heating. A key feature of the design is the ability to inject faults. On demand, the system allows simulating both software and hardware anomalies—ranging from generating invalid data and artificially delaying messages to simulating complete infrastructure crashes.

4 APPLICATION DESCRIPTION AND IMPLEMENTATION

The application was completely implemented as a fully containerized open-source solution. The entire system is divided into separate projects managed via Docker Compose, guaranteeing 100% portability, service isolation, and easy replicability of the laboratory environment (Docker 2026).

4.1 Edge Layer Implementation and Double MQTT Architecture

The local building utilizes a lightweight Docker image based on python:3.9-slim to run the simulation scripts and the Edge Controller. The Edge Controller is the heart of local intelligence and implements a critical double MQTT architecture, utilizing two parallel MQTT clients:

- local_client: Communicates exclusively with the local broker (Eclipse Mosquitto) inside the building's isolated network. It processes granular telemetry from all sensors at one-second intervals.
- cloud_client: Maintains an encrypted connection through the VPN to the central cloud. It publishes regular heartbeat messages, opens/closes incidents in the cloud registry, and sends only smartly aggregated data, radically conserving WAN network bandwidth.

The client includes a Last Will and Testament mechanism, ensuring that if an unexpected outage or connection breakdown occurs, the cloud broker immediately detects the OFFLINE status and the cloud core automatically initiates a highest-priority (P1) incident—loss of facility monitoring.

4.2 Cloud Core Engine and Data Quality Control Pipeline

The central cloud core (ITIL Core Engine) is implemented in Python and functions as the main service layer. A new and essential element of the implementation is a dedicated protective layer of the cloud data pipeline (validator.py). This component intercepts incoming MQTT messages from buildings before they are stored or forwarded to analytics modules, subjecting them to strict validation:

- Syntactic Check: Verifying correct JSON structure, the presence of mandatory metadata (building identifier, timestamp), and correct data types.
- Semantic Check: Detecting time drift between the generation time on the edge device and the arrival time in the cloud, uncovering hidden synchronization or network delay issues.
- Physical Validation: Checking whether measured values fall within physically possible boundaries (e.g., a room temperature showing a sudden jump of 50 °C within a single second), which indicates a failure of the hardware sensor itself.

If a message fails validation, it is immediately diverted to the error history and is not passed further into automation schemes or linear regression predictive models. A cloud State Tracker monitors the frequency of these errors within a sliding time window. If the number of faulty messages exceeds defined thresholds (DQ_P3_THRESHOLD and DQ_P2_THRESHOLD), the system automatically escalates a data quality incident because the data loses trustworthiness and cannot be safely used to manage the building.

4.3 Observability and Data Model

An InfluxDB time-series database was deployed to store telemetry data and operational logs at both local and cloud levels. The translation of asynchronous messages is handled by Telegraf middleware, which parses topics and transforms them into tags and metrics. The visualization layer is implemented in Grafana, where two types of dashboards are created: detailed technical monitoring for engineers at the local level, and managerial-dispatch reporting for tracking SLA fulfillment, service availability, and incident overviews at the cloud level.

5 EXPERIMENTS AND RESULTS EVALUATION

To experimentally verify the functionality of the entire proposed concept and its real link to ITIL 4 practices, an extensive set of comprehensive testing scenarios was conducted. The behavior of the entire ecosystem was rigorously monitored and logged using a unified logging model managed via .env configuration.

5.1 Test: Reference Autonomous Operation of the Building

The aim of the first experiment was to verify the Edge layer‘s basic capability to function autonomously and correctly synchronize its state after connectivity was restored. The building started in local mode without an immediate connection to the cloud. During this operation, the Edge Controller successfully detected local anomalies and stored them in a local incident ledger (ledger.json).

Following a simulated establishment of the VPN and MQTT connection with the cloud, immediate and flawless synchronization took place. A one-time transfer of 15 pending incident records was delivered to the cloud layer. The central dispatch then issued reverse control commands for the ventilation system (HVAC), leading to environmental stabilization and the successful, automatic closure of all incidents. The incident closure success rate reached 100%, precisely demonstrating the high reliability of local autonomy and distributed storage.

5.2 Test: MQTT Data Quality and Escalation of Data Quality Incidents

This key experiment focused on verifying the protective function of the validator.py component in the cloud data pipeline. Faulty and delayed sensor data were intentionally injected into the system to test the cloud‘s ability to detect a decline in data stream trustworthiness.

The test results confirmed the 100% success of the design. The system intercepted invalid messages, diverted them, and prevented their hazardous influence on automated building control. When the error limit within the sliding window was exceeded, the cloud immediately and correctly

generated a dedicated Data Quality incident (initially with P3 priority, subsequently escalated to P2). This experimentally demonstrated that the system can effectively protect management integrity and proactively alert the operator that telemetry data from the given facility is unreliable.

5.3 Test: Complex Physical Disaster in the Facility

To demonstrate the robustness of the solution in critical situations, a scenario simulating a discrete physical disaster—specifically, a massive water leak in a data center facility (DC_Brno_1)—was executed. This type of event poses an immediate and fatal risk to service availability, hardware safety, and overall business continuity.

The system‘s reaction took place across three fully coordinated layers:

- Edge Layer: The sensor immediately detected critical humidity, the Edge Controller received the error flag, instantly activated emergency mode, and rapidly generated a local incident with the highest priority, P1.
- Cloud Layer: The disaster message was transferred to the cloud as a priority using QoS 1. The central ITIL Core Engine immediately registered the building as an affected Configuration Item (CI), logged the P1 incident in the global registry, triggered a visual alarm on the Grafana dispatch dashboard, and started the countdown for evaluating strict SLA limits.
- Network Layer: VPN and network logs (vpn_audit.log) confirmed that stable and prioritized communication between the edge location and the cloud SLA engine occurred throughout the incident without any packet loss.

The incident was held open and active in the system until the disaster condition was physically resolved in the simulator and a confirming clear_fault command was sent. Upon receiving it, the cloud core correctly closed the incident, calculating the precise duration and final MTTR metric. This test successfully proved that the proposed model does not represent merely an isolated technical solution, but forms an integrated, auditable, and fully traceable incident process from the onset of a physical failure to its procedural closure.

6 EVALUATION AND DISCUSSION

Experimental data clearly confirms that linking the ITIL 4 methodology with the technical capabilities of the Internet of Things brings an entirely new quality to building management. The main benefits of the designed and implemented solution can be summarized into three key points:

- Operational Continuity and Autonomy: Thanks to the implementation of the Edge Controller with a local database and decision logic, the building is capable of operating safely and regulating its internal environment even when completely disconnected from the cloud. No data is lost, as incidents are stored locally and automatically synchronized immediately upon network restoration.
- High Data Trustworthiness: Deploying the validation component validator.py resolves a critical vulnerability in many IoT systems, which is blind trust in incoming data. The capability to detect syntactic, semantic, and physical anomalies and translate them into data quality incidents radically increases the security and stability of automated management.

- Managerial Oversight and SLA Measurability: Translating raw telemetry into standardized incidents with a clearly defined priority (P1 to P3) allows operations managers to allocate resources efficiently, track the actual availability of buildings as corporate services, and exactly evaluate compliance with contractual obligations to both suppliers and customers.

Regarding limitations and constraints, it must be acknowledged that the implemented solution was operated in a laboratory setup. In a real production deployment, it would be essential to eliminate certain intentional simplifications—most notably, implementing strict security for MQTT brokers (disabling anonymous access, introducing device authentication via certificates, TLS/MQTTS encryption, and detailed ACL rules for topic access). However, these steps were intentionally omitted from our laboratory version to preserve maximum code clarity and log readability, which in no way diminishes the validity of the scientific conclusions reached.

7 CONCLUSIONS

The goal of this research was to design, implement, and experimentally validate an innovative way in which the principles and practices of the international ITIL 4 framework can be effectively utilized to manage modern IoT smart building infrastructure supported by predictive analytics. The results of our work clearly show that the ITIL 4 methodology can be applied meaningfully and successfully far beyond its traditional domain of managing purely software and IT information systems. In the context of the Internet of Things and smart buildings, ITIL 4 functions as a robust translation and process framework that helps transform disparate technical data from hundreds of sensors and operational events into a coherently managed, transparent, and measurable corporate service. This approach creates a highly stable environment where it is possible not only to reliably monitor the current physical state of buildings but also to proactively evaluate the real impacts of incidents, effectively support operator decision-making, continuously increase operational data trustworthiness, and establish a firm foundation for deploying advanced predictive and preventive management for modern corporate facilities.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ADHIKARI, Deepak, JIANG, Wei, ZHAN, Jinyu, RAWAT, Danda B., BHATTARAI, Asmita. (2024). Recent advances in anomaly detection in Internet of Things: Status, challenges, and perspectives. *Computer Science Review*. 54, 100665. ISSN 1574-0137. https://doi.org/10.1016/j.cosrev.2024.100665

CIRANI, Simone, FERRARI, Gianluigi, PICONE, Marco, VELTRI, Luca. (2019). *Internet of things: architectures, protocols and standards*. Hoboken: Wiley. ISBN 978-1-119-35967-8

DOCKER. (2026). The Docker platform. *Docker Docs*. [accessed 2026-04-10]. https://docs.docker.com/get-started/docker-overview/

GÉRON, Aurélien. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. 3rd ed. Sebastopol: O'Reilly. ISBN 978-1-098-12561-5.

GRIFFOR, Edward. (2017). *Handbook of Cyber-Physical Systems*. Cham: Springer. ISBN 978-3-319-51827-5

IBM. (2026). What is predictive analytics? *IBM Topics*. [accessed 2026-03-15]. https://www.ibm.com/topics/predictive-analytics.

ITIL Foundation. (2019). *ITIL 4 Edition, 2019*. London: AXELOS / TSO. ISBN 978-0-113-31599-4

Contact information

Polina Komárková: e-mail: polia.tselisheva@mail.ru

LIGHTNING NETWORK AS AN ALTERNATIVE PAYMENT INFRASTRUCTURE FOR SMALL AND MEDIUM-SIZED ENTERPRISES

Tomáš Lacina[1], Petr Jedlička[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_019

ABSTRACT

This paper presents the design, implementation, and economic evaluation of a payment module that enables merchants to accept Bitcoin payments through the Lightning Network using a self-hosted Bitcoin node. While the Bitcoin base layer suffers from limited throughput, relatively high transaction fees, and longer confirmation times, the Lightning Network addresses these limitations through off-chain payment channels, enabling near-instant settlement at minimal cost. The study combines a theoretical analysis of the Bitcoin protocol and the Lightning Network with a comparative evaluation of traditional payment methods and cryptocurrency-based payments. Primary data were collected from the official pricing schedules of eight payment terminal providers and six payment gateway providers operating in the Czech Republic. A full Bitcoin and Lightning Network node was deployed on an HP 260 G2 Mini platform running Umbrel, Bitcoin Core, LND, and BTCPay Server. Based on this infrastructure, a web application named Flash was developed to support payment processing and transaction management. The functionality and usability of the application were verified through pilot user testing. The results indicate that transaction costs associated with the Lightning Network are several orders of magnitude lower than those of conventional payment terminals and payment gateways while maintaining immediate settlement. Furthermore, operating a self-hosted payment infrastructure requires only a modest initial investment. The study identifies merchant technical expertise and the limited adoption of Lightning Network wallets among consumers as the principal barriers to broader deployment. The findings suggest that the Lightning Network represents a technically feasible and economically attractive payment alternative for small and medium-sized enterprises seeking to reduce transaction costs and increase payment-processing autonomy.

Keywords: Lightning Network, Bitcoin, payment infrastructure, transaction costs, decentralized payment systems, BTCPay Server, small and medium-sized enterprises, cryptocurrency payments

JEL Code: E42, G23, L86



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Cashless payment methods have been steadily increasing their share of transactions in the Czech Republic. According to the payment behavior survey conducted within the Prosperity Index project, 60% of Czech customers prefer payment cards when shopping online, while 12% choose bank transfers and 9.3% use QR-code payments (Index Prosperity, 2024). This trend is consistent with international findings demonstrating that the expansion of digital payment instruments systematically displaces cash transactions, with transaction speed and perceived security representing key determinants of payment method selection (Schuh & Stavins, 2016; Arango-Arango & Suárez-Ariza, 2020). The diversification of available payment methods is not only a matter of customer convenience but also has economic implications for merchants, who bear transaction costs associated with each cashless payment.

Traditional payment infrastructure based on card networks, payment gateways, and bank transfers combines card scheme fees, interchange fees, and margins charged by payment terminal providers. The resulting fee for a single transaction charged by a typical payment terminal provider in the Czech Republic ranges from 0.75% to 1.95% (Dotypay Fees, 2025; KB SmartPay, 2023; ČSOB START Tariff, 2022). Transaction fees charged by payment gateways range from 0.5% to 1.99% and are often supplemented by setup fees or monthly subscription charges (GoPay, 2025; Barion, 2025). Furthermore, settlement to the merchant's bank account may be delayed by one to three business days.

One of the primary motivations behind the creation of Bitcoin was the development of a payment system that does not require a trusted third party. In this framework, peer-to-peer transactions are verified through cryptographic mechanisms and a decentralized network of nodes (Nakamoto, 2008). However, the base layer of the Bitcoin blockchain exhibits several well-known limitations for everyday retail use. It processes only tens of transactions per second, maintains an average block interval of ten minutes, and transaction fees may rise to several dollars during periods of network congestion, making micropayments economically inefficient (Novák, 2023).

A response to these limitations is the Lightning Network, a second-layer solution built on top of the Bitcoin blockchain. The Lightning Network operates through shared payment channels established between pairs of nodes, where balance allocations are continuously updated while only the opening and closing of channels are recorded on the main blockchain (Antonopoulos *et al.*, 2021). As a result, transactions can be settled almost instantly with fees measured in satoshis, representing fractions of a cent. The Lightning Network is among the most advanced layer-two protocols built on blockchain infrastructure and has attracted increasing academic attention in recent years from the perspectives of security, network topology, and economic behavior (Gudgeon *et al.*, 2020; Zabka *et al.*, 2022). Empirical analyses indicate that despite a growing number of nodes and channels, actual retail adoption remains limited.

Despite its evident technical advantages, the adoption of cryptocurrency payments among retail merchants remains relatively low. Empirical research identifies perceived complexity, uncertainty regarding customer behavior, and limited visibility of merchant-side benefits as the primary barriers (Jonker, 2019). From the consumer perspective, cryptocurrency holders still represent a relatively small segment of the population, and their motivations for ownership are heterogeneous, combining speculative interests with distrust of the traditional financial system (Auer & Tercero-Lucas, 2022). A specific challenge associated with the Lightning Network is the technical complexity of operating a self-hosted node, the lack of documented reference implementations of integrated payment modules, and limited awareness of the associated economic implications (Bouška, 2023). Nevertheless, operating a proprietary Bitcoin node represents a crucial prerequisite for achieving full sovereignty over payment acceptance. Without a self-hosted node, merchants remain dependent

on third-party infrastructure within the Lightning Network ecosystem, which partially undermines the primary benefits of decentralized architecture in terms of resilience and competitiveness.

In the context of small and medium-sized enterprises (SMEs), the choice of payment infrastructure acquires a significant economic dimension. Reducing transaction costs and shortening settlement times directly contribute to profit margins, particularly for businesses operating with low unit margins or high transaction volumes. The adoption of blockchain technologies in business has become an established field of research emphasizing the critical role of technical readiness and supporting infrastructure (Wamba *et al.*, 2020). Furthermore, independence from card schemes and banking intermediaries increases operational resilience against technical failures or regulatory disruptions affecting these providers.

The objective of this paper is to present a comprehensive design and reference implementation of a payment module enabling merchants to accept Bitcoin payments through the Lightning Network via their own Bitcoin node, supplemented by a quantitative comparison of transaction costs with traditional payment methods in the Czech Republic. Specifically, the paper (i) summarizes the cost structures associated with payment terminals, payment gateways, bank transfers, and cryptocurrency payments; (ii) describes the proposed system architecture integrating a Bitcoin node, LND, and BTCPay Server with a web application; and (iii) presents the results of pilot usability testing of the implemented application. The outcome is a set of recommendations for merchants considering the inclusion of Lightning Network payments within their payment portfolio.

2 METHODOLOGY AND DATA

To achieve the stated objective, an engineering design and information systems development methodology was employed, combined with a qualitative and quantitative comparison of payment methods. The research process was structured into five consecutive stages:

(1) Deployment and configuration of a self-hosted Bitcoin node.
(2) Deployment of a Lightning Network node and establishment of payment channels.
(3) Requirements analysis and architectural design of the payment module.
(4) Implementation and pilot testing of the web application.
(5) Comparative evaluation of payment methods.

At the infrastructure level, a compact HP 260 G2 Mini desktop computer was used, equipped with an Intel Core i3-6100U processor, 16 GB DDR4 RAM, and a 1.92 TB enterprise-grade SSD (SanDisk MLC with power-loss protection and DRAM cache). The hardware platform was selected based on requirements for long-term stability, sufficient storage capacity relative to the current size of the Bitcoin blockchain (approximately 745 GB at the time of the study), and low power consumption.

The device was configured with Umbrel version 1.4.1, a specialized Linux distribution designed for operating self-hosted Bitcoin and Lightning Network nodes through containerized deployment. The Bitcoin layer was provided by the reference implementation Bitcoin Core, while the second layer was implemented using LND (Lightning Network Daemon) version 0.18.5. To generate invoices and integrate Lightning Network payments with the application, the open-source BTCPay Server platform was deployed. Within the Umbrel environment, BTCPay Server operates as a containerized service and automatically integrates with the remaining node components. The selection of LND reflects its status as the most widely deployed Lightning Network implementation in the production network (Gudgeon *et al.*, 2020).

On top of this infrastructure, a web application named Flash was designed and implemented. The system architecture follows a three-tier model consisting of:

- • the presentation layer (frontend),
- • the application layer (backend),
- • the data layer.

The frontend was developed using React, with Vite serving as the build tool and Tailwind CSS providing styling functionality. The server-side application runs on Node.js using the Express framework and exposes interfaces both to the client application and to BTCPay Server. Data persistence is handled through a MySQL relational database, while user authentication is implemented using JSON Web Tokens (JWT).

The data model includes entities representing users, stores, products, categories, receipts, receipt items, price-change history, VAT rates, and currencies. The system distinguishes between two user roles: a standard user (cashier) and an administrator with extended CRUD privileges for user accounts, categories, and products.

3 RESULTS

The deployed Bitcoin node achieved full synchronization with the Bitcoin blockchain and demonstrated stable operation. The acquisition cost of the hardware solution amounted to approximately CZK 5,100, consisting of CZK 1,500 for the computer, CZK 600 for 16 GB DDR4 memory, and CZK 3,000 for the SSD storage device. The average power consumption of the system is approximately 12 W, which, at an electricity price of CZK 6.5 per kWh, corresponds to annual operating costs of approximately CZK 683. Consequently, energy consumption does not constitute a significant component of the total cost of node operation.

The software layer of the node was deployed in a complete configuration. The Bitcoin Node application, implementing Bitcoin Core, performs validation of the entire blockchain history and actively maintains peer-to-peer connections with other nodes in the network. The Lightning Node application, implementing LND version 0.18.5, provides the functionality of the second-layer payment network. During operation, two public payment channels were established. At the time of verification, balances amounted to 29,341 satoshis in the on-chain wallet and 179,670 satoshis in the Lightning Network wallet, with two active channels and five connected peers. BTCPay Server was automatically integrated with the LND node within the Umbrel environment and served as the payment gateway responsible for invoice generation and payment-status monitoring.

The implemented Flash web application supports the principal operational processes required by merchants, including order creation, selection of products from a categorized inventory, shopping-cart management, settlement through two payment methods (cash and Bitcoin Lightning Network), management of VAT rates and discounts, and receipt generation. The administrative role provides complete management of user accounts, product categories, and inventory items, including price-change history.

For Lightning Network payments, the application invokes the BTCPay Server interface, which generates an invoice containing a QR code. The customer scans the QR code and completes the payment using a Lightning Network wallet. Upon confirmation of payment receipt by BTCPay Server, the application records the transaction data in the database and generates a receipt.

Tab. 1 Multi-criteria comparison of payment methods from the merchant's perspective

Criterion	Cash	Card payment	Payment gateway	Bank transfer	QR payment	BTC	BTC LN
Payment settlement speed	★★★★★	★	★	★★★★★	★★★★★	★★★	★★★★★
Transaction fees	★★★★★	★★	★★	★★★★★	★★★★★	★★★★★	★★★★★
Technical complexity	★★★★★	★★★	★★★	★★★★	★★★★	★★★	★★★★
Security	★★★	★★★★★	★★★★★	★★★★★	★★★★★	★★★★★	★★★★★
Suitability for small business	★★★★★	★★★	★★★	★★★	★★★★	★★★★★	★★★★★

The key outcome of the comparative evaluation is the quantitative comparison of transaction costs presented in Table 1. The values for card payments were obtained from calculations performed using the Dotypay fee calculator for a MasterCard debit card, while the payment gateway values were based on the promotional fee rate of 1.25% published by PayU.

Tab. 2 Comparison of Merchant Transaction Costs for Selected Payment Amounts

Payment method	CZK 100	CZK 1,000	CZK 10,000	CZK 25,000	Settlement time
Cash	CZK 0	CZK 0	CZK 0	CZK 0	Immediate
Payment Card (POS Terminal)	CZK 2.33	CZK 14.65	CZK 137.86	CZK 343.21	1–3 days
Payment Gateway	CZK 1.25	CZK 12.50	CZK 125.00	CZK 312.50	1–3 days
Bank Transfer / QR Payment	CZK 0	CZK 0	CZK 0	CZK 0	Immediate to 1 day
Bitcoin (On-Chain)*	CZK 0	CZK 0	CZK 0	CZK 0	~10 minutes
Bitcoin Lightning Network	Few satoshis	Few satoshis	Tens of satoshis	Tens of satoshis	Immediate

* For on-chain Bitcoin transactions, the transaction fee is paid by the sender (customer); the merchant incurs no fee for receiving the payment.

Source: Author's calculations based on fee schedules published by payment service providers (Dotypay, 2025; KB SmartPay, 2023; ČSOB, 2022; GoPay, 2025; Barion, 2025; PayU, 2025).

The results demonstrate that for transactions in the range of several thousand Czech crowns—typical of retail purchases—the transaction costs of card payments amount to approximately CZK 14 per transaction, whereas Lightning Network transaction costs remain several orders of magnitude lower, measured in only a few satoshis. Given that average retail profit margins are often only a few percentage points, this difference represents a meaningful component of the economic competitiveness of small businesses. For higher transaction values, the potential savings increase proportionally with transaction volume.

The multi-criteria evaluation further indicates that cash payments and the Lightning Network share the advantages of negligible fees and immediate settlement. However, they differ substantially in terms of technical complexity and customer anonymity. Card payments and payment gateways offer a high degree of convenience and broad customer acceptance but at the cost of transaction fees and reduced privacy. Bank transfers executed through QR codes represent the most cost-effective cashless option for merchants within the traditional banking infrastructure, although they do not facilitate cost-efficient operation outside the SEPA area.

Pilot usability testing of the Flash application involving five participants confirmed the comprehensibility and usability of the interface. All participants successfully completed all ten assigned tasks without external assistance. Completion times varied among participants, with more pronounced differences observed during administrative operations. Older participants generally required more time to navigate the interface. User feedback indicated a need for greater visual emphasis on inventory-management controls; however, overall usability evaluations were positive. Given the limited sample size, the findings should be interpreted as indicative results characteristic of a pilot study.

Additional evidence of growing cryptocurrency payment adoption in the Czech Republic can be found in publicly available industry data. Confirmo, a cryptocurrency payment processor, reported settlement volumes of USD 383 million in 2024, representing a year-on-year increase of 260% (Confirmo Turnover, 2025). Similarly, the Czech e-commerce retailer Alza reported cryptocurrency payment volumes of approximately CZK 201 million in 2021, corresponding to roughly 0.5% of total annual revenue (Cryptocurrency Payments on Alza.cz, 2022).

4 DISCUSSION AND CONCLUSIONS

The conducted study demonstrates that the Lightning Network operated through a self-hosted Bitcoin node represents a technically feasible and economically advantageous alternative to traditional payment infrastructures for retail transactions. For typical transactions ranging from hundreds to thousands of Czech crowns, Lightning Network transaction costs are several orders of magnitude lower than those associated with payment terminals and payment gateways, while maintaining virtually instantaneous settlement. The initial investment required for the node hardware amounts to only several thousand Czech crowns, while annual electricity costs remain in the range of a few hundred crowns. In the context of small and medium-sized enterprises, such an investment can be recovered within a relatively short period through savings on transaction fees alone.

From the perspective of the economic competitiveness of small businesses, a proprietary payment infrastructure offers three principal advantages. First, reducing marginal transaction costs improves the sustainability of business models operating with low unit profit margins, such as small cafés, bars, or other hospitality establishments, where card-payment fees may significantly reduce the gross profit generated by individual sales. Second, the immediate availability of funds,

compared to the two- or three-day settlement periods commonly associated with card-payment systems, improves operational cash flow. The importance of these two factors is consistent with empirical findings indicating that settlement speed and perceived security significantly influence payment-method selection from both the consumer and merchant perspectives (Schuh & Stavins, 2016). Third, independence from third-party infrastructure enhances operational resilience against technical failures affecting payment gateways and card-processing systems. From a sustainability perspective, it is also noteworthy that a self-hosted node does not require data-centre infrastructure and can operate efficiently on low-power hardware.

Despite these economic benefits, several significant barriers to wider adoption remain. On the merchant side, the primary obstacle continues to be technical competence. Although platforms such as Umbrel are designed to simplify deployment and management, operating a fully functional Lightning Network node—including channel balancing, channel-state backup, and liquidity management—requires a deeper understanding of the protocol or access to external technical support. This finding aligns with broader research on blockchain adoption in business environments, where technical readiness and supporting infrastructure are repeatedly identified as critical success factors (Wamba *et al*., 2020).

Empirical research on cryptocurrency-payment adoption among merchants further suggests that perceived complexity and uncertainty regarding customer demand often outweigh purely financial incentives (Jonker, 2019). From the consumer perspective, the adoption of Lightning Network wallets remains limited. Although the availability of both custodial and non-custodial wallet solutions continues to grow, the majority of consumers have not yet interacted with this technology. Contemporary research further indicates that motivations for cryptocurrency ownership are heterogeneous and combine speculative interests with distrust of the traditional financial system. Consequently, only a subset of cryptocurrency holders is motivated to actively use digital assets for everyday payments (Auer & Tercero-Lucas, 2022).

This dual-sided adoption barrier reflects a classic network-effect problem. The economic value of accepting Lightning Network payments increases as the proportion of customers possessing compatible wallets grows, while customer incentives to use such wallets increase as merchant acceptance expands.

The limitations of the present study should be interpreted in four dimensions:

- First, the pilot usability testing involving five participants represents an exploratory validation exercise intended primarily to identify major usability issues rather than provide statistically representative measurements of user experience. A more rigorous evaluation would require a substantially larger sample size and the incorporation of standardized usability assessment instruments.
- Second, the multi-criteria comparison of payment methods combines qualitative expert evaluation with quantitative cost calculations. The subjective component of the assessment should be replaced in future research by a structured evaluation framework validated through expert-panel review.
- Third, the economic advantage of the Lightning Network was quantified primarily through transaction-fee comparisons. In real-world deployment scenarios, merchants must also consider the volatility of the Bitcoin exchange rate relative to the Czech koruna, which may introduce additional costs or risks.

Fourth, the study does not address the regulatory dimension, which is essential for production deployment in commercial environments. Furthermore, empirical analyses of Lightning Network topology indicate that network capacity is concentrated among a limited number of large nodes (Zabka *et al.*, 2022). For newly entering merchants, this implies the necessity of carefully selecting channel management and connectivity strategies.

Based on the findings presented above, several practical recommendations can be formulated:

- For merchants considering the inclusion of Lightning Network payments within their payment portfolio, a sensible initial approach is to operate a self-hosted node for low-value transactions, where transaction fees constitute a significant proportion of unit profit margins. For more advanced deployments, combining a self-hosted node with automatic conversion of cryptocurrency into fiat currency through a third-party service may reduce exposure to exchange-rate volatility while preserving the economic advantages of low transaction fees.
- For the research community, the study opens several promising avenues for future investigation, including:
  - longitudinal analyses of the actual operating costs of proprietary payment infrastructures,
  - quantitative studies of network effects in retail Lightning Network adoption,
  - comparisons of energy consumption and overall carbon footprints among alternative payment infrastructures.

Given the ongoing development of layer-two protocols (Gudgeon *et al.*, 2020), a natural continuation of this research would be a comparative analysis of the Lightning Network and competing layer-two solutions from the perspective of retail-payment applications.

4.1 Conclusion

In conclusion, the proposed and implemented reference architecture of a payment module demonstrates both the technical feasibility and the economic viability of accepting Bitcoin payments through the Lightning Network using a self-hosted Bitcoin node within a small- or medium-sized-business environment.

The principal contribution of the study lies in providing an accessible, reproducible, and well-documented implementation that lowers the entry barrier for merchants considering the integration of Lightning Network payments into their payment mix. In addition, the study offers an empirical comparison of transaction costs directly relevant to the Czech market environment.

Future research should focus on the long-term economic implications of node operation, the evolving regulatory framework governing cryptocurrency payments, and customer-side barriers to adoption. Such investigations would contribute to a more comprehensive understanding of the role that decentralized payment infrastructures may play within the broader digital-payments ecosystem.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ANTONOPOULOS, A. M., OSUNTOKUN, O., PICKHARDT, R. 2021. *Mastering the Lightning Network: A Second Layer Blockchain Protocol for Instant Bitcoin Payments*. Sebastopol: O'Reilly & Associates. ISBN 978-1-492-05486-3

ARANGO-ARANGO, C. A., SUÁREZ-ARIZA, N. F. 2020. Digital payments adoption and the demand for cash: New international evidence. *Journal of Payments Strategy & Systems*. 14(4), 392–410. ISSN 1750-1806

AUER, R., TERCERO-LUCAS, D. 2022. Distrust or speculation? The socioeconomic drivers of U.S. cryptocurrency investments. *Journal of Financial Stability*. 62, 101066. ISSN 1878-0962.

BARION. 2025. *Barion*. [Accessed 2025-05-08]. https://www.barion.com/cs/

BOUŠKA, K. 2023. *Velká kniha o Bitcoinu a kryptoměnách, aneb, Finanční gramotnost pro každého*. Praha: Braiins Publishing. ISBN 978-80-908709-5-6

ČSOB. 2022. *Ceník POS START*. [Accessed 2025-05-08]https://platbakartou.csob.cz/wp-content/uploads/2022/12/cenik-POS-START.pdf

CZECHCRUNCH. 2022. Češi a krypto na Alze: Loni zaplatili bitcoinem za počítače, čipy i Pokémony 201 milionů, sílí ale i obavy. *CzechCrunch*. [Accessed 2025-05-08] https://cc.cz/cesi-a-krypto-na-alze-loni-zaplatili-bitcoinem-za-pocitace-cipo-pokemony-201-milionu-sili-ale-i-obavy/

CZECHCRUNCH. 2025. Česká kryptobrána vyrostla o 260 procent na devět miliard. Letos chce být Confirmo ještě větší. *CzechCrunch*. [Accessed 2025-05-08] https://cc.cz/ceska-kryptobrana-vyrostla-o-260-procent-na-devet-miliard-letos-chce-byt-confirmo-jeste-vetsi/

DOTYPAY. 2025. Poplatky. *CzechCrunch*. [Accessed 2025-05-08]. https://dotypay.com/poplatky/

GOPAY. 2025. *GoPay*. [Accessed 2025-05-08]. https://www.gopay.com/cs/

GUDGEON, L., MORENO-SANCHEZ, P., ROOS, S., MCCORRY, P., GERVAIS, A. 2020. SoK: Layer-Two Blockchain Protocols. In: BONNEAU, J., HENINGER, N. (eds.). *Financial Cryptography and Data Security*. FC 2020. Lecture Notes in Computer Science, vol. 12059. Cham: Springer, pp. 201–226. ISSN 0302-9743

INDEX PROSPERITY. 2024. *Platební chování*. [Accessed 2025-05-08]. Available at: https://www.indexprosperity.cz/2024/platebni-chovani/

JONKER, N. 2019. What drives the adoption of crypto-payments by online retailers? *Electronic Commerce Research and Applications*. 35, 100848. ISSN 1567-4223.

KB SMARTPAY. 2023. *Přehled nákladových složek celkového poplatku za transakci*. [Accessed 2025-05-08]. Available at: https://www.kbsmartpay.cz/wp-content/uploads/2023/07/KBSmartPay_Prehled-nakladovych-slozek-celkoveho-poplatku-za-transakci-22011.pdf

NAKAMOTO, S. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*. [Accessed 2025-05-08]. Available at: https://bitcoin.org/bitcoin.pdf

NOVÁK, M. 2023. *Lightning network: platby budoucnosti*. Praha: Brains Publishing. ISBN 978-80-908709-4-9.

PAYU. 2025. *PayU*. [Accessed 2025-05-08]. https://czech.payu.com/

SCHUH, S., STAVINS, J. 2016. How do speed and security influence consumers' payment behavior? *Contemporary Economic Policy*. 34(4), 595–613. ISSN 1074-3529

WAMBA, S. F., QUEIROZ, M. M., ROUBAUD, D., PAPPAS, I. O. 2020. Blockchain adoption in operations and supply chain management: empirical evidence from an emerging economy. *International Journal of Production Research*. 58(7), 2055–2072. ISSN 0020-7543

ZABKA, P., FÖRSTER, K.-T., DECKER, C., SCHMID, S. 2022. Empirical Evaluation of Nodes and Channels of the Lightning Network. *Pervasive and Mobile Computing*. 83, 101584. ISSN 1574-1192

Contact information
Petr Jedlička: e-mail: petr.jedlicka@mendelu.cz

FINE-TUNING ROBERTA FOR SENTIMENT ANALYSIS OF SUPPORT CONVERSATIONS WITH PRIVACY-PRESERVING ANONYMISATION

Vojtech Mahdal[1], Oldrich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_020

ABSTRACT

This article describes a prototype system for processing support-conversation text. The system accepts a text input, classifies sentiment into three classes (positive, neutral, negative), masks selected sensitive items such as names, e-mails, phone numbers, URLs, usernames and long numeric strings, and stores an anonymised record. The sentiment classifier is based on RoBERTa and was fine-tuned in three stages on public chat and dialogue datasets. Evaluation showed that the second stage reached the best score on its internal test split, but its accuracy on a separate combined test set fell from 83.41% to 70.06%. After a final fine-tuning stage on a smaller dataset with a higher share of neutral examples, the model reached 92.82% accuracy on 1,573 independent samples. The anonymisation layer is rule-based and is most suitable for structured patterns; names, especially sentence-initial capitalised words, remain the main limitation. The result is a working demonstration consisting of a FastAPI service, a JavaScript dashboard and an SQLite log, containerised with Docker and deployed on Hugging Face Spaces.

Keywords: sentiment analysis, RoBERTa, fine-tuning, anonymisation, GDPR, customer support
JEL Code: C88

1 INTRODUCTION

Customer-support conversations can form a substantial body of free text that organisations may want to analyse, for example to understand recurring problems, customer attitudes and the tone of communication. At the same time, such texts may contain personal data, including names, e-mail addresses or phone numbers. This paper therefore treats sentiment analysis and anonymisation as linked tasks: the system extracts a three-class sentiment label from a text input and masks selected sensitive items before the result is stored (Mahdal, 2026).

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

The thesis behind this paper (Mahdal, 2026) sets a narrow, practical goal: take a piece of conversation text, judge its sentiment, strip the sensitive parts, and keep only what is safe to store. Sentiment was chosen as the first structured signal to extract because it is useful on its own and because it lets the result be checked against published three-class models. The work is meant as a teaching example of the whole chain, from picking a base model to deploying a running service, rather than as a finished commercial product.

1.1 Extracting structure from conversations

Pulling structured facts out of free text sits in the wider area of information extraction, which recent surveys frame increasingly around large language models (Xu *et al.*, 2024). Sentiment analysis is one of its older and better-understood branches: Liu (2012) treats it as the task of deciding the polarity that a span of text expresses. Conversational settings add their own twist, because turn-taking and short replies carry context that a single sentence loses (Mariani *et al.*, 2023). We kept three sentiment classes rather than a positive–negative split, so that messages with no clear stance — a large share of support traffic — have a label of their own.

The model family used here goes back to the transformer (Vaswani *et al.*, 2017) and to BERT-style pretraining. Among the available encoders, RoBERTa (a Robustly optimised BERT approach) drops the Next Sentence Prediction objective, switches to dynamic masking and trains on a much larger corpus, roughly 160 GB against 16 GB for BERT (Kotamraju, 2022). For a three-class polarity task that trade favours accuracy over the lighter, faster DistilBERT, which is why RoBERTa was selected. The Hugging Face Transformers library (Hugging Face, 2022) supplied both the pretrained weights and the training loop.

1.2 Anonymisation as a constraint, not an afterthought

Because the stored text must be GDPR-compliant, anonymisation is part of the pipeline rather than a later clean-up step. The field offers a spread of options, from rule-based masking to learned models, with recent benchmarks comparing the two on the same data (Asimopoulos *et al.*, 2024; Shen *et al.*, 2024). The thesis takes the simpler route: a set of regular expressions that replace e-mails, URLs, phone numbers, usernames, long digit strings and capitalised name sequences with placeholder tokens. This is deliberately modest, and Section 4 is honest about where it slips.

2 METHODOLOGY

The pipeline runs in Python and has four parts: load data, mask sensitive items, fine-tune the classifier, then serve and log results. Before any model sees the text, a mask_sensitive function applies the regular expressions, turning e-mails into `<EMAIL>`, URLs into `<URL>`, phone numbers into `<PHONE>`, usernames into `<USER>`, long digit sequences into `<NUMBER>` and runs of capitalised words into `<NAME>`. Labels were mapped to integers (negative 0, neutral 1, positive 2) and text was tokenised with the pretrained RoBERTa tokenizer using padding and truncation.

Data were split in a stratified way, with roughly 70–80 % for training, 10–20 % for validation and the rest for testing. Class weighting was added to push back against imbalance, which mattered a great deal here: the first dataset was 75.78 % neutral, so without weighting the model could score well by guessing neutral. Training used the Hugging Face Trainer with a learning rate of 2e-5,

three epochs and weight decay of 0.1. The work ran on MetaCentrum (CESNET) JupyterHub with a 6-core CPU, 32 GB RAM and Nvidia A40 GPU.

The classifier was not trained once but in three stages, each on a different dataset, building on the previous weights:

Stage 1 used `chat-sentiment-analysis` (Hugging Face, about 28.8k rows), first plain and then with class weights. **Stage 2** added EmpatheticDialogues (about 25k dialogues, 32 emotions remapped to three classes). The **final stage** added the small Chat Sentiment Dataset (about 500 rows), brought in specifically because the model still read the neutral class poorly. Why the middle stage hurt rather than helped is the more interesting part of the story, and Section 3 reports it with the numbers.

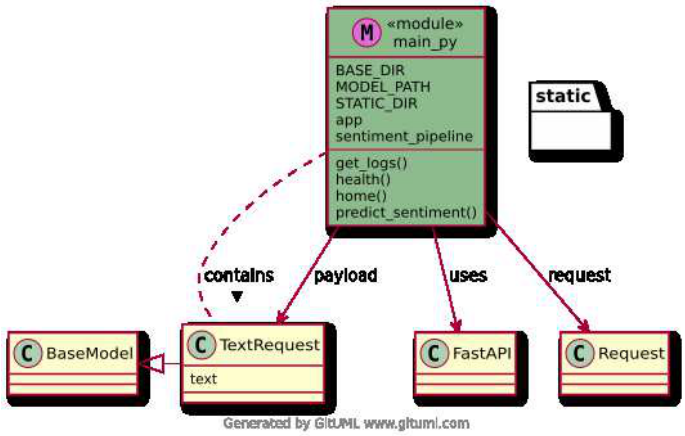


Fig. 1: UML class diagram of the main application module, showing the FastAPI service, the sentiment-analysis pipeline and the predict_sentiment endpoint (adopted from Mahdal, 2026)

The trained model was published to the Hugging Face Hub and loaded at runtime by a REST API built with FastAPI, which exposes a POST /predict endpoint and OpenAPI docs. A JavaScript dashboard sits on top of the API. Each prediction is written to an SQLite log through a thread-safe, regex-based anonymisation module, so the database never holds the raw sensitive text. The whole thing was containerised with Docker and deployed to Hugging Face Spaces. Figure 1 shows the class layout of the main module.

3 RESULTS

Two evaluation contexts have to be kept apart, because mixing them would misread the work. The first is the internal test split measured after each fine-tuning stage (Table 1). The second is a separate, combined test set of 1 573 unique samples that the model never saw during training (Table 2). The internal split and the independent set do not agree, and that disagreement is the point.

On the internal split the second stage looks like the winner, with an F1 of 0.9560. Taken at face value that would suggest stopping there. The independent set tells the opposite story.

Version 2, the one that scored highest on the internal split, dropped to 70.06% on the independent set, a fall of 13.35% against version 1 (Mahdal, 2026). The final version recovered and went past both, reaching 92.82%, which is an improvement of 22.76% over version 2 and the model shipped in the deployed service. The counts are consistent with the totals (1102 + 471 = 1573; 1460 + 113 = 1573).

The error breakdown in Table 3 shows where the trouble lived. Version 2 confused neutral with the two poles in 353 cases, far more than the other versions; almost all of its damage was neutral text pushed into positive or negative. The final version cut those mistakes sharply, and what remains is dominated by a single residual type, neutral read as positive (80 cases). Figure 2 shows the same picture as a confusion matrix for the final model.

Against two community RoBERTa sentiment models (`cardiffnlp/twitter-roberta-base-sentiment-latest` and `j-hartmann/sentiment-roberta-large-english-3-classes`) the trained model read standalone sentiments more accurately, the gap being clearest on the neutral class. This comparison is qualitative: the thesis presents it as side-by-side figures and does not report numeric accuracy for the two baselines, so none is quoted here.

The running service behaves as intended. On the dashboard the sentence “Everything is set up correctly!” was scored Positive with a confidence of 99.67 %, and a POST request with “FastAPI works correctly.” returned the matching positive label. Figure 3 shows the deployed interface.

Tab. 1 Metrics on the internal test split after each fine-tuning stage (not the independent set)

Metric	Stage 1 baseline	Stage 1 weighted	Stage 2 (dataset2v2)	Final (dataset3)
Eval loss	0.4481	0.5240	**0.2030**	0.3346
Accuracy	0.9243	0.9277	**0.9561**	0.9153
Precision	0.9282	0.9286	**0.9560**	0.9160
Recall	0.9243	0.9277	**0.9561**	0.9153
F1-score	0.9257	0.9281	**0.9560**	0.9152
Epochs	3	3	**3**	3

Tab. 2 Table 2: Accuracy on the combined independent test set (1 573 samples)

Model version	Accuracy	Correct	Incorrect
Version 1	83.41%	—	—
Version 2	70.06%	1102	471
Final version	**92.82%**	**1460**	**113**

Tab. 3 Most frequent error types by model version (count of class confusions)

Error (confusion)	Version 1	Version 2	Final version
neutral positive	65	212	80
neutral negative	54	141	19
positive neutral	72	88	5
negative neutral	68	26	7
positive negative	2	3	1

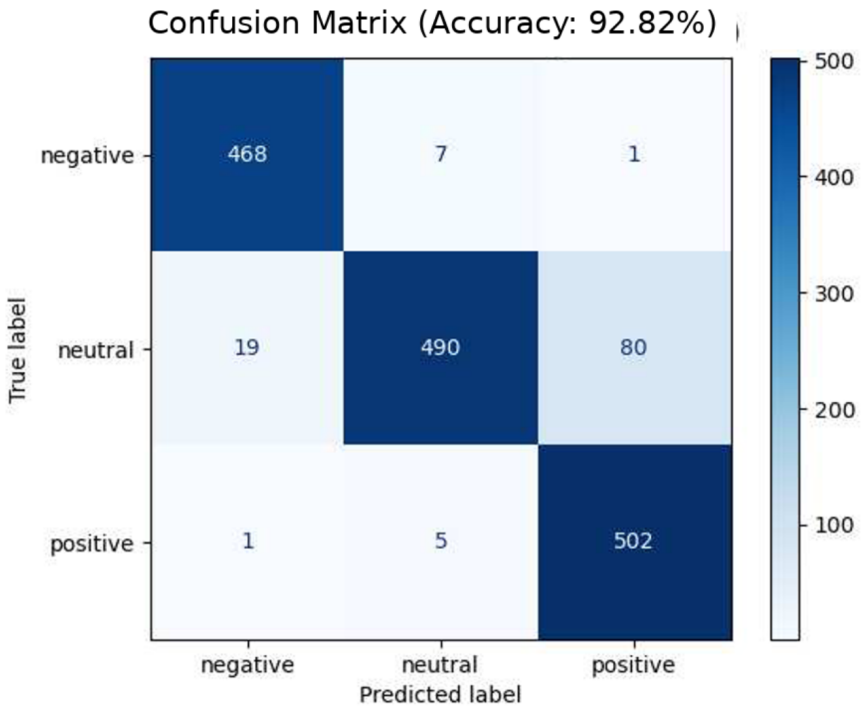


Fig. 2: Confusion matrix of the best-performing sentiment classification model on the three-class task (negative/neutral/positive), reaching 92.82% accuracy (adopted from Mahdal, 2026)

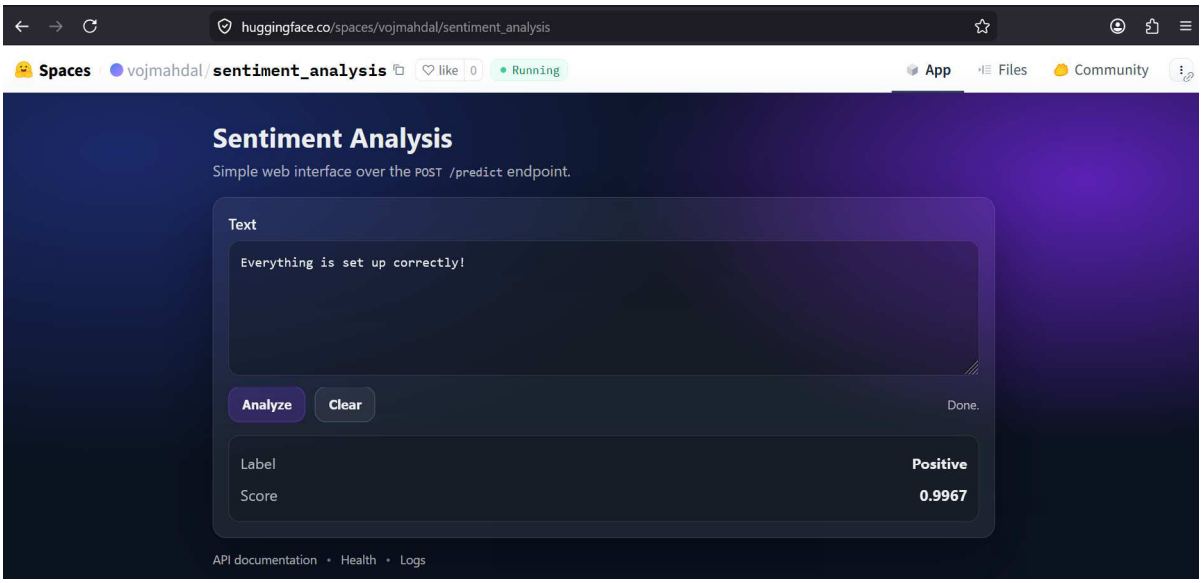


Fig. 3: Deployed web interface (Hugging Face Spaces) over the POST /predict endpoint, returning the predicted sentiment label and confidence score for an input sentence (adopted from Mahdal, 2026).

4 DISCUSSION

The gap between Table 1 and Table 2 is the main result worth discussing. The second stage achieved the strongest internal-split metrics, but the independent test set showed a much weaker result. The thesis links this behaviour to the different class distribution after remapping Empathetic Dialogues from 32 emotions into three sentiment classes, especially to weaker representation of neutral examples. It is therefore more accurate to describe this as a plausible explanation supported by the error pattern, not as a proven causal mechanism (Mahdal, 2026).

A plausible interpretation is that the final dataset helped mainly because of its class composition rather than its size. The Chat Sentiment Dataset contained a higher share of neutral examples than the preceding stage, and after fine-tuning on it the number of neutral-class confusions decreased from 353 in version 2 to 99 in the final version. The conclusion should therefore be formulated cautiously: in this experiment, adding a smaller but more balanced dataset improved generalisation on the independent test set, especially for the neutral class.

The anonymisation layer should be presented as a rule-based prototype rather than a fully reliable GDPR solution. Regular expressions are appropriate for structured patterns such as e-mails, URLs, phone numbers, usernames and long numeric strings, but the thesis itself identifies names as a difficult case, particularly when a capitalised word appears at the beginning of a sentence. For stronger anonymisation, the rule-based layer should be complemented or replaced by a trained named-entity recognition model, as suggested by recent anonymisation research (Asimopoulos *et al.*, 2024; Shen *et al.*, 2024).

One more caveat bounds the whole result. The independent test set was built from generated and public text made of short, plain sentences, so 92.82 % describes the model on simple, non-domain text. Sarcasm, irony, long compound sentences and the vocabulary of a specific product line were not tested. We would not expect the same figure on raw, messy support logs, and the work does not claim it.

5 CONCLUSION

The work produced a functional prototype for sentiment analysis and anonymised storage of support-conversation text. A RoBERTa classifier fine-tuned in three stages reached 92.82 % accuracy on a combined independent test set of 1,573 samples, and the model was deployed through a FastAPI service with a JavaScript dashboard and an SQLite log. The main transferable finding is that inter-nal validation metrics alone were not sufficient: the second model version looked strongest on its internal split but performed poorly on the independent test set. The results suggest that the neutral class and class balance require special attention in three-class sentiment analysis. The main limitations are that anonymisation remains rule-based and incomplete for names, and that the reported accuracy was measured on generated and public short-text samples rather than on real support logs. Future work should therefore evaluate the system on domain-specific conversations and replace or supplement regex masking with a trained entity-recognition component.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Mahdal, 2026) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

ASIMOPOULOS, D., SINIOSOGLOU, I., ARGYRIOU, V., KARAMITSOU, T., FOUNTOUKIDIS, E., GOUDOS, S. K., MOSCHOLIOS, I. D., PSANNIS, K. E., SARIGIANNIDIS, P. 2024. Benchmarking Advanced Text Anonymisation Methods: A Comparative Study on Novel and Traditional Approaches. Version 1. *arXiv*:2404.14465 https://doi.org/10.48550/ARXIV.2404.14465

HUGGING FACE. 2022. Transformers Documentation. *Hugging Face*. https://huggingface.co/docs/transformers/v4.17.0/en/index

KOTAMRAJU, S. 2022. Everything You Need to Know About ALBERT, RoBERTa and DistilBERT. *Towards Data Science*. https://towardsdatascience.com/everything-you-need-to-know-about-albert-roberta-and-distilbert-11a74334b2da/

LIU, B. 2012. *Sentiment Analysis and Opinion Mining*. San Rafael: Morgan & Claypool. https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

MAHDAL, V. 2026. *Analysis and extraction of data from conversations between users and support services* [in Czech]. Master's thesis, head: Trenz, O. Brno: Mendel University in Brno. .

MARIANI, M. M., HASHEMI, N., WIRTZ, J. 2023. Artificial Intelligence Empowered Conversational Agents: A Systematic Literature Review and Research Agenda. *Journal of Business Research*. 161, 113838. https://doi.org/10.1016/j.jbusres.2023.113838

SHEN, W. *et al*. 2024. *GDPR Compliance via NLP*. New York: ACM. https://doi.org/10.1145/3690931.3691009

VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L., POLOSUKHIN, I. 2017. Attention Is All You Need. *Advances in Neural Information Processing Systems.* 30.

XU, D. *et al*. 2024. Large Language Models for Generative Information Extraction: A Survey. *Frontiers of Computer Science*. 18, 186357. https://doi.org/10.1007/s11704-024-40555-y

Contact information

Vojtech Mahdal: e-mail: xmahdal@node.mendelu.cz

Oldrich Trenz: oldrich.trenz@mendelu.cz

INCOME, ATTITUDES, AND FOOD WASTE: SEGMENTATION OF CZECH HOUSEHOLDS

Anna Martinková[1], Veronika Antošová[1], Lucie Veselá[1]

[1]Department of Marketing and Trade, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_021

ABSTRACT

Household food waste is a significant environmental and economic problem in the Czech Republic — it is estimated that households generate a significant portion of total food waste. The present article examines whether household income affects the level of food waste and attitudes towards it, and complements this survey on consumer segmentation based on purchasing and consumption behavior. The empirical basis is a questionnaire survey conducted in January and February 2025 among the population of the Czech Republic (n = 294), supplemented by semi-structured in-depth interviews (n = 6). The statistical dependence between the monitored variables was verified by Pearson's chi-square test of independence, and the strength of the dependence was verified by Cramer's coefficient; hierarchical cluster analysis using the Ward method was used to segment respondents. The results showed that the proportion of food thrown away per month was not statistically dependent on income, while attitudes towards waste did depend on income and age. A significant predictor of lower waste was the frequency of checking the refrigerator. Cluster analysis identified three household segments with different purchasing, waste and perception approaches. The findings suggest that quantitatively measured waste levels and attitudes towards the problem may diverge, which has implications for the development of targeted communication strategies.

Keywords: food waste, households, income, consumer behavior, segmentation, cluster analysis, attitudes, Czech Republic
JEL Code: D12, Q53, M31

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Food waste is currently perceived as one of the most serious cross-cutting problems of the food system, combining environmental impacts, economic losses, and the ethical dimension of food inaccessibility for part of the population. The United Nations Sustainable Development Goal 12.3 aims to halve global food waste at the retail and consumer levels by 2030 (Eurostat, 2024). Within the European Union, food waste generation is continuously monitored in the context of this commitment both in the processing sector and at the levels of distribution and households, with households consistently representing the single largest source of waste (Eurostat, 2024).

The situation in the Czech Republic is comparable. Previous studies repeatedly document that the majority of food waste is generated at the final consumption stage (Kormaňáková, Remešová and Vančová, 2021; Nováková, Hák and Janoušková, 2021). Consequently, attention is logically focused on the purchasing and consumption behaviour of Czech households and on the socioeconomic factors shaping such behaviour.

The economic dimension of the issue is significant: the value of food ending up in municipal waste represents a portion of household expenditure that can be successfully reduced through relatively simple organisational adjustments. The environmental dimension—greenhouse gas emissions associated with the production, transportation, and disposal of unconsumed food—makes household food waste one of the measurable contributors to the carbon footprint of final consumption (Aloysius *et al.*, 2023). Alongside the economic and environmental dimensions, a social dimension also plays a role, as food waste contrasts with the growing number of households facing income insecurity (Czech Statistical Office, 2024).

Among the socioeconomic predictors of food waste, household income is frequently discussed. However, the literature does not provide a clear conclusion regarding the relationship between income and the level of food waste. Some studies demonstrate that higher income correlates with larger amounts of discarded food or with less cautious stock management (Schanes, Doberning and Gözet, 2018; Bravi *et al.*, 2020). Other studies, however, point out that the relationship is not straightforward: the influence of income overlaps with household size, life stage, demographic structure, and attitudes toward consumption (Aloysius *et al.*, 2023; Gonzáles-Santana *et al.*, 2022). Furthermore, Qi, Roe and van Rijnsoever (2016) show that higher-income households may frame food waste as an acceptable side effect of quality eating habits, meaning that attitudes toward the issue may diverge from observable behaviour indicators.

Another recurring topic in the literature is the role of discount promotions and retail marketing campaigns. The classical thesis (Schanes, Doberning and Gözet, 2018) describes discounts as incentives for impulsive purchasing, leading to overconsumption and subsequent waste; however, more recent empirical studies (Tsalis, Bouptrup Jensen and Aschemann-Witzel, 2024) question this straightforward mechanism. A third consistently documented factor is household organisation itself—particularly stock control, shopping planning, and the use of shopping lists (Kubíčková, Veselá and Kormaňáková, 2021; Kubíčková *et al.*, 2022).

In the Czech context, empirical evidence regarding the relationship between these factors and actually measured (or at least self-reported) household behaviour remains limited. Existing studies either rely on aggregated statistics (Czech Statistical Office, 2024) or focus on specific subpopulations and regions (Nováková, Hák and Janoušková, 2021; Kormaňáková, Remešová and Vančová, 2021). Less attention has been paid to the contemporary verification of quantitative hypotheses concerning the effects of income and accompanying factors, as well as to qualitative insight into the mechanisms underlying observed behavioural patterns.

Therefore, the aim of this article is to:

(1) (verify whether the income level of Czech households is statistically associated with the self--reported proportion of discarded food and attitudes toward food waste;
(2) examine complementary hypotheses concerning the effects of age, education, stock control, and the use of discounts; and
(3) identify household segments differing in the level of food waste and attitudes toward it based on purchasing and consumption variables.

The conclusions of the article are discussed with regard to the possibilities of targeted communication aimed at reducing household food waste and build upon the general framework of consumer decision-making (Kotler and Keller, 2013), although the primary focus remains empirical and quantitative.

2 METHODOLOGY AND DATA

To fulfil the research objectives, a combination of quantitative and qualitative research methods was employed in a sequential design.

The quantitative survey was conducted in the form of an online questionnaire using Google Forms, distributed primarily through social media and targeted email communication during the period from 20 January to 28 February 2025. The questionnaire was structured into three sections: sociodemographic data, purchasing habits (shopping frequency, influence of discounts, stock control, planning), and food waste behaviour itself (types of discarded food, reasons for disposal, attitudes). Prior to the main data collection, pilot testing was carried out on a sample of 22 respondents, and the feedback obtained was incorporated into the final version of the questionnaire. The minimum respondent age was set at 15 years.

A total of 303 responses were collected; after excluding irrelevant and incomplete responses, the final sample consisted of 294 units. Gender and the number of dependent children in the household were used as quota characteristics; the sample distribution corresponded to the 2021 Population Census data with a maximum deviation of 4.3 percentage points (Czech Statistical Office, 2024). Women represented 51.5% of the sample and men 48.5%. Regarding age structure, the most represented groups were 15–25 years (34.0%) and 26–35 years (29.6%), whereas the groups aged 56–65 years and 66 years and above together accounted for only 7.5% of the sample. This uneven age distribution resulted from the exclusively online distribution of the questionnaire and is discussed as a limitation in the Discussion section. Monthly household income most frequently fell within the range of CZK 50,001–70,000 (29.3%), while households with income below CZK 25,000 represented 10.6% of the sample.

Statistical processing relied on basic descriptive statistics and Pearson's chi-square test of independence. The significance level was set at $\alpha = 0.05$; the null hypothesis of independence between two categorical variables was rejected when the p-value was lower than this threshold. The strength of association was quantified using Cramér's V coefficient. In accordance with the objectives of the article, the following hypotheses were tested (formulated as null hypotheses):

- ***H01: Household expenditure on food is not dependent on the number of household members.***
- ***H02: The frequency of using discount promotions is not dependent on household income.***
- ***H03: The percentage of discarded food is not dependent on household income.***

- ***H04: The percentage of discarded food is not dependent on the frequency of checking food stocks in the refrigerator.***
- ***H05: The percentage of discarded food is not dependent on the use of discount promotions.***
- ***H06: Attitudes toward the problem of food waste are not dependent on household income level.***
- ***H07: Attitudes toward the problem of food waste are not dependent on age.***
- ***H08: Attitudes toward the problem of food waste are not dependent on the highest level of education attained.***

The formulation of the hypotheses was based on previous literature—H02 follows Gupta (2016), H04 follows the methodology of Kubíčková *et al.* (2022), H05 follows Schanes, Doberning and Gözet (2018), H06 follows Qi, Roe and van Rijnsoever (2016), and H07 and H08 again follow Kubíčková *et al.* (2022). H01 and H03 test general assumptions regarding the relationship between household size, household income, and consumer behaviour.

Hierarchical cluster analysis was employed among the multivariate methods for respondent segmentation. Input variables—including shopping frequency and types of purchases, use of discounts and loyalty programmes, stock control, frequency of discarding different food categories, reasons for disposal, and attitudes toward the issue—were transformed onto a unified 0–1 scale. Clustering was performed using Ward's method with Euclidean distance.

The qualitative research followed the quantitative phase and took the form of semi-structured in-depth interviews. Six interviews were conducted—two representatives from each identified cluster—with the aim of deepening the understanding of attitudes and motivations typical for individual segments. The interviews were thematically structured around food evaluation, sensitivity to discounts, barriers to sustainable purchasing, and information needs. The characteristics of the sample and the limitations of the research design are systematically discussed in the Discussion and Conclusions section.

3 RESULTS

The purchasing and consumption behaviour of respondents corresponds to patterns typical of Czech urban and mixed environments. Most respondents purchase food once or twice a week, with supermarkets representing the dominant shopping location, followed by small retail stores. Among the factors influencing purchasing decisions, quality and price were evaluated as the most important, followed by discounts and shelf life, while brand emerged as the least significant factor. A shopping list is always prepared by 17% of respondents, and another 40% prepare one for more than half of their shopping trips; however, only one quarter of those who create shopping lists follow them strictly. The remaining three quarters admitted being tempted by discounts or purchasing additional items during shopping. Discount promotions are used during every shopping trip by 26% of respondents, and during more than half of shopping trips by another 26%; only 2% reported never using discounts. Food waste is considered a serious problem by 78% of respondents, while 16% perceive it as not very serious and 6% do not consider it a problem at all.

The results of hypothesis testing concerning the relationships among the examined variables are summarised in Table 1. Of the eight tested hypotheses, four were rejected (H01, H04, H06, H07).

Tab. 1 Summary results of hypothesis testing (Pearson's chi-square test, Cramer's coefficient V; N = 294)

Hypothesis (H0)	p-value	Cramer's coefficient
H01: Spending × Number of Household Members	0,0001	0,212
H02: Frequency of Discounts × Income	0,9694	0,104
H03: % Food Waste × Income	0,6366	0,132
H04: % Food Waste × Inventory Control	0,0224	0,190
H05: % Food Waste × Discount Usage	0,3242	0,124
H06: Attitude towards Waste × Income	0,0093	0,212
H07: Attitude towards Waste × Age	0,0248	0,187
H08: Attitude towards Waste × Education	0,3589	0,122

The key finding is that the direct relationship between household income level and the self-reported proportion of discarded food (H03) was not confirmed—the p-value substantially exceeds the significance threshold, and the value of Cramér's coefficient indicates only a very weak potential association. In contrast, attitudes toward the problem of food waste (H06) were found to depend on income level—the relationship is statistically significant and of moderate strength according to Cramér's coefficient. Attitudes are also related to age (H07), but not to the highest level of educational attainment (H08).

Among the variables associated with household organisation, the frequency of checking food stocks in the refrigerator (H04) was confirmed as a significant predictor: the more frequently households check their food supplies, the lower the proportion of monthly purchased food they ultimately discard. Neither the effect of discount usage itself on the level of food waste (H05) nor the relationship between the frequency of using discounts and income level (H02) was statistically confirmed.

Cluster analysis identified three respondent segments with n = 68 (Cluster 1), n = 113 (Cluster 2), and n = 113 (Cluster 3). The segments differ markedly both in shopping habits and in the level and perception of food waste (see Figure 1).

Segment A (Cluster 1) consists predominantly of younger respondents (most strongly represented by the 26–35 age group) living in two- to three-member urban households; it contains the highest proportion of households with incomes exceeding CZK 100,000. These respondents shop more frequently for smaller purchases, make substantial use of online food delivery services, and are relatively more influenced by brand in their purchasing decisions. They exhibit the highest level of food waste among all segments—the average self-reported proportion of food discarded monthly reaches 0.29 (on a 0–1 scale). At the same time, this segment demonstrates the most indifferent attitude toward the issue: a smaller proportion of respondents consider food waste a serious problem compared to the other segments. They also make the least use of loyalty and discount programmes and prefer smaller food package sizes as a suitable measure for reducing waste.

Segment B (Cluster 2) comprises the youngest respondents (predominantly aged 15–25 years), with a higher proportion of university-educated individuals and a balanced income distribution. This group demonstrates the most active use of discounts, loyalty programmes, and mobile applications, most frequently checks food supplies before shopping, and reports consistent planning

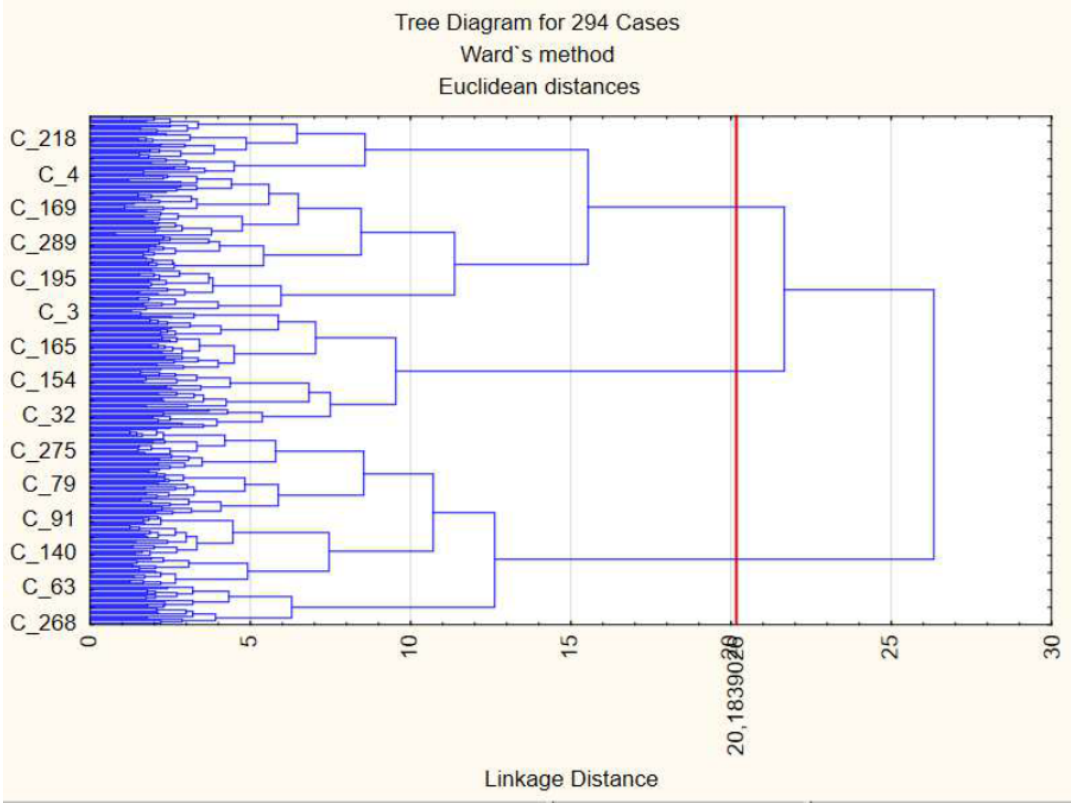


Fig. 1: Hierarchical cluster analysis dendrogram

behaviour. The average proportion of discarded food (0.19) is lower than in Segment A. In terms of attitudes, this segment perceives food waste as the most serious problem and considers better shopping planning to be the most effective preventive measure.

Segment C (Cluster 3) represents the oldest age group, with a higher representation of individuals aged 56 years and above, the highest proportion of households with monthly incomes below CZK 35,000, and more frequent residence in rural areas. This segment is characterised by careful shopping planning, consistent monitoring of expiration dates, and the use of discounted food products nearing the end of their shelf life. It demonstrates the lowest level of food waste (0.08). Food waste is perceived as a problem; however, the motivation to reduce it stems primarily from financial caution and practical experience rather than environmental convictions.

Across all segments, the most common reasons for discarding food include failure to consume food in time, exceeding the use-by date, and reaching the minimum durability date. The most frequently discarded categories are bakery products, fruit, vegetables, and dairy products. The qualitative interviews confirmed these quantitative patterns and supplemented them with motivational differences: whereas respondents from Segment A typically do not perceive their behaviour as wasteful, respondents from Segment B intentionally prevent waste through various organisational practices, and respondents from Segment C prefer traditional approaches such as freezing, composting, and feeding leftovers to animals.

4 DISCUSSION AND CONCLUSIONS

The empirical findings of this study offer a useful counter to the hypothesis that higher household income directly leads to higher food waste. The statistical dependence between income and the declared proportion of monthly food waste (H03) was not confirmed in the Czech sample, not even as a weak tendency. This result is in line with the conclusions of some foreign literature (Schanes, Doberning and Gözet, 2018; Aloysius *et al.*, 2023), which emphasizes that income should be interpreted together with the life stage, size and structure of the household and with subjective attitudes. At the same time, it is necessary to point out a possible methodological explanation: subjective self-estimates of the proportion of food waste tend to be systematically lower than objectively measured values (Kubíčková, Veselá and Kormaňáková, 2021), and the underestimation may be different in different income groups. Moreover, the influence of income may be better reflected in the absolute quantity or value of food thrown away than in its share in monthly purchases — which the survey on which the article is based did not follow.

In contrast, the dependence of the attitude towards the problem of waste on income (H06) was statistically confirmed. This finding corresponds to the work of Qi, Roe and van Rijnsoever (2016), according to which households with higher incomes more often frame waste as a side effect of quality food and therefore as more acceptable. Qualitative interviews support this interpretation: respondents in Segment A typically do not perceive waste as waste, attribute the problem to retail chains and perceive their own behavior as neutral. There is therefore a tension between quantitatively measured behavior and its subjective framing, which in itself represents a challenge for communication in the field of waste reduction.

Among other results, it is worth highlighting the confirmed connection between the frequency of checking the stock in the refrigerator and the share of food thrown away (H04). This finding is consistent with the conclusions of Kubíčková *et al.* (2022) and the broader literature on the role of household organization in reducing waste. The practical implication is available: a simple routine of checking inventory before shopping is among the least expensive measures that can be addressed through a communication campaign.

The data did not confirm the effect of the use of discount events on the level of waste (H05) or the relationship between the frequency of discounts and income (H02). This is in line with more recent evidence (Tsalis, Bouptrup Jensen and Aschemann-Witzel, 2024), which questions the simple thesis of "discounts = excessive purchases = waste". Qualitative interviews in this study demonstrate that discounts affect different segments in different ways: respondents with lower incomes actively seek out discounts and integrate them into planning (Segment C), while for respondents with higher incomes, discounts tend to modify specific purchasing decisions in an unplanned manner (Segment A). The insignificant aggregate result may thus conceal conflicting mechanisms within the population.

Segmentation based on cluster analysis yielded a structured description of three distinct types of households that systematically differ from each other in terms of both waste rates (0.29 / 0.19 / 0.08), attitudes towards the problem and preferred solutions. This distinction has practical implications: measures that one segment perceives as relevant (better shopping planning — Segment B) are not necessarily effective for the other (Segment A prefers smaller packaging; Segment C actively seeks discounted food before the end of its shelf life). For retail chains and public campaigns, this suggests that communication should be differentiated by segment — in particular, increasing the visibility and availability of discounted food before the end of its shelf life, expanding the range of smaller packages of perishable food, and supplementing communication with information on proper storage (Huang *et al.*, 2021; Corsini *et al.*, 2023; Eičaité *et al.*, 2022).

The study findings should be interpreted with a few limitations in mind. The sample of 294 respondents does not have a representative age distribution — a consequence of the questionnaire being distributed exclusively online — and the proportion of seniors is significantly lower than in the population. The subjective estimate of the proportion of food wasted in monthly purchases, although methodologically economical, does not capture the absolute quantity or value of food wasted. The six in-depth interviews provide only illustrative insight into the mechanisms identified in the cluster analysis, not their saturated description. Further research should focus on (i) replication on a larger and more age-representative sample, (ii) addition of objective measurement of waste (diary or weight methods), and (iii) longitudinal verification of the relationship between attitude and behavior within individual segments.

In terms of the objectives of this study, it can be stated that empirical evidence has been obtained that the relationship between income and food waste is not linear in the Czech context: the level of income is not primarily reflected in the declared share of food thrown away, but rather in the household's attitude towards the problem of waste and — through this attitude and related purchasing habits — in segment-specific behavioral patterns. This finding is the basis for a differentiated communication strategy aimed at reducing food waste in the Czech Republic.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ALOYSIUS, N., ANANDA, J., MITSIS, A., PEARSON, D. (2023). Why people are bad at leftover food management? A systematic literature review and a framework to analyze household leftover food waste generation behavior. *Appetite*. 186, 106577. ISSN 0195-6663. https://doi.org/10.1016/j.appet.2023.106577

BRAVI, L., FRANCIONI, B., MURMURA, E., SAVELLI, E. (2020). Factors affecting household food waste among young consumers and actions to prevent it. A comparison among UK, Spain and Italy. *Resources, Conservation & Recycling*. 153, 104586. ISSN 0921-3449. https://doi.org/10.1016/j.resconrec.2019.104586

CORSINI, F., ANNESI, N., ANNUNZIATA, E., FREY, M. 2023. Exploring success factors in food waste prevention initiatives of retailers: the critical role of digital technologies. *British Food Journal*. 126 (5): 1941–1957. https://doi.org/10.1108/BFJ-01-2023-0034

ČESKÝ STATISTICKÝ ÚŘAD. 2024. Životní podmínky domácností — EU-SILC. *ČSÚ*. [Accessed 2025-03-05] https://csu.gov.cz/zivotni-podminky-domacnosti-eusilc

EIČAITÉ, O. *et al*. 2022. Food waste in the retail sector: A survey-based evidence from Central and Eastern Europe. *Journal of Retailing and Consumer Services*. 69. ISSN 0969-6989. https://doi.org/10.1016/j.jretconser.2022.103116

EUROSTAT. 2024. Food waste and food waste prevention — estimates. *Eurostat*. [Accessed: 2025-05-09] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Food_waste_and_food_waste_prevention_-_estimates

GONZÁLES-SANTANA, R. A. *et al*. 2022. Dimensions of household food waste focused on family and consumers. *Critical Reviews in Food Science and Nutrition*. 62(9), 2342–2354. ISSN 1040-8398. https://doi.org/10.1080/10408398.2020.1853033

GUPTA, S. 2016. Efficacy of Promotional Offers in Poor Households: Insights from the Bottom of the Pyramid. In: KIM, K. K. (ed.). *Celebrating America's Pastimes: Baseball, Hot Dogs, Apple Pie and Marketing?* Cham: Springer International Publishing, pp. 805–806. ISBN 978-3-319-26646-6. https://doi.org/10.1007/978-3-319-26647-3_173

HUANG, I. Y. *et al*. 2021. Food waste management: A review of retailers' business practices and their implications for sustainable value. *Journal of Cleaner Production*. 285. ISSN 0959-6526. https://doi.org/10.1016/j.jclepro.2020.125484

KORMAŇÁKOVÁ, M., REMEŠOVÁ, M., VANČOVÁ, T. 2021. Food waste in municipal mixed waste produced at household level: empirical evidence from the Czech Republic. *Journal of Material Cycles and Waste Management*. 23(4), 1348–1364. ISSN 1438-4957. https://doi.org/10.1007/s10163-021-01214-7

KOTLER, P., KELLER, K. L. 2013. *Marketing management*. Praha: Grada. ISBN 978-80-247-4150-5.

KUBÍČKOVÁ, L., VESELÁ, L., KORMAŇÁKOVÁ, M. 2021. Food Waste Behaviour at the Consumer Level: Pilot Study on Czech Private Households. *Sustainability*. 13(20). ISSN 2071-1050. https://doi.org/10.3390/su132011311

KUBÍČKOVÁ, L., VESELÁ, L., MORÁVKOVÁ, M., SYROVÁTKA, P. 2022. *Návrh postupu pro účinnou komunikační kampaň v oblasti redukce plýtvání potravinami v domácnostech*. Project TAČR č. TL02000092. [Accessed 2025-04-15] https://neplytvejpotravinami.mendelu.cz/wp-content/uploads/2022/06/Metodika_Vystup-projektu_final.pdf

NOVÁKOVÁ, P., HÁK, T., JANOUŠKOVÁ, S. 2021. An Analysis of Food Waste in Czech Households — A Contribution to the International Reporting Effort. *Foods*. 10(4). ISSN 2304-8158. https://doi.org/10.3390/foods10040875

QI, D., ROE, B. B., VAN RIJNSOEVER, F. 2016. Household Food Waste: Multivariate Regression and Principal Components Analyses of Awareness and Attitudes among U.S. Consumers. *PLOS ONE*. 11(7). ISSN 1932-6203. https://doi.org/10.1371/journal.pone.0159250

SCHANES, K., DOBERING, K., GÖZET, B. 2018. Food waste matters — A systematic review of household food waste practices and their policy implications. *Journal of Cleaner Production*. 978–991. ISSN 0959-6526. https://doi.org/10.1016/j.jclepro.2018.02.030

TSALIS, G., BOUPTRUP JENSEN, B., ASCHEMANN-WITZEL, J. 2024. The relationship between retail price promotions and household-level food waste: Busting the myth with behavioural data? *Waste Management*. 173, 29–39. ISSN 0956-053X. https://doi.org/10.1016/j.wasman.2023.10.032

Contact information

Veronika Antošová: e-mail: veronika.antosova@mendelu.cz

CONTROL OF INTEGRATING SYSTEMS USING A TWO-DEGREE-OF-FREEDOM PID CONTROLLER

Jakub Meduna[1], Radim Farana[2]

[1]Department of Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

[2]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_022

ABSTRACT

This article deals with the comparison of a two-degree-of-freedom (2DOF) PID controller and the classical 1DOF PID controller for the control of integrating systems. The work responds to the structural limitation of the 1DOF controller, which cannot simultaneously optimize setpoint tracking and disturbance rejection. Three typical integrating plants (with transport delay, oscillatory, non-oscillatory) and four controller variants tuned by the SIMC method and the PID Tuner application were tested in the MATLAB Simulink environment. Control quality was assessed using the IAE and ITAE criteria. The serial 2DOF controller improved ITAE by 28 to 54% compared with the best 1DOF solution; the optimal tuning method, however, depends on the plant structure.

Keywords: integrating systems, 2DOF PID, SIMC, PID Tuner, IAE, ITAE
JEL Code: C88

1 INTRODUCTION

Systems with integrating behaviour belong among the more difficult control tasks. Unlike proportional plants they have no self-regulating capability: once disturbed from equilibrium, the output keeps rising or falling until the controller intervenes. Typical examples include the liquid level in a tank without outflow, motor position control or temperature regulation in a greenhouse, where heat losses fully determine the energy balance.

Industrial practice relies on the PID controller with one degree of freedom (PID 1DOF), which takes the control error **e=w-y** as the only input. This structure has one inherent limitation: tuning the controller for fast response to a setpoint change yields a high gain and consequently high sensitivity to disturbances — and vice versa. Aström and Hägglund (1995) labelled this trade-off as the fundamental limitation of single-input controllers. The aim of the present work is to verify experimentally whether PID 2DOF removes this trade-off and, if so, which tuning method should be selected (Sladká 2020).

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1.1 PID 2DOF and its advantages

PID 2DOF has structurally two parallel paths: an input filter for the setpoint w with weights b and c for the proportional and derivative term, and a classical PID channel for the control error e (Šulc and Vítečková 2004; Wang 2020). The serial variant proposed by Vítečková and Víteček (2011) introduces the discrete transfer functions of the input filter $G_F(z)$ and of the controller $G_R(z)$:

$$G_F(z) = \frac{A_g^{(1)} z^2 - B_g^{(1)} z + C_g^{(1)}}{A_g z^2 - B_g z + C_g}, \quad G_R(z) = \frac{K_p (A_g z^2 - B_g z + C_g)}{Z_2 z^2 - Z_1 z + Z_0} \tag{1}$$

This structure allows the tracking mode (through the filter) and the disturbance-rejection mode (through the controller) to be tuned independently. The dissertation by Sladká (2020) demonstrated the applicability of 2DOF on typical industrial processes; the open question remained the behaviour on integrating plants, where long-term disturbance rejection is usually the dominant task.

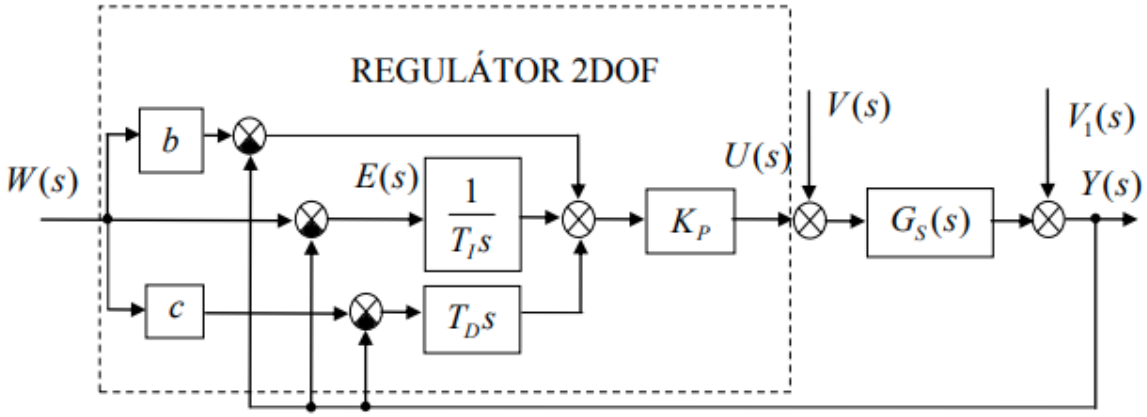


Fig. 1: Block diagram of the control loop with an analog PID 2DOF controller. The set-point *W(s)* is weighted by the coefficients *b* (proportional) and *c* (derivative), the control error *E(s)* enters the integral path, and the resulting actuating signal *U(s)* drives the plant $G_s(s)$ with additive disturbance *V(s)* (adopted from Meduna 2024).

1.2 Tuning methods

Hundreds of empirical rules exist for 1DOF controllers (O'Dwyer 2009). The present work compares two methods used in practice. The SIMC (Simple Internal Model Control) method derives analytical expressions for the parameters K_p, T_I, T_D from the internal model of the plant with a single tuning parameter T_W (Víteček and Vítečková 2011). PID Tuner from the MATLAB Control System Toolbox searches the parameter space automatically; for the 2DOF case it adds parameters T_o (time response) and T_b (robustness) (Mathworks 2024).

2 METHODOLOGY

Simulation circuits were built in MATLAB Simulink with a sampling period of T=0.1 s for three typical integrating plants.

Integrating plant with transport delay $(T_s = 1 \cdot \mathrm{s}, T_1 = 12 \cdot \mathrm{s}, T_d = 4 \cdot \mathrm{s})$:

$$G_{s1}\left(s\right) = \frac{1}{12s^2+s} e^{-4s}. \tag{2}$$

In agriculture this corresponds, for example, to material transport that introduces a noticeable delay between actuator action and output response.

Oscillatory integrating plant $(T_s = 1\mathrm{s}, T_0 = 5\mathrm{s}, \xi = 0.8)$:

$$G_{s2}\left(s\right) = \frac{1}{25s^3+8s^2+s}. \tag{3}$$

Non-oscillatory integrating plant $(T_s = 1\mathrm{s}, T_1 = 10\mathrm{s}, T_2 = 6\mathrm{s})$:

$$G_{s3}\left(s\right) = \frac{1}{60s^3+16s^2+s}. \tag{4}$$

Four controllers were run in parallel for each plant: PID 1DOF (SIMC and PID Tuner) and serial PID 2DOF (SIMC and PID Tuner). A signal shaper acting as a D/A converter followed the controller output. A disturbance *v(t)* entered the loop in front of the plant as a step change at *t* = 200 s. To test robustness, one time constant was changed at *t* = 250 s (for the transport-delay plant T_1 from 12 s to 14 s), simulating plant ageing.

Control quality was assessed by four metrics: settling time with 1% tolerance, overshoot, and the integral criteria IAE and ITAE:

$$I_{IAE} = \int_0^\infty |e(t)|dt, \qquad I_{ITAE} = \int_0^\infty t\,|e(t)|dt. \tag{4}$$

IAE penalises the absolute magnitude of the control error, ITAE weights later errors more heavily.

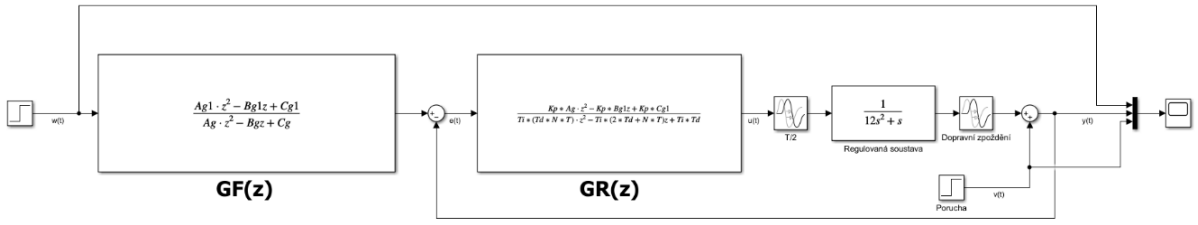


Fig. 2: Schematic of the simulation circuit with a serial PID 2DOF controller realised by transfer functions $G_F(z)$ and $G_R(z)$ in MATLAB Simulink (adopted from Meduna 2024).

3 RESULTS

Tables 1 to 3 summarise the results for a setpoint step at $t = 0$ s; the best variant for each plant is shown in bold.

The serial 2DOF controller won across all three plants. For the transport-delay and non-oscillatory plants the best method was SIMC, for the oscillatory plant PID Tuner. The robustness test under parameter change, however, yielded a different picture for the transport-delay plant: after the change of T_1 from 12 s to 14 s the most stable result came from PID 1DOF tuned by PID Tuner (settling 75.4 s vs 105.8 s for SIMC 2DOF), while the 2DOF controllers lost 50 to 130 % of their original speed. For the oscillatory plant PID Tuner 2DOF held the leading position even after the parameter change (ITAE 180 vs 285 for 1DOF Tuner) (Meduna, 2026).

Tab. 1 Plant with transport delay

Tuning method	Settling [s]	Overshoot [%]	IAE	ITAE[s]
PID Tuner 1DOF	83.42	33.9	14.12	253.11
SIMC 1DOF	94.89	29.4	15.70	315.52
PID Tuner 2DOF	77.38	8.7	17.86	228.27
SIMC 2DOF	**45.40**	**11.9**	**10.95**	**147.73**

Tab. 2 Oscillatory integrating plant

Tuning method	Settling [s]	Overshoot [%]	IAE	ITAE[s]
PID Tuner 1DOF	96.79	28.2	12.34	242.50
SIMC 1DOF	121.22	21.4	17.38	473.08
PID Tuner 2DOF	**64.71**	**4.7**	**11.95**	**111.49**
SIMC 2DOF	96.38	7.7	14.68	363.85

Tab. 3 Non-oscillatory integrating plant

Tuning method	Settling [s]	Overshoot [%]	IAE	ITAE[s]
PID Tuner 1DOF	124.61	25.9	16.73	451.93
SIMC 1DOF	140.23	27.0	22.56	699.64
PID Tuner 2DOF	89.79	11.2	20.28	319.73
SIMC 2DOF	**62.66**	**8.4**	**15.62**	**324.95**

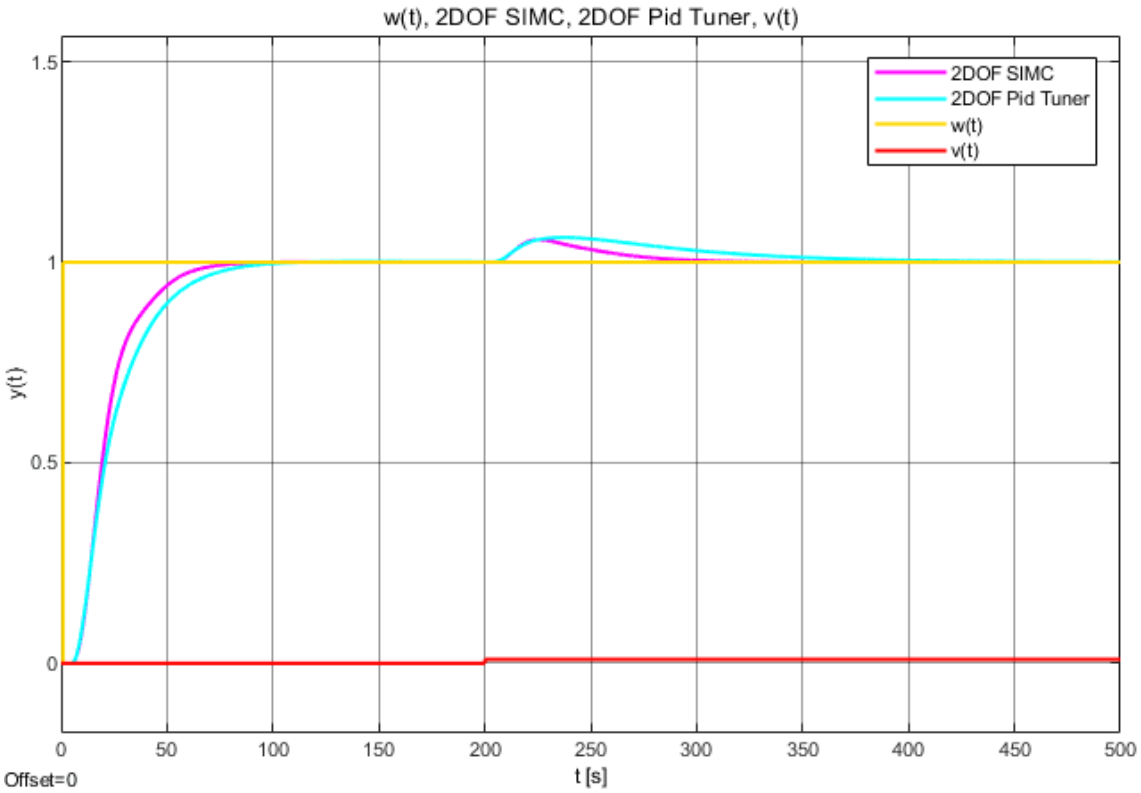


Fig. 3: Control response for the integrating plant with transport delay — comparison of PID 2DOF tuned by SIMC against PID Tuner. The disturbance step $v(t)$ enters at $t = 200$ s (adopted from Meduna 2024).

4 DISCUSSION

The structural advantage of the 2DOF controller was confirmed quantitatively: overshoot dropped from 25–35 % to 5–12% and settling time shortened by 25–50% across all three plants. The reason is the independent tuning of the tracking mode through the input filter G_F and of the disturbance--rejection mode through the controller G_R. A 1DOF controller must choose a higher gain to achieve the same disturbance rejection, which yields an oscillatory response to a setpoint change.

The choice of tuning method depends on plant structure. SIMC produces deterministic, reproducible results whenever an analytical formula exists for the given structure; for the oscillatory integrating plant the formula has to be approximated from a 1st-order template instead of the original 2nd-order one, which explains why PID Tuner outperformed it in this case. The practical rule reads: if SIMC covers the structure, use it as the primary method; otherwise reach for PID Tuner.

Robustness comes at a price. A 2DOF SIMC controller tuned precisely to specific dynamics deteriorated more after a 17% change of T_1 than a less precisely tuned 1DOF Tuner. For processes with substantial model uncertainty (such as agricultural greenhouses with seasonal parameter drift), the more conservatively tuned 1DOF or a 2DOF with an adaptive identification mechanism is preferable.

5 CONCLUSION

The work compared four PID controller variants on three typical integrating plants in MATLAB Simulink. For integrating systems, the serial PID 2DOF is structurally the better choice than the classical PID 1DOF: across all tested plants, it improved ITAE by 28 to 54% and reduced overshoot fivefold. The choice of tuning method, however, depends on the structure — SIMC for transport--delay and non-oscillatory plants, PID Tuner for oscillatory ones. Robustness against parameter changes is lower for 2DOF than for a conservatively tuned 1DOF, which represents the price for precision. Future extensions may include verification of 2DOF on non-linear plants with actuator saturation, adaptive tuning from a continuously identified model, and experimental verification on a physical system such as DC motor position control.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

REFERENCES

ASTRÖM, K. J., HÄGGLUND, T. 1995. *PID Controllers: Theory, Design, and Tuning*. 2nd ed. Research Triangle Park, NC: Instrument Society of America. ISBN 1-55617-516-7

MATHWORKS. 2024. *PID Controller Tuner App*. https://www.mathworks.com/help/control/getstart/tune-pid-controller-to-balance-tracking-and-disturbance-rejection.html

MEDUNA, J. 2026. *Control of integrating systems using a two-degree-of-freedom PID controller* [in Czech]. Master's thesis, head: Farana, R. Brno: Mendel University in Brno.

O'DWYER, A. 2009. *Handbook of PI and PID Controller Tuning Rules*. 3rd ed. London: Imperial College Press. ISBN 978-1-84816-242-6

SLADKÁ, K. 2020. *Tuning of Digital and Analog Two-Degree-of-Freedom Controllers* [in Czech]. PhD Dissertation. Ostrava: VŠB – Technical University of Ostrava, Faculty of Mechanical Engineering.

ŠULC, B., VÍTEČKOVÁ, M. 2004. *Theory and Practice of Control Loop Design* [in Czech]. Prague: Czech Technical University Publishing House. ISBN 80-01-03007-5

VÍTEČEK, A., VÍTEČKOVÁ, M. 2011. *Selected Methods of Controller Tuning* [in Czech]. Ostrava: VŠB – Technical University of Ostrava. ISBN 978-80-248-2503-8

WANG, L. 2020. *PID Control System Design and Automatic Tuning using MATLAB/Simulink*. London: John Wiley & Sons. ISBN 978-1-119-46934-6

Contact information

Jakub Meduna: e-mail: jakubmeduna390@seznam.cz

Radim Farana: e-mail: radim.farana@mendelu.cz

PRACTICAL BLOCKCHAIN DEPLOYMENT FOR TICKET MANAGEMENT: A CASE STUDY WITH SOLIDITY SMART CONTRACTS

Jan Mrákava[1], Radim Farana[2]

[1]Department of Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

[2]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_023

ABSTRACT

This article deals with the design and implementation of a web application for ticket management based on a local Ethereum-compatible blockchain. The work responds to the gap between marketing claims about blockchain as a universal document verification tool and limited knowledge of real transaction costs. The frontend uses React.js, the backend Node.js with Express and MongoDB, and the logic consists of two Solidity 0.8.10 smart contracts (EventContract, TicketContract) deployed through Truffle into the Ganache environment (Chain ID 1337). Measurements showed that creating a single event with a typical description costs 0.0078 ETH (~546 CZK at the 2024 exchange rate), while ticket purchase and resale require an order of magnitude lower fees. The main finding: transaction cost is dominated by the length of text strings, which for document verification in general means storing only hashes and short identifiers on the chain.

Keywords: blockchain, Ethereum, smart contracts, Solidity, document verification, gas cost
JEL Code: O33

1 INTRODUCTION

Two widely reported cases illustrate the same problem. In 2017, tickets for the Rammstein concert in Prague reappeared in secondary sale at several times the original price, and the organizer had no way to verify who had actually sold what to whom. In 2019, the forensic expert Jiří Závora presented a forged document to 66 offices in the Czech Republic; in 51 cases the officials accepted it as authentic. Both cases share the same root — there is no single source of truth about which document (a ticket, a diploma, a power of attorney) is genuine and to whom it currently belongs.

Blockchain — a distributed and cryptographically secured registry — appears as a candidate solution thanks to record immutability and smart contracts for automation. The aim of this work was to build a functional prototype that demonstrates this approach on a concrete use case (ticket management), to measure real transaction costs, and to generalize the results for document verification at large.

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

2 BUILDING BLOCKS AND CONSENSUS

A cryptographically secured chain of timestamps for digital documents was described by Haber and Stornetta (1991); Merkle (1988) added a tree structure of hashes for efficient integrity verification. These concepts were combined by Nakamoto (2008) in the Bitcoin whitepaper. To reach consensus about network state, various algorithms are used: Proof-of-Work (high security at the cost of energy consumption comparable to the annual demand of Pakistan according to CCAF 2023), Proof-of-Stake (energy efficient), and Proof-of-Authority (for private networks with validated identity of validators, highly scalable). Ganache uses a simplified PoA variant — confirmation latency is eliminated and transactions confirm within milliseconds.

3 EXISTING DOCUMENT VERIFICATION SYSTEMS

Academic reviews (Catalini and Gans 2017; Houben and Snyers 2018) and commercial systems confirm the same pattern — the document is hashed and only the digest with metadata is stored on the chain. Dock uses decentralized identifiers DID with Issuer/Holder/Verifier roles. The Czech Certoo uses the Litecoin blockchain with a maximum file size of 500 MB. Diplomachain cooperates with the Czech Ministry of Industry and Trade, the Chamber of Commerce, CTU Prague and TBU Zlín. Blockademia runs on Cardano with the ACI token. None of these systems stores entire documents on the chain — only the hash, identifier, and metadata, which ensures a constant transaction cost regardless of document size.

4 METHODOLOGY

4.1 Technology Stack

For the prototype, a combination enabling rapid development and measurement of real transaction costs was chosen: React.js (frontend), Node.js with Express.js and MongoDB (backend with user data), Ganache (local development blockchain with ten preset accounts of 100 fictitious ETH each), Truffle (compilation, migration, and contract testing), Solidity 0.8.10 (with built-in overflow protection), and MetaMask (browser wallet). Ganache listens on `127.0.0.1:8545` with Chain ID 1337; MetaMask is configured with the network name “TicketBlock”.

4.2 Two Smart Contracts

The application architecture separates event management and ticket management into two independent contracts. EventContract owns the Event structure with the `eventID` field (bytes32), name, date, ticket count, price, description, category, and owner address; the key function `buyTicket(eventID)` delegates the call to TicketContract. TicketContract owns the `Ticket` structure with `ticketID` (bytes32), `eventID`, owner, purchase date, and flags `isRedeemed`, `isValid`, and `forSale`; key functions include `createNewTicket`, `transferTicket`, `setTicketForSale`, and `verifyTicket`. The separation has a practical reason — events and tickets have different lifecycles, which would lead to more complex logic and higher gas overhead in a monolithic contract.

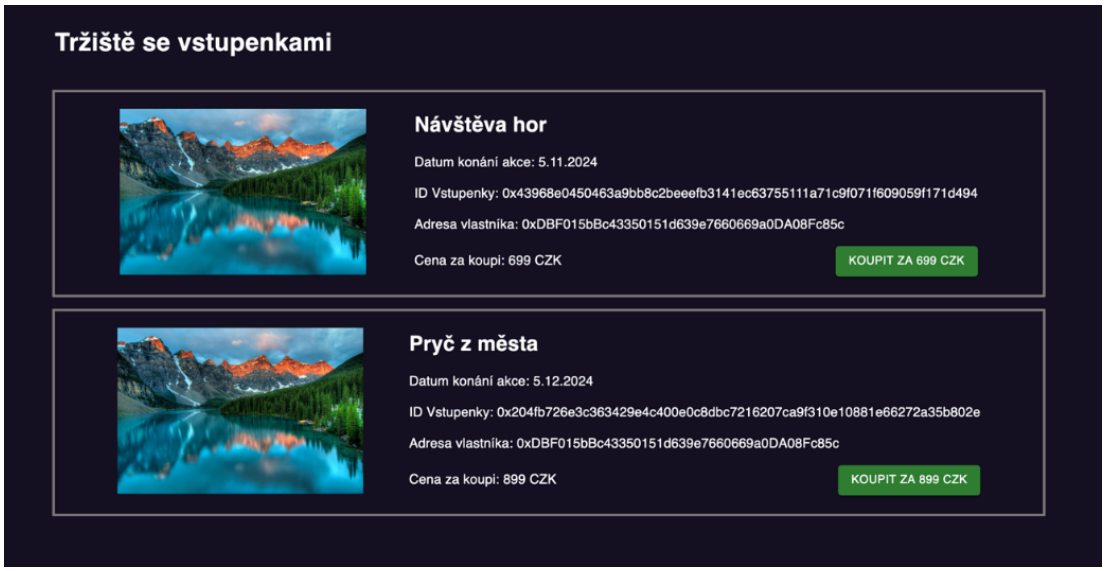


Fig. 1: Resale marketplace in the web application — each entry shows the event name, on-chain event ID, current owner address (Ethereum account) and asking price in CZK, which is converted to ETH/wei at confirmation time.

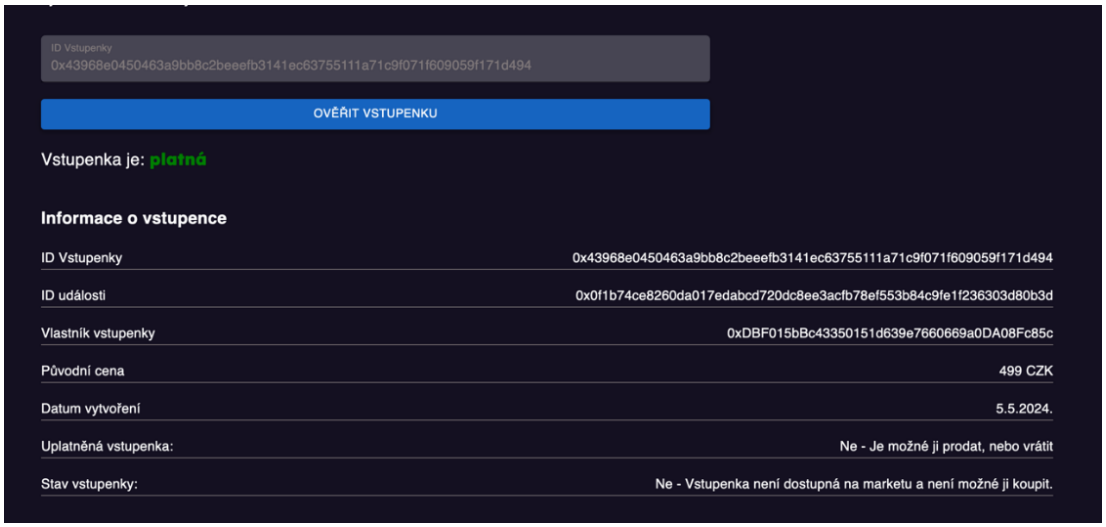


Fig. 2: Result of ticket verification after entering a valid ID — the system immediately returns the status (“valid”), event ID, owner address, original price, creation date and redemption status. Verification is a view function that consumes no gas and can be invoked by anyone.

4.3 Identifier Generation and Audit Trail

Identifiers are built by hashing the combination of attributes using the Keccak-256 function, which produces a 32-byte output (bytes32):

(1)

„eventID“=“keccak256“ („abi.encodePacked“ („eventName“ ,“block.timestamp“)).

(2)

„ticketID“=“keccak256“ („abi.encodePacked“ („eventID“ ,“userAddress“ ,“block.timestamp“))

The collision risk at this size is negligible; before writing into the map, `require(!eventExists[_eventID])` is checked, which would formally catch a collision should it occur. Each contract emits structured events permanently written into the blockchain as an audit trail — TicketContract emits ten event types (`TicketCreated`, `TicketBought`, `TicketTransferred`, `TicketRedeemed`, and others). The frontend obtains the history by calling `getPastEvents(‚allEvents‘, { filter: { ticketID }, fromBlock: 0, toBlock: ‚latest‘ })`.

4.4 Gas Cost Measurement

After truffle migrate, the addresses of deployed contracts and their ABI are generated. Web3.js initializes the blockchain connection; state-changing functions use `.send({ from: userAddress, gas: ‚500000‘ })`, view functions only `.call()`. For each key operation (event creation with various description lengths, purchase, resale, verification), gas usage, gas price, and total cost in ETH and CZK were measured according to the Coinbase 2024 exchange rate.

5 RESULTS

5.1 Transaction Costs

For the full workflow (event creation → purchase → resale → verification), the values in Tab. 1 were measured. The cost of event creation is dominated by description length, while purchase and resale have a relatively constant cost given by hashing and map update. In Solidity, each additional byte of a string represents about 17 gas units, which scales the cost linearly with the volume of text data (Mrákava, 2026).

Tab. 1 Measured transaction costs (Ganache, gas price 2.5 gwei, Coinbase 2024 exchange rate)

Operation	Cost (ETH)	Cost (CZK)	Note
Event creation — short description (30 chars)	~0.002	~140	minimal payload
Event creation — typical description (200 chars)	0.0078	~546	gas usage ~3,305,129
Event creation — long description (500 chars)	~0.015	~1,050	festival program
Ticket purchase	~0.001	~70	no payload
Ticket resale	~0.001	~70	no payload
Verification (view)	0	0	off-chain

5.2 Stability and Auditability

The application was tested on 30 tickets across 5 events; all transactions completed successfully and no identifier collision occurred. In a simulation of two concurrent purchases of the last available ticket, the second purchase reverted with "Not enough tickets left". For a randomly selected ticket, four history records were retrieved through `getEventsForTicketID` (creation, two transfers, redemption) including transaction hashes, block hashes, and timestamps. Third-party verification without access to the UI was possible through a Web3.js call directly to Ganache — auditability without trusting the system operator is the main value blockchain adds over a centralized solution.

6 DISCUSSION

The measurements revealed a practical boundary for economic deployment: the size of data stored on chain must be as small as possible. For tickets this was not a problem — ID, owner, and state flags fit into tens of bytes. For event creation with a long description, the cost rises sharply because each byte of text is expensive on Ethereum. For document verification in general, a practical pattern emerges: the document is stored off chain (IPFS, server, archive), its hash (SHA-256 or keccak-256) is computed, and only the hash, identifier, issuer, and timestamp are stored on chain. Verification then means comparing the hash from the file with the hash on the chain.

Economics matters more than technology. The cost of 546 CZK per event may seem negligible, but for an organizer with 1,000 events per year this means 546,000 CZK only for the blockchain, and at tickets priced 50-200 CZK the viability is questionable. For high-value documents with long validity (diplomas), the situation differs — a one-off cost against the document's value is reasonable, which explains the interest of CTU and TBU in Diplomachain. For production deployment, Polygon would be worth considering, as it preserves the Ethereum API while reducing fees by an order of magnitude. Permanent on-chain storage also conflicts with the right to be forgotten under GDPR; Czech law does not explicitly regulate blockchain (Hospodka 2020) and no Czech court has officially accepted blockchain as evidence (Hanák 2022).

7 CONCLUSION

The work showed that a functional blockchain application for ticket management can be built with a common stack (React, Node.js, MongoDB, Truffle, Solidity) in a reasonable time, and measured specific transaction costs that marketing materials rarely disclose. The cost of event creation ranges from ~140 CZK for a short description to ~1,050 CZK for a long one — text dominates the gas cost. Architectural separation into two contracts (EventContract, TicketContract) reduces overhead in repeated operations and simplifies code audit. For document verification at large the rule holds: store only hashes and metadata on chain, keep content off chain. Future extensions include deploying an identical contract on Polygon to measure real transaction costs under competitive operation and implementing a combined on-chain hash plus off-chain payload mechanism with deletion option for GDPR compliance.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

REFERENCES

CATALINI, C., GANS, J. S. 2017. *Some Simple Economics of the Blockchain*. NBER Working Paper No. 22952. National Bureau of Economic Research. https://doi.org/10.3386/w22952

CAMBRIDGE CENTRE FOR ALTERNATIVE FINANCE (CCAF). 2023. Cambridge Bitcoin Electricity Consumption Index. *CCAF*. https://ccaf.io/cbnsi/cbeci

COINBASE. 2024. ETH/CZK: Convert Ethereum (ETH) to Czech Koruna (CZK). *Coinbase*. https://www.coinbase.com/converter/eth/czk

HABER, S. and STORNETTA, W. S. 1991. How to Time-Stamp a Digital Document. *Journal of Cryptology*. 3(2), 99-111. ISSN 0933-2790. https://doi.org/10.1007/BF00196791

HANÁK, O. 2022. What Use Does Blockchain Have in Law? [in Czech]. *Kroupa & Helan*. https://kroupahelan.cz/blog/jake-vyuziti-ma-blockchain-v-pravu/136

HOSPODKA, J. 2020. *Legal Aspects of Cryptocurrencies* [in Czech]. Master's thesis. Prague: Charles University.

HOUBEN, R., SNYERS, A. 2018. *Cryptocurrencies and Blockchain: Legal Context and Implications for Financial Crime, Money Laundering and Tax Evasion*. Brussels: European Parliament TAX3 Study.

MERKLE, R. C. 1988. A Digital Signature Based on a Conventional Encryption Function. In: *Advances in Cryptology — CRYPTO '87*. Berlin: Springer, pp. 369-378. ISBN 978-3-540-18796-7. https://doi.org/10.1007/3-540-48184-2_32

MRÁKAVA, J. 2026. *Practical Blockchain Deployment for Ticket Management with Smart Contracts in Solidity* [in Czech]. Master's thesis, head: FARANA, R. Brno: Mendel University in Brno.

NAKAMOTO, S. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*. https://bitcoin.org/bitcoin.pdf

ZHAI, X., PANG, S., WANG, M., QIAO, S., LV, S. 2021. TVS: A Trusted Verification Scheme for Office Documents Based on Blockchain. *Complex & Intelligent Systems*. 8, 3941-3954. ISSN 2199-4536. https://doi.org/10.1007/s40747-021-00617-1

Contact information

Jan Mrákava: e-mail: janmrakava@gmail.com

Radim Farana: e-mail: radim.farana@mendelu.cz

GREENWASHING IN THE MARKETING COMMUNICATION OF SUSTAINABLE FASHION COMPANIES: THE ATTITUDES OF GENERATION Y AND Z

Gabriela Navrátilová[1], Veronika Antošová[1], Lucie Veselá[1]

[1]Department of Marketing and Trade, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_024

ABSTRACT

The paper deals with the perception of the credibility of environmental claims of fash-ion brands among Czech consumers of generation Y and Z and the possibilities of communication that minimizes the risk of greenwashing. The aim is to identify the key factors influencing the purchasing behaviour of these generations in the sustainable fashion market and to describe their attitudes towards the credibility of sustainability marketing messages. A combination of situational analysis of secondary data and a primary questionnaire survey in the period March–April 2024 is used. The sample con-sists of 368 respondents born between 1980 and 2010 selected according to their age. Dependencies are tested by Pearson's chi-square independence test at a significance level of 5%, the intensity of association is measured by Cramér's V. It was found that 68% of respondents are more likely to distrust marketing claims about the sustainabil-ity of fashion brands, and 65.5% say they would not buy from a brand convicted of greenwashing. Confidence in environmental claims is not statistically dependent on the age of the respondent or the frequency of purchases of slow fashion, but it is weakly but statistically significantly dependent on the knowledge of the concept of greenwash-ing and on the frequency of its feeling when buying clothes. Transparent communica-tion about the origin of materials and production processes increases consumer trust for 58.7% of respondents. The results support the conclusion that greenwashing risk reduction is based primarily on the transparency, authenticity and traceability of the message, not on the intensity of the communication itself.

Keywords: greenwashing, sustainable fashion, slow fashion, marketing communication, generation y, generation z, consumer trust, transparency
JEL Code: M31, M37, Q56

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Sustainability has become an important part of corporate strategies and consumer decisions in recent years. In response to the growing demand for environmentally friendly products, companies are adapting their products, production processes and marketing communications to reflect environmental issues. However, when these claims are unfounded, simplistic or misleading, greenwashing occurs, which is defined in the literature as the dissemination of misleading or false information about a company's environmental activities or about the environmental benefits of its products and services (de Freitas Netto *et al.*, 2020). This phenomenon undermines consumer trust, damages brand reputations, and weakens the position of honest producers, whose claims are lost in a flood of generic phrases such as „sustainable" or „natural."

The fashion industry is one of the sectors where the issue of greenwashing is particularly serious. The global fashion industry is responsible for 4–10% of global greenhouse gas emissions from human activities, and the entire value chain is characterised by high water consumption, pollution of water resources and textile waste production (Niinimäki *et al.*, 2020). Almost 60% of all newly produced garments are thrown away within one year of production, and textile materials make up a substantial part of mixed waste, making fashion one of the most unsustainable consumer segments. Fast fashion builds on low prices, short production cycles and frequent changes of collections, which motivates impulsive purchases and short lifespans of clothes. An alternative is the concept of slow fashion, which emphasizes quality, production ethics and a longer product life cycle (Jung and Jin, 2014; Štefko and Steffek, 2018). In response to the increased demand for sustainable products, some fast fashion brands have expanded their offerings to include so-called sustainable collections, but their credibility is often questioned in connection with greenwashing. In this context, there is a growing focus on how brands communicate their environmental commitments and how consumers respond to such communications.

Generation Y (born approximately 1980–1995) and Generation Z (born approximately 1996–2010) represent segments with high sensitivity to environmental issues and corporate social responsibility. Generation Z is often referred to as the „green generation" — sustainability influences their purchasing decisions, prefers manufacturers with responsible strategies, and actively monitors the environmental impacts of their behavior (Wawer *et al.*, 2022). Generation Y is already fully involved in the market in terms of purchasing power, and both generations account for a growing share of the sustainable clothing market. At the same time, skepticism about environmental claims is increasing: consumers are increasingly critical of general „green" formulations and are more likely to verify that real brand practices correspond to them (Samal and Bhalala, 2023). Adamkiewicz *et al.* (2022) confirm that reducing greenwashing, together with increasing the transparency of marketing communications, significantly contributes to strengthening trust in green products.

On the regulatory level, the European Commission responded to the expansion of greenwashing in 2023 with a proposal for a directive on green claims, which aims to increase the reliability, comparability and verifiability of environmental communications (European Commission, c1995-2023). At the same time, however, there is a lack of primary data on how generations Y and Z in the Czech Republic perceive the credibility of fashion brands' marketing claims and how these perceptions are related to their knowledge of greenwashing and experience with it. International studies suggest that familiarity with greenwashing leads to higher skepticism and a departure from the brands in which it has been detected (Nosto, 2022; Zhang *et al.*, 2018), but the Czech

empirical evidence in the fashion segment remains limited. Existing domestic studies focus mainly on general consumer attitudes towards sustainability, not on the relationship between knowledge of greenwashing, the perceived credibility of environmental claims and subsequent purchasing decisions. This post tries to fill that gap.

The aim of the paper is to identify the key factors influencing the purchasing behavior of generations Y and Z on the Czech fashion market and to evaluate their attitudes towards the credibility of marketing claims about the sustainability of products and brands. In particular, attention is paid to whether trust in environmental claims depends on age, on the frequency of purchases of slow fashion, on the knowledge of the concept of greenwashing and on the frequency of its feeling when buying clothes. Based on these findings, communication principles for sustainable fashion businesses are formulated that minimize the risk of greenwashing.

2 METHODOLOGY AND DATA

To fulfil the aim of the paper, a combination of situational analysis of secondary data and primary quantitative questionnaire survey was chosen. The situational analysis was based on freely available statistical sources, in particular data from the Czech Statistical Office on the age distribution of the population (CZSO, 2022), and provided a framework for defining target generations and their quota representation. The main empirical conclusions are based on the primary investigation.

The target population consists of Czech consumers born between 1980 and 2010, who correspond to generations Y and Z. To capture the internal heterogeneity of these cohorts, the population was divided into four age quotas: 2003–2010, 1996–2002, 1988–1995 and 1980–1987. The relative abundance of individual quotas in the basic population was determined according to the CZSO as of 31 January 2022 (see Table 1).

Data collection took place electronically via the Google Forms platform in the period from 25 March to 16 April 2024. The questionnaire was preceded by a pilot phase on two people from each age quota to identify ambiguities or errors in wording. The distribution took place via social networks Facebook, Instagram and the streaming platform Twitch, as well as by e-mail and in person at the youth home. A total of 376 responses were obtained, of which 368 valid observations remained after discarding incomplete or inconsistent records. In terms of representativeness, the sample in the 2003–2010 and 1988–1995 quotas is in very good agreement with the basic population; Larger deviations are evident in the 1996–2002 quota (7.04 percentage point higher) and in

Tab. 1 Structure of quota sampling by age — comparison of the population of the CZSO and the question-naire survey

Generations	Year of birth	Population of the Czech Republic (%)	Surveys (%)
Generation Z	2003–2010	23,76	25,00
Generation Z	1996–2002	18,23	25,27
Generation Y	1988–1995	27,35	26,63
Generation Y	1980–1987	30,66	23,10

Source: own processing according to the Czech Statistical Office (2022) and data from the questionnaire sur-vey (n = 368).

the 1980–1987 quota (7.56 percentage point lower).

The questionnaire contained a total of 22 questions structured into three thematic blocks. The first block (9 questions) examined the knowledge of the concept of greenwashing, the assessment of its severity, the frequency of its feeling when buying clothes, perceived suspicious practices and claims, and the impacts of detecting greenwashing on purchasing decisions and trust in the brand. The second block (4 questions) concerned shopping behaviour in relation to slow fashion and the importance of individual factors and aspects of sustainability in the selection of clothing. The third block (3 questions) focused on preferred communication channels, verification of sustainability claims and the importance of selected elements in sustainable fashion advertising. The conclusion of the questionnaire included identification questions. A seven-point scale was used to evaluate preferences and attitudes.

MS Excel and TIBCO Statistica 14 were used for statistical evaluation. The analysis included descriptive statistics and pivot tables for dependency testing. The existence of dependencies between variables was verified by Pearson's chi-square independence test at the chosen significance level of $\alpha = 0.05$ (Tahal, 2022). In the case of statistically significant dependence, the intensity of the association was measured by the Cramér coefficient V (Hendl, 2015), which takes values from the interval (0; 1), where a zero value means no dependence and a value close to one strong dependence.

Based on existing knowledge about the relationship of trust, greenwashing awareness and buying behavior (Jacobson and Harrison, 2022; Samal and Bhalala, 2023), four null hypotheses have been established:

- ***H01:*** *The degree of trust in sustainability marketing claims from fashion brands does not depend on the age of the consumer.*
- ***H02:*** *There is no relationship between the degree of trust in marketing claims from fashion brands and the frequency of purchases of slow fashion.*
- ***H03:*** *There is no correlation between the degree of trust in sustainability marketing claims made by fashion brands and the knowledge of the concept of greenwashing.*
- ***H04:*** *There is no correlation between the level of trust in marketing claims from fashion brands and the frequency of feeling greenwashed when buying clothes.*

3 RESULTS

The sample (n = 368) consists of 30.4% men, 68.5% women and 1.1% persons of other gender identity. In terms of economic status, the most numerous are employees and the self-employed (48.4%), students (42.4%) and people on maternity or parental leave (7.6%). The distribution across the four age quotas ranges from 23.1 to 26.6 %, ensuring comparability of findings across generations.

3.1 Awareness of greenwashing and perception of its seriousness

A total of 72.6% of respondents have heard of the term greenwashing, with 55.2% saying they know exactly what the term means. The youngest group of 14–21 years of age has the lowest knowledge. Greenwashing is perceived as a serious problem — 48% of respondents rate it on a scale of 1–7 at levels 6 and 7. Three-quarters of participants (69%) report that they encounter greenwashing when shopping either frequently (for about half of purchases) or rarely (less than half of purchases); 16.6% of respondents believe that they have never encountered it.

3.2 Trust in the environmental claims of fashion brands

A seven-point rating of trust in sustainability marketing claims from fashion brands (1 = no trust, 7 = full trust) shows significant skepticism: 68% of respondents chose values 1, 2 or 3, i.e. rather distrust. Less than 2% of respondents chose high trust values (6 and 7).

3.3 Hypothesis Testing Results

The relationship between confidence in environmental claims and selected variables was tested by a chi-square independence test. The null hypotheses H01 and H02 cannot be rejected because the corresponding p-values exceed the determined significance level of 5% (H01: $p = 0.24137$; H02: $p = 0.43007$). Therefore, it cannot be concluded that confidence in marketing claims about sustainability is statistically dependent on the age of the consumer or on the frequency of purchases of slow fashion. On the other hand, hypotheses H03 and H04 are rejected: for H03 a p-value of 0.00045 and a Cramér V = 0.2184 were found, for H04 a p-value of 0.04753 and a Cramér V = 0.1578. In both cases, it is a weak, statistically significant dependence. The pivot tables show that the better the respondent knows the concept of greenwashing and the more often they feel it when buying clothes, the less confidence they have in marketing claims about sustainability.

3.4 Implications of greenwashing for purchasing decisions and brand reputation

When asked directly whether a respondent would buy products from a brand that noticed greenwashing, 65.5% answered that they would rather or definitely not buy them. Repeated greenwashing is perceived as a very negative impact on brand reputation (levels 6 and 7) by 51.4% of

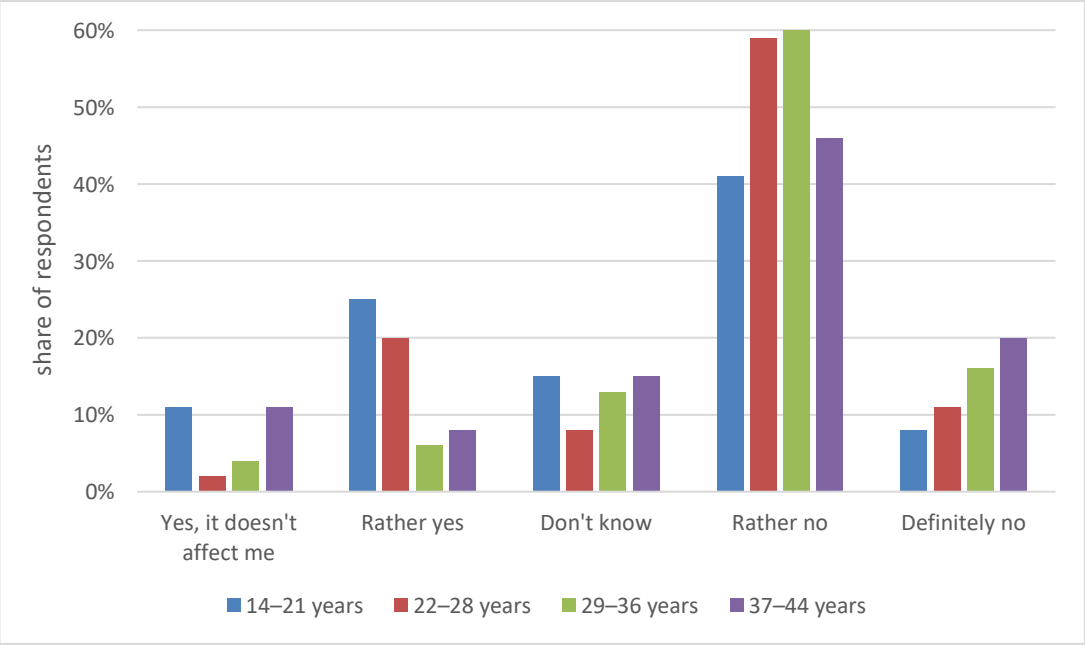


Fig. 1: Effect of greenwashing detection on the decision to purchase brand products by age group (n = 368).

respondents. A total of 35.3% of participants said that they had personally identified greenwashing in a brand they considered sustainable. Of this subgroup, 50.8% say that disclosure has fundamentally eroded their trust in the brand and reduced their willingness to endorse or recommend it in the future (levels 6 and 7). The distribution of responses by age group is illustrated in Figure 1.

3.5 Key factors when choosing clothes

The seven-point scale of importance is dominated by functional-qualitative factors: comfort (average 6.21), quality (5.99), size (5.95) and material (5.14). The price averages 5.09; sustainability 4.44; mark 2.98; Recommendations from influencers only 1.97. This preference structure shows that sustainability is not the top priority among the separately assessed factors, but it is significantly linked to the attributes that consumers put first — quality and durability. Among the aspects of sustainability, respondents consider quality and durability to be the most important (73.9% report values 6-7), natural materials and timeless design. The least importance is given to recycled materials and the general socially responsible behavior of the brand.

3.6 Perceived suspicious practices and claims

When evaluating specific marketing formulations typical of greenwashing, respondents perceive the use of unknown eco-certificates and general „green promises for the future“ (for example, reducing the carbon footprint by the following year) as the most suspicious. These findings suggest that non-specific aspirational formulations without a third party as a verifying actor are perceived as a red flag by consumers in this age group.

3.7 Communication channels and verification of claims

When shopping for sustainable fashion, respondents across age groups especially prefer social media, recommendations from friends and family, and retailer websites — these three channels dominate all four age subgroups. The younger segment of 14-21 years also adds recommendations from influencers as an important channel, while the oldest group of 37-44 years more often appreciates informational emails and newsletters. To verify sustainability claims, respondents most often use brand and certification websites; About a quarter of respondents (97 out of 368 people) also verify information on the brand‘s social networks and a slightly smaller share (71 people) with sustainable fashion influencers. Almost a third of respondents (30.2%) do not verify sustainability claims at all, with this share being highest in the youngest group of 14-21 years. The importance of selected elements in sustainable fashion advertising — information on materials, price, explanation of environmental terms, certificates and links to more detailed information — varies across age groups and is summarised in Figure 2.

Brand transparency has a strong effect on trust: 58.7% of respondents gave a score of 6 or 7 on a scale of 1-7 when asked whether transparent communication about the origin of materials and production processes increases their trust in a brand‘s sustainability claims.

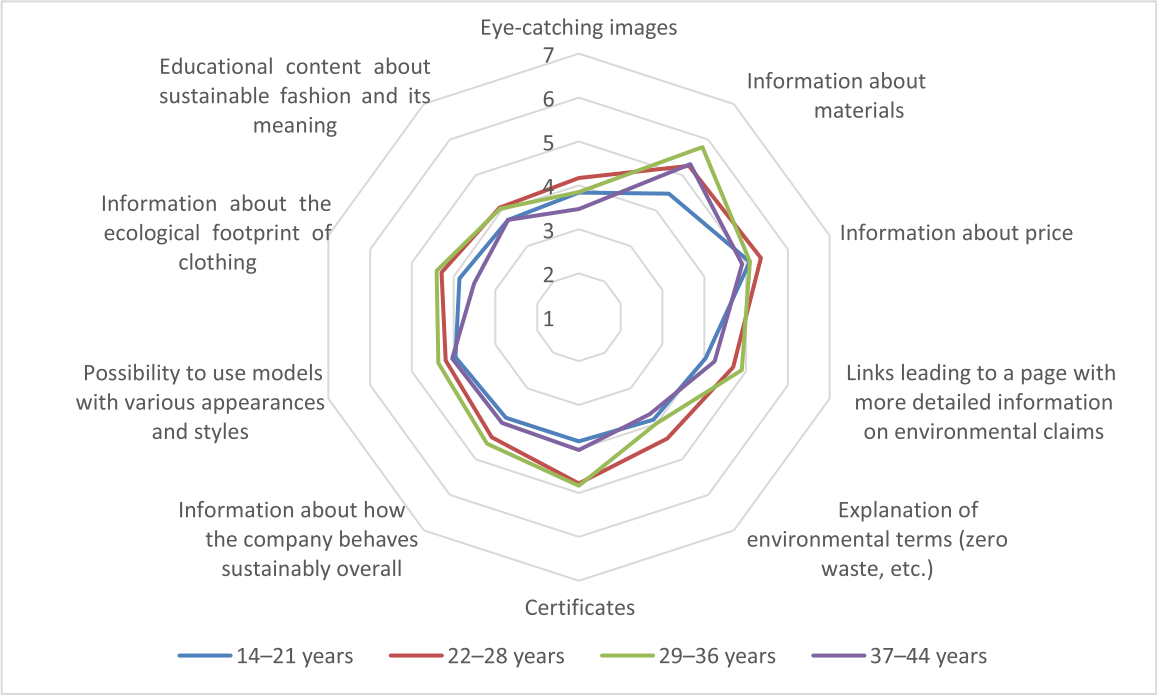


Fig. 2: Importance of individual elements in sustainable fashion advertising by age group (n = 368).

4 DISCUSSIONS AND CONCLUSIONS

The findings of the paper confirm that the Czech generation Y and Z are mostly sceptical about the environmental claims of fashion brands: 68% of respondents tend to distrust and 65.5% would not buy from a brand convicted of greenwashing. This finding is consistent with international evidence that familiarity with greenwashing increases skepticism about green claims and leads to a willingness to break the purchasing relationship with the problematic brand (Samal and Bhalala, 2023; Nosto, 2022). Also, the strong negative impact of exposed greenwashing on brand reputation corresponds to a model according to which greenwashing erodes trust and loyalty and leads to a deterioration in the customer relationship (Zhang *et al.*, 2018).

A surprising finding is that confidence in marketing sustainability claims in this study is not statistically dependent on the age of the respondent (H01 cannot be rejected). Generations Y and Z therefore form a relatively homogeneous whole within the Czech fashion market in terms of the level of scepticism towards green claims. Also, the frequency of slow fashion purchases is not related to trust in marketing claims (H02 cannot be dismissed) — this means that the decision to buy slow fashion products is not primarily driven by deep trust in the brands‘ message, but rather by other motives, such as quality, durability or natural materials, which dominate the ranking of factors in clothing purchases in this study. This conclusion corresponds to the view of Domingos *et al.* (2022), according to which slow-fashion consumers combine an emphasis on uniqueness, quality and authenticity with an interest in ethical and environmental values.

On the contrary, hypotheses H03 and H04 were rejected with a weak, statistically significant dependence. Respondents who are more familiar with the term greenwashing or, according to their own experience, encounter it more often when buying, have reduced trust in marketing

claims about sustainability. The weak intensity of the association (Cramér's V = 0.2184 for H03 and 0.1578 for H04) suggests that knowledge of greenwashing influences trust, but is not its sole or dominant source; The general level of media literacy, value orientation or experience with specific brands probably also play an important role.

A high proportion of respondents (58.7%) for whom transparent communication about the origin of materials and production processes increases trust in the brand is in line with the conclusions of Adamkiewicz *et al.* (2022), according to which more transparent marketing and reduction of greenwashing strengthen trust in green products. This result is a well-usable signal in practice: transparency is not just a defense against legal risks, but an active tool for building brand credibility.

4.1 Recommendations for practice

Based on empirical findings, four key communication principles for sustainable fashion businesses can be formulated that minimize the risk of greenwashing:

1. **Transparency and traceability.** Sustainability communications should be supported by easily traceable information on the origin of materials, production processes and certifications. A suitable tool is the publication of product stories based on life cycle analysis (LCA) on the brand's website, or linking the product to a detailed description via a QR code.
2. **Authenticity.** Brand credibility is strengthened by open communication about gradual steps, achieved goals and uncoordinated areas. Vague or aspirational promises for the future, on the other hand, are worth avoiding — according to the survey, respondents perceive them as clearly suspected of greenwashing.
3. **Caution in visual communication.** Visual greenwashing — the overuse of green color, nature motifs, leaves, and general „eco" aesthetics with no substantiated content — is a separate form of greenwashing (de Freitas Netto *et al.*, 2020). Therefore, visuals and text claims should be pre-tested with representatives of the target group.
4. **Differentiation of channels by generation.** For the youngest segment of 14-21 years old, social networks (Instagram, YouTube) are key, for older representatives of Generation Y, sellers' websites and recommendations from friends and family are becoming increasingly important. Collaborating with selected micro-influencers in the sustainability segment can strengthen brand credibility, provided that the influencer has an authentic relationship to the topic — the influencer's perceived authenticity towards sustainability is key to follower trust, while combining sustainable and unsustainable partnerships weakens credibility (Jacobson and Harrison, 2022).

4.2 Contribution limits

The study should be interpreted with three main limitations in mind. First, the sample is not fully representative, with the representation of respondents in two of the four age quotas deviating from the population by more than 7 percentage points. Second, the segmentation was carried out exclusively by age, without psychographic or behavioral variables that can influence the perception of greenwashing more significantly than age; Further research could use cluster analysis methods over attitude and value scales. Third, the perception of what is and is not greenwashing is itself subjective and controversial in the literature (de Freitas Netto *et al.*, 2020); Therefore, it is advisable to supplement the quantitative view with qualitative methods, such as semi-structured interviews, in follow-up research.

4.3 Conclusion

The paper presents an empirical picture of the trust of generations Y and Z in the environmental claims of fashion brands on the Czech market. Trust is low (68% tend to distrust), statistically dependent on knowledge and the perceived frequency of greenwashing, but not on age or frequency of slow fashion purchases. For sustainable fashion brands, this means that preventing a loss of trust is not about increasing the volume of sustainability messages, but about their quality — transparency, authenticity, traceability and visual caution.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ADAMKIEWICZ, J., KOCHAŃSKA, E., ADAMKIEWICZ, I., ŁUKASIK, R. M. 2022. Greenwashing and sustainable fashion industry. *Current Opinion in Green and Sustainable Chemistry*. 38, 100710.

CZECH STATISTICAL OFFICE. 2022. *Age composition of the population of the Czech Republic as of 31.1.2022*. Prague: Czech Statistical Office.

DE FREITAS NETTO, S. V., SOBRAL, M. F. F., RIBEIRO, A. R. B., SOARES, G. R. L. 2020. Concepts and forms of greenwashing: a systematic review. *Environmental Sciences Europe*. 32(1), 19.

DOMINGOS, M., VALE, V. T., FARIA, P. 2022. Slow fashion consumer behavior: a literature review. *Sustainability*. 14(5), 2860.

EUROPEAN COMMISSION. c1995–2023. *Green Claims*. Brussels: European Commission.

HAMEED, I., HYDER, Z., IMRAN, M., SHAFIQ, K. 2021. Greenwash and green purchase behavior: an environmentally sustainable perspective. *Environment, Development and Sustainability*. 23(9), 13113–13134.

HENDL, J. 2015. *Overview of statistical methods: data analysis and meta-analysis*. 5th ed. Prague: Portal.

JACOBSON, J., HARRISON, B. 2022. Sustainable fashion, social media, influencers and content creation calibration. *International Journal of Advertising*. 41(1), 150–177.

JUNG, S., JIN, B. 2014. A theoretical investigation of slow fashion: sustainable future of the apparel industry. *International Journal of Consumer Studies*. 38(5), 510–519.

NIINIMÄKI, K., PETERS, G., DAHLBO, H., PERRY, P., RISSANEN, T. *et al*. 2020. The environmental price of fast fashion. *Nature Reviews Earth & Environment*. 1(4), 189–200.

NOSTO. 2022. *Sustainable Fashion Report*. London: Nosto.

SAMAL, S. D., BHALALA, T. 2023. An awareness towards greenwashing of product and its impact on consumers' buying behaviour in Surat City. I*nternational Journal of Engineering and Management Research*. 13(2).

ŠTEFKO, R., STEFFEK, V. 2018. Key issues in slow fashion: current challenges and future perspectives. *Sustainability*. 10(7), 2270.

TAHAL, R. 2022. *Marketing research: procedures, methods, trends*. 2nd ed. Prague: Grada Publishing.

VELASCO-MOLPECERES, A., ZARAUZA-CASTRO, J., PÉREZ-CURIEL, C., MATEOS-GONZÁLEZ, P. 2023. Slow fashion as a communication strategy of fashion brands on Instagram. *Sustainability*. 15(1), 423.

WAWER, M., GRZESIUK, K., JEGOROW, D. 2022. Smart mobility in a smart city in the context of Generation Z sustainability, use of ICT, and participation. *Energies*. 15(13), 4651.

ZHANG, L., LI, D., CAO, C., HUANG, S. 2018. The influence of greenwashing perception on green purchasing intentions: the mediating role of green word-of-mouth and moderating role of green concern. *Journal of Cleaner Production*. 187, 740–750.

Contact information

Lucie Veselá: ee-mail: lucie.vesela@mendelu.cz

INTERACTIVE TRANSFORMATION OF LAYERS IN 3D TILES FORMAT WITHIN A WEB GIS APPLICATION

Markéta Nedělová[1], Ivo Pisařovic[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_025

ABSTRACT

The paper deals with the design and implementation of an extension to the Spatial Lens web application, which is used to prepare 3D educational materials displayed in virtual reality. The subject of the extension is editing spatial layers stored in the 3D Tiles format, specifically their displacement, rotation and scaling directly in the scene preview, which eliminates the need to use external tools. Three-dimensional geodata is not always perfectly prepared for a specific scene and may have a different axis orientation or incorrect scale. The entire 4×4 transformation matrix is not transferred between the frontend and backend, but a more economical data structure containing the position (longitude, latitude and altitude), rotation and scale. This solution reduces cumulative rounding errors, increases accuracy for large numbers, and facilitates input validation. The backend recalculates the resulting matrix each time from the beginning from the saved original matrix and a stable pivot point, thus ensuring idempotence of repeated saves. The usability of the extension was verified through scenario-based user testing with ten participants. The results confirm the usability of the tool while identifying areas for further improvement of the user experience.

Keywords: 3d tiles, cesiumjs, web gis application, spatial data transformation
JEL Code: C88, L86

1 INTRODUCTION

The development of geospatial technologies has led to the wider use of three-dimensional geographic data in a number of areas, such as spatial planning, environmental monitoring and navigation systems (Biljecki *et al.*, 2015). The ability to visualize and manipulate 3D geographic data allows for a more realistic and detailed representation of the real world, which supports decision-making and analysis (Ketzler *et al.*, 2020). This is accompanied by the growing importance of web geographic information systems (Web GIS), which make complex spatial datasets accessible directly in the browser without the need to install specialized desktop software; map creation on the web has undergone rapid development in recent decades from simple static map services to fully-fledged interactive systems (Veenendaal *et al.*, 2017).

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

One such platform is Spatial Lens, formerly known as Spatial Data Service, designed for preparing 3D educational courses that are subsequently displayed in virtual reality (Chylíková, 2025). The system builds on the work of Spatial Data Browser for Metaverse, which was created as part of a bachelor's thesis by Mitter (2022) as a client-server application for visualizing 3D spatial data in virtual reality using the 3D Tiles format and open web standards. Chylíková (2025) subsequently expanded this system to support the creation of educational materials.

Three-dimensional geodata and 3D models are not always perfectly prepared for a specific scene. For example, they may have different axis orientations or incorrect scales. If editing is not available directly in the application, the user must use an external tool, which complicates the entire workflow. The ability to move, rotate, and scale a layer allows you to adapt the scene to a specific learning goal. The 3D Tiles standard is designed primarily for streaming and visualization, not for editing geometry (Open Geospatial Consortium, 2023). However, it is possible to move, rotate, or scale an entire tileset by editing the root transformation matrix in the tileset.json file.

The main objective of the paper is to describe the design and implementation of an extension to the Spatial Lens application that allows users to perform editing operations with spatial layers (shift, rotate, and scale 3D objects) directly in the application, with a live preview of the scene. The usability of the proposed solution was subsequently evaluated through user testing.

2 METHODOLOGY AND DATA

The CesiumJS library (CESIUM, 2026) uses several coordinate systems to work with the position of objects. The geodetic system determines the position of a point using longitude, latitude, and altitude, taking into account the fact that the Earth is not a perfect sphere. The Earth-Centered, Earth-Fixed (ECEF) system describes the position in Cartesian coordinates relative to the center of a reference ellipsoid and rotates with the Earth, so that a point fixed on the surface has constant coordinates in this system (Van Sickle, 2017). Because the global Z axis points to the North Pole and does not coincide with the local vertical at any point, the East-North-Up (ENU) system is introduced for local orientations and transformations. In CesiumJS, it is generated by the `Transforms.eastNorthUpToFixedFrame function`, which returns a 4×4 transformation matrix.

In computer graphics, linear transformations in three-dimensional space are usually represented by 4×4 matrices in homogeneous coordinates and are divided into three basic types: translation, rotation, and scaling. Rotation can be expressed in several ways — using a transformation matrix, a triple of Euler angles, or quaternions. Euler angles are intuitive for users, but they can lead to a phenomenon called gimbal lock, in which one degree of freedom is lost. Quaternions prevent this problem and serve as a robust intermediate step in the calculations (Žára *et al.*, 2005). However, for the user interface, Euler angles were chosen because they provide an intuitive way for users to input precise numerical values via sliders.

The entry point for each tileset is the tileset.json file, whose `root.transform` property contains a 4×4 matrix stored in column order that places the local coordinate system of the model in the ECEF system (Open Geospatial Consortium, 2023). The primary format for the content of the tiles is the native glTF 2.0 in the 3D Tiles 1.1 specification (The Khronos Group, 2021). Due to the enormous scale of the virtual globe, 32-bit floating-point numbers cannot be reliably used for coordinates; therefore, CesiumJS operates at the processor level with 64-bit precision and employs techniques such as Relative-to-Center and Relative-to-Eye (Cozzi and Ring, 2011).

The above features lead to a key design decision: the backend does not receive the entire 4×4 matrix from the frontend, but the more efficient data structure shown in Listing 1, which contains position, scale, and rotation.

```
"transform": {
    "position": {
            "longitude": 16.613117421768894,
            "latitude": 49.19078110708189,
            "altitude": 220.1149262056975
    },
    "scale": { "x": 1, "y": 1, "z": 1 },
    "rotation": { "yaw": 90, "pitch": 0, "roll": 0 }
}
```

Listing 1 Transform data structure designed for communication between frontend and backend.

There are four reasons for this solution. The structure can be easily extended with future features, such as pivot point selection. It also prevents the accumulation of errors: if the frontend were to repeatedly multiply the current matrix and send it for storage, this would lead to gradual distortion (skew or rotation drift) after several saves. The third reason is accuracy for large numbers — ECEF coordinates are in the millions of meters, and transmitting a 4×4 matrix in JSON could lead to a loss of decimal precision, while transmitting geographic coordinates is more secure. The last reason is easier validation and security of inputs on the backend side.

The application is built on the Vue.js framework (frontend) and FastAPI (backend), the MongoDB database is used for data storage. Data validation on the backend is provided by the Pydantic library, and the vue-cesium component library is used on the frontend to work with CesiumJS.

The usability of the extension was verified by a scenario usability test combined with the think-aloud method. The method is based on cognitive psychology (Ericsson and Simon, 1980) and has become a standard in the field of usability testing (Nielsen, 1994). Evaluating the usability of web mapping applications through a combination of expert judgment and user testing is an established procedure in this domain (Nivala *et al.*, 2008). Ten people participated in the testing as part of our research; the total time to complete the task was measured, and success was monitored qualitatively and supplemented with a questionnaire.

3 RESULTS

The proposed solution consists of two tightly coupled parts. On the backend, a dedicated computation pipeline translates the compact Transform structure into a precise 4×4 root matrix. On the frontend, the same transformation parameters drive a real-time scene preview, giving the user immediate visual feedback. The following subsections describe both layers in detail, followed by the results of the usability evaluation.

3.1 Calculating the transformation on the backend

The main computational logic is located on the backend in the `geo_utils.py` module, whose functions are called when the `PUT /layers/{id}` endpoint is queried. On the first save, the root matrix `root.transform` generated by the processing pipeline is stored in the `asset.extras.originalRoot`

array and is never overwritten on subsequent saves. This value serves as a stable reference point representing the original position of the model. Each subsequent save recalculates the entire transformation matrix from the beginning from this reference, rather than incrementally modifying the previously saved matrix. This prevents the accumulation of rounding errors that would otherwise manifest as gradual geometric deformations.

A prerequisite for proper rotation and scaling is the determination of a stable pivot point. Without it, rotation around an arbitrary origin would cause the model to "orbit" through space, rather than rotating in place. The pivot point is defined as the center of the enclosing body of the tile set expressed in ECEF coordinates, calculated once from `originalRoot` and constant throughout the lifetime of the set. The resulting matrix is then composed according to relation (see Listing 2), where `ENU(pos)` denotes the ENU matrix at the target position, `HPR` the rotation matrix, `Scale` the scale matrix, and `bsCenter` the center of the enclosing body:

```
new_root = ENU(pos) × HPR × Scale × inv(ENU(bsCenter)) × originalRoot
```

Listing 2 Composing the resulting matrix.

This logic is written in the `geo_utils.py` module as shown in Listing 3.

```
new_root = ( +
    enu_target
    @ rot_4x4
    @ scale_4x4
    @ np.linalg.inv(enu_bs_center)
    @ original_root
)
data["root"]["transform"] = new_root.flatten(order='F').tolist()
```

Listing 3 Folding the resulting root matrix on the backend ("full-bake" method).

A less obvious complication arises from the region type wrapping body, which uses geographic coordinates instead of Cartesian coordinates local to the tiles. CesiumJS derives tile visibility from these coordinates independently of `root.transform`, so after saving a new position, the culling system continues to evaluate visibility at the original position and the model sometimes disappears from view. The solution is to replace the region definition with an oriented wrapping box expressed in the local tile space, which is correctly transformed using `root.transform` according to the 3D Tiles specification (Open Geospatial Consortium, 2023). This conversion only happens once, on the first save, when the region type wrapping body is detected.

3.2 Real-time preview on the frontend

The frontend dynamically updates the 3D preview of the layer based on how the user interacts with the controls. When the tileset is finished loading, the `onReady` callback explicitly downloads the `tileset.json` file from the server. This additional request is necessary because the CesiumJS rendering engine optimizes the tile hierarchy internally, often omitting its own extras array, which holds the original transformation parameters.

To make the operations idempotent, the model matrix is recomputed from scratch from the full stored transformation state (position, rotation, and scale) for each change using the

`calculateModelMatrix function` (Listing 4). Because the rendered state is always a pure function of the stored values, rather than an accumulation of incremental changes, repeated save-load cycles produce the same visual result. This approach eliminates the initial problem of errors gradually accumulating as the scale changes.

```
export const calculateModelMatrix = (transform: Transform): Matrix4 | null => {
    const targetPos = Cartesian3.fromDegrees(
           transform.position.longitude,
           transform.position.latitude,
           transform.position.altitude
    );
    const newENU = Transforms.eastNorthUpToFixedFrame(targetPos);
    const scaleMatrix = Matrix4.fromScale(
           new Cartesian3(transform.scale.x, transform.scale.y, transform.scale.z),
           new Matrix4()
    );
    const hpr = new HeadingPitchRoll(
           CesiumMath.toRadians(transform.rotation.yaw),
           CesiumMath.toRadians(transform.rotation.pitch),
           CesiumMath.toRadians(transform.rotation.roll)
    );
    const rotationMatrix4 = Matrix4.fromRotation(
           Matrix3.fromHeadingPitchRoll(hpr, new Matrix3()),
           new Matrix4()
    );
    const localTransform = Matrix4.multiply(
           rotationMatrix4, scaleMatrix, new Matrix4()
    );
    return Matrix4.multiply(newENU, localTransform, new Matrix4());
};
```

Listing 4 Calculating the model matrix on the frontend for real-time preview.

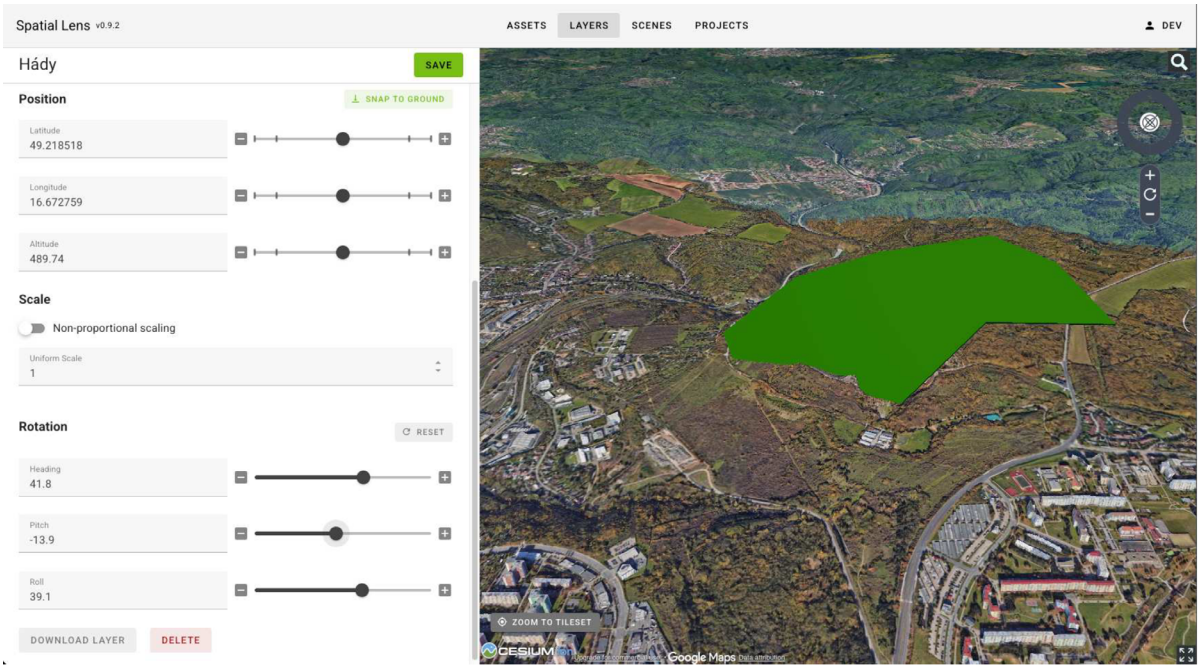


Fig. 1: The resulting implementation of a layer detail with position, scale and rotation settings and a 3D scene preview. (Nedělová, 2026)

The transformation control user interface (see Figure 1) includes separate components for translation, rotation, and scaling. For precise adjustments, a custom `VariableSlider` component has been created that supports dynamic step sizes — both linear and logarithmic behavior based on the distance of the slider from the center. This allows the user to quickly make coarse adjustments as well as fine micro-adjustments. The `flyToLayer` function dynamically calculates a new bounding sphere for objects outside the originally defined area from the transformed coordinates and the changed radius to correctly point the camera at the current spatial extent.

3.3 User testing results

Ten people (aged 26 to 40) participated in the test. Six of them were educators from primary school to university. The remaining four were technically proficient users with experience with GIS applications. All participants successfully completed the assigned tasks, in some cases with minor assistance. The duration of the testing phase ranged from 15.22 to 42.31 minutes, with an average of 27.33 minutes and a median of 25.90 minutes. The upper limit fell on the first session, during which several critical problems were identified and subsequently eliminated.

The evaluation of transformation functions and related interface aspects from the questionnaire is summarized in Table 1. Columns 1 to 5 indicate the number of participants who chose a given evaluation on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree).

The results show that the best rated were rotation adjustment (4.3) and confidence in future work efficiency (4.5). Scale adjustment received 3.8 points; position adjustment (3.6) and interface orientation (3.4) received the lowest ratings with a greater variance of responses. Qualitative findings indicated that participants were unsure of which direction the layer would move, as it depended on the current camera orientation; most had to move the slider and observe the result. Three participants preferred the increase and decrease buttons to dragging the slider. Three participants experienced incorrect caching behavior, where the layer did not reflect the changes after saving, and a hard page reload was required. The most frequently mentioned wish was the ability to edit layers directly in the scene preview.

Tab. 1 Evaluation of questionnaires related to transformations (selection from results).

Questionnaire item	1	2	3	4	5	Average
Future work efficiency	0	0	1	3	6	4.5
Rotation adjustment	0	0	0	7	3	4.3
Scaling	1	1	1	3	4	3.8
Position adjustment	0	2	2	4	2	3.6
Interface orientation	0	3	2	3	2	3.4

4 DISCUSSION AND CONCLUSIONS

The aim of the paper was to describe the design and implementation of an extension to the Spatial Lens application that allows moving, rotating and scaling layers in 3D Tiles format directly in the scene preview. The key design decision was to transfer the economical Transform structure instead of the entire 4×4 matrix and to recalculate the resulting matrix on the backend from scratch, always from the saved original matrix and a stable pivot point. This "full-bake" approach has proven itself in practice: repeated save-load cycles led to identical visual results, thus confirming the idempotence of operations and eliminating the accumulation of round-off errors and initial distortion when changing the scale. At the same time, it was necessary to solve a less obvious complication with the region-type wrapping body, which led to the model disappearing from the view after its movement.

The results of the user testing confirm the usability of the tool. Of the transformation functions, rotation was rated best (4.3), followed by scaling (3.8) and position adjustment (3.6). The high rating of confidence in future work efficiency (4.5) indicates that participants expected to complete the task significantly faster in subsequent sessions, which indicates a manageable learning curve. However, the lower and more scattered ratings of position adjustment and orientation in the interface indicate room for improvement of the user experience, especially regarding the dependence of the direction of movement on the camera orientation. The finding that interactive work with 3D objects places different demands on users than working with static displays is consistent with research on the differences between interactive and static 3D maps (Herman *et al.*, 2018).

The testing resulted in several recommendations for further development. It would be beneficial to add an interface with a manipulator (gizmo) for direct manipulation of the object, preset camera modes, support for keyboard shortcuts for placing layers, and a snap to terrain function. Although the sample of ten participants was sufficient for identifying key usability issues (Nielsen, 1994), a larger and more diverse group would strengthen the generalizability of the findings. The developed extension is functional, meets the set goals, and represents a solution that can be built on; future development should therefore focus on implementing improvements based on feedback from testing.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education, and also supported by grant No. IGA24-PEF-TP-003 (Using metaverse technologies for working with spatial data) of the Internal Grant Agency FBE MENDELU.

Anthropic Claude was used to summarize a part of the final thesis (Nedělová, 2026) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

BILJECKI, F., STOTER, J., LEDOUX, H., ZLATANOVA, S., ÇÖLTEKIN, A. 2015. Applications of 3D City Models: State of the Art Review. I*SPRS International Journal of Geo-Information*. 4(4), 2842–2889.

CESIUM. 2026. *CesiumJS Documentation*. Exton, PA: Cesium GS, Inc. https://cesium.com/learn/cesiumjs/ref-doc/

CHYLÍKOVÁ, E. 2025. *Supporting education through virtual reality and geospatial data*. Diploma thesis. Brno: Mendel University in Brno, Faculty of Business and Economy.

COZZI, P., RING, K. 2011. *3D Engine Design for Virtual Globes*. 1st Edition. Boca Raton, FL: CRC Press.

ERICSSON, K. A., SIMON, H. A. 1980. Verbal reports as data. *Psychological Review*. 87(3), 215–251.

HERMAN, L., JUŘÍK, V., STACHOŇ, Z., VRBÍK, D., RUSSNÁK, J., ŘEZNÍK, T. 2018. Evaluation of User Performance in Interactive and Static 3D Maps. ISPRS *International Journal of Geo-Information*. 7(11), 415. https://doi.org/10.3390/ijgi7110415

KETZLER, B., NASERENTIN, V., LATINO, F., ZANGELIDIS, C., THUVANDER, L., LOGG, A. 2020. Digital Twins for Cities: A State of the Art Review. *Built Environment*. 46(4), 547–573.

MITTER, M. 2022. *Spatial Data Browser for Metaverse*. Bachelor thesis. Brno: Mendel University in Brno, Faculty of Business and Economy.

NEDĚLOVÁ, Markéta. 2026. *Editing Spatial Data Using a Web Application*. Diploma thesis. Brno: Mendel University in Brno.

NIELSEN, J. 1994. *Usability Engineering*. San Francisco, CA: Morgan Kaufmann.

NIVALA, A.-M., BREWSTER, S., SARJAKOSKI, LT 2008. Usability Evaluation of Web Mapping Sites. *The Cartographic Journal*. 45(2), 129–138. https://doi.org/10.1179/174327708X305120

OPEN GEOSPATIAL CONSORTIUM. 2023. *3D Tiles Specification 1.1*. Arlington, VA: Open Geospatial Consortium. http://www.opengis.net/doc/cs/3D-Tiles/1.1.

THE KHRONOS GROUP. 2021. *glTF 2.0 Specification: Version 2.0.1*. Beaverton, OR: The Khronos Group. https://registry.khronos.org/glTF/specs/2.0/glTF-2.0.html

VAN SICKLE, J. 2017. Two Coordinate Systems. In: *Basic GIS Coordinates*. 3rd Edition. Boca Raton, FL: CRC Press, pp. 47–68.

VEENENDAAL, B., BROVELLI, M. A., LI, S. 2017. Review of Web Mapping: Eras, Trends and Directions. I*SPRS International Journal of Geo-Information*. 6(10), 317. https://doi.org/10.3390/ijgi6100317

ŽÁRA, J., BENEŠ, B., SOCHOR, J., FELKEL, P. 2005. *Modern computer graphics*. 2nd Edition. Brno: Computer Press.

Contact information

Markéta Nedělová: e-mail: xnedelov@mendelu.cz

Ivo Pisařovic: e-mail: research@pisarovic.cz

A WEB-BASED FINANCIAL PLANNING APPLICATION FOR ENTERPRISE BUDGETING AND INVESTMENT MANAGEMENT: A DESIGN AND IMPLEMENTATION CASE STUDY

Ondřej Pavliík[1], Oldřich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_026

ABSTRACT

A mid-sized company that tripled its number of organizational units over five years outgrew its spreadsheet-based budgeting. Centre managers planned in their own Excel files, the finance department merged them by hand, and errors surfaced only after the fact. This paper reports the design and implementation of an on-premise web application that sits between the company ERP system and the centre managers, replacing the manual consolidation with a single validated planning workflow. The work follows a standard software-engineering process: analysis of the existing routine, requirements (22 functional, 12 non-functional), a make-or-buy assessment of three commercial vendors, UI design in Figma, implementation, testing, and deployment. The application is built on ASP.NET Core (.NET 8) with Razor Pages, a Microsoft SQL Server database, SOAP integration with the ERP, integrated Windows authentication, and Power BI reporting embedded through OAuth 2.0. None of the three vendors covered the on-premise requirement, customization, and integration at once, so a tailored solution was developed and is now in production. We describe the architecture, the integration and authentication decisions, and five concrete problems solved during development. No quantitative performance metrics were measured; the contribution is the engineering design.

Keywords: financial planning, ASP.NET Core, ERP integration, on-premise deployment, Power BI, Razor Pages
JEL Code: C88

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Budgeting in a small company often starts as a single shared spreadsheet. It works until the company grows. The firm behind this case study tripled its number of organizational units in five years (Pavlík, 2025), and the spreadsheet routine did not scale with it. Each centre manager planned in a separate Excel file, the files arrived in different layouts, and the finance department merged them manually into a company-wide plan. Mistakes in a single cell propagated into the consolidated figures and were usually caught only during a later review, when correcting them meant another round of e-mails.

The aim of the work was to design and implement a web application for managing and planning corporate budgets and investments that acts as an intermediary between the company ERP system and the centre managers. It was to replace the manual consolidation with a centralized overview and validate entries as they are made, rather than after submission. Two constraints shaped every later decision: the system had to run fully on-premise, and it had to fit the company's existing Microsoft-based infrastructure instead of forcing new processes onto the staff.

1.1 Why not a commercial product

The cost of bad financial data is a recurring theme in the management-information literature: data that cannot be trusted forces manual rework and erodes the value of every report built on top of it (Redman, 2008; Laudon and Laudon, 2020). Centralizing the planning data was therefore the first goal, ahead of any new feature. The make-or-buy question came next. Romero *et al.* (2021) show that financial web applications are commonly assembled on top of a business-intelligence layer rather than written from scratch, which is one argument for buying.

Three commercial vendors active on the Czech market were assessed against the company's key requirements, together with the planning module of the existing ERP. The comparison was qualitative; it is reproduced in Section 3 as Table 1. None of the options met the on-premise requirement, the demand for deep customization, and integration with the existing systems at the same time. Cloud dependence was the most frequent disqualifier. The ERP planning module was too rigid to model the company's centre structure. This led to the decision to build a tailored solution, which the requirements analysis (Sommerville, 2015; Gorbachenko, 2021) then turned into 22 functional and 12 non-functional requirements.

2 METHODOLOGY

The project followed a sequential software-engineering process with an iterative testing phase: analysis of the current process, requirements definition, UI design, implementation, testing, and deployment. The non-functional requirements fixed the technology before the first line of code, because the firm already ran a Microsoft stack and Power BI licences and wanted to keep them.

The backend is an ASP.NET Core application on .NET 8, built as a monolith with Razor Pages and server-side rendering and hosted by the integrated Kestrel server (Freeman, 2023; Karpenko, 2025). A monolith was a deliberate choice for a single internal tool with one team; the trade-offs against a service-oriented split are well documented (Richards and Ford, 2020; Newman, 2021; Bass *et al.*, 2012), and none of the drivers for splitting applied here. Persistent data lives in Microsoft SQL Server. The schema was designed first as a conceptual model abstracted over four main sections

of the system and then as a physical ER diagram. Database rows are mapped onto domain models through a dedicated mapping layer, and the services are wired together with dependency injection.

Integration with the ERP uses a SOAP API with scheduled synchronization. The SOAP client is configured dynamically rather than from a fixed endpoint, so the same code reaches different ERP environments without recompilation, and the channel is secured for the data it carries. Sensitive configuration is kept out of source control with User Secrets, and the application logs through Serilog.

Authentication reuses the company's Active Directory. With integrated Windows authentication, a user who is already signed in to the domain reaches the application without logging in again. Reporting is handled by Power BI, embedded under OAuth 2.0 in the Service Principal mode, where the application authenticates as itself and requests an embedded token carrying an Effective Identity so that each manager sees only the figures for their own centres. The deployment topology, including the link to the reporting service, is shown in Figure 1.

The frontend stays inside the same Razor Pages application. Layout uses Tailwind CSS with Heroicons, and the dynamic parts are plain JavaScript, without a separate framework. The interface was designed first in Figma, where the colour scheme, logo, wireframes, and clickable prototypes were checked with users before implementation; Figure 3 shows one of the resulting screens, the

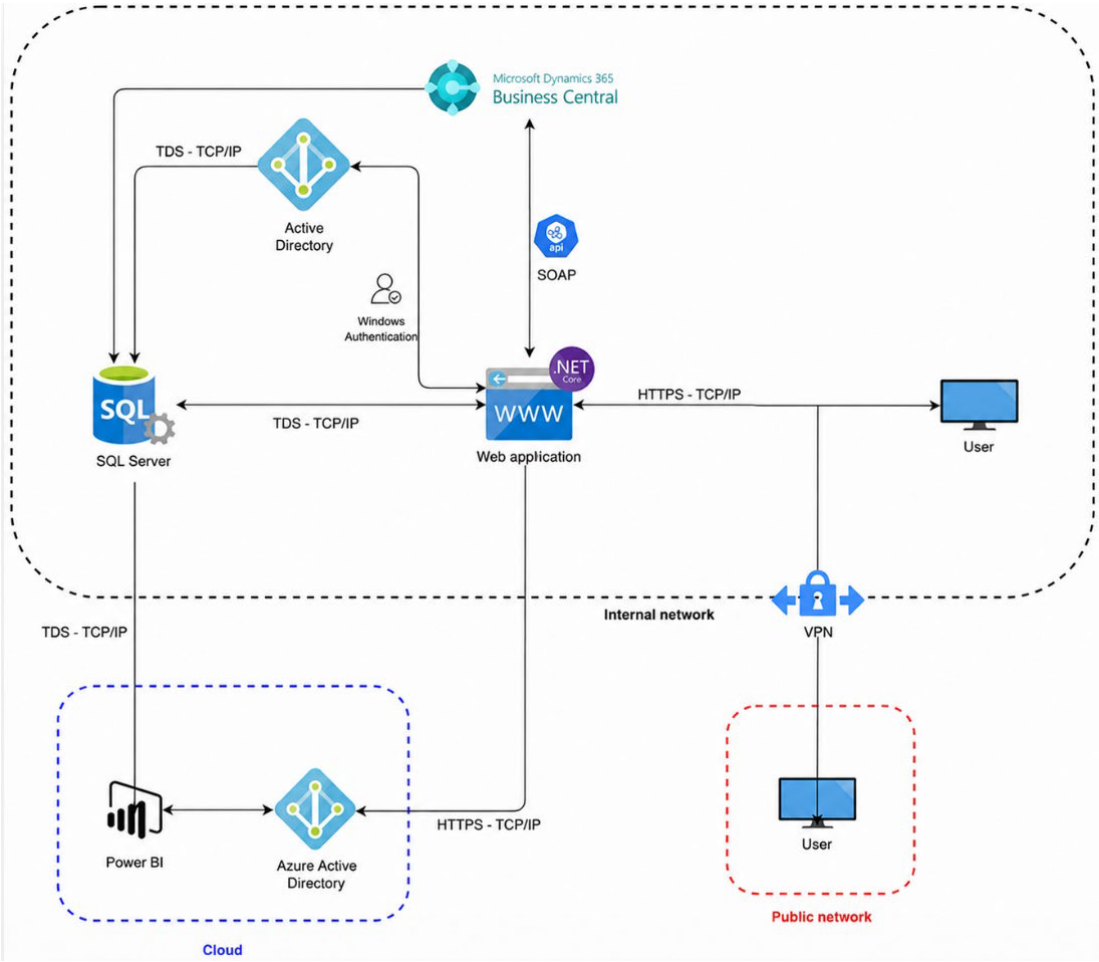


Fig. 1: Deployment architecture of the financial-planning web application integrating the ERP system, SQL Server, Active Directory and Power BI across the internal network and VPN (adopted from Pavlík, 2025).

investment-creation form. Reporting needed two Power BI variants, because paginated reports render differently from interactive ones and require their own embedding path.

The application is deployed on-premise on IIS, installed with the .NET Core Runtime and the Hosting Bundle and run from a dedicated application pool with in-process hosting. Historical data from the old spreadsheets was migrated during rollout, and every module was first released to a separate UAT (User Acceptance Testing) server. Testing combined MSTest unit tests on the key components, use-case scenarios that replay real planning situations, and isolated checks of the API against unexpected inputs. The thesis reports no test counts, coverage, or timing figures, and none are claimed here (Patton, 2002).

3 RESULTS

The outcome of the project is a working, deployed application, so the results are described qualitatively. Table 1 documents the vendor assessment that led to the in-house decision. It is the single table in the thesis and reports ordinal judgements, not measurements; no row is a winner, because every commercial option was rejected.

The tailored application replaced the Excel-based routine in production. Centre managers now enter their plans through the cost-planning screen shown in Figure 2, the entries are validated on the spot, and the consolidated view is built from the database instead of by hand. The application reads source data from the Business Central database and uploads planned data back to Business Central over SOAP, signs users in through their domain accounts, and presents the financial statements through embedded Power BI. The company reported that the time saved in some areas was higher than it had expected; that statement was not quantified in the thesis, and no figure is attached to it here.

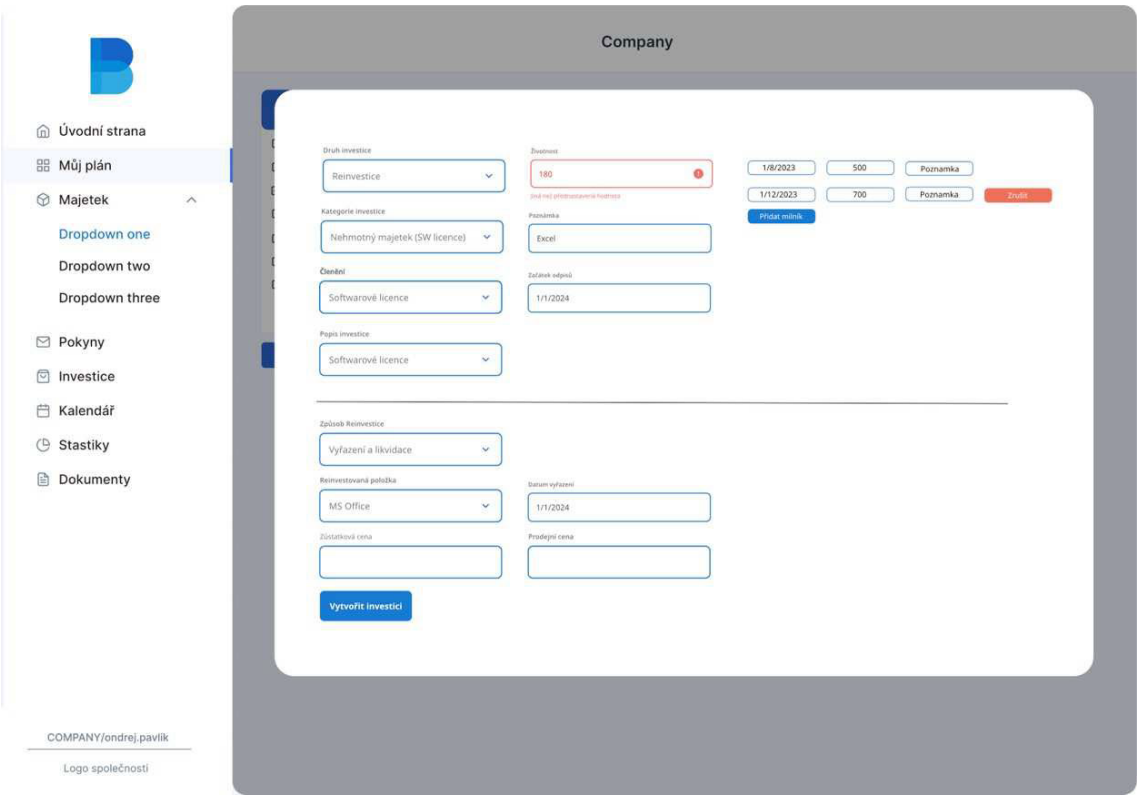
Five concrete problems were solved during development, and they are part of the result because they shaped the working system. The default .NET Core limit on the number of form fields was reached by the largest plans and had to be raised explicitly. Loading large plans was slow until the database indexing was reworked (Mukherjee, 2019). Power BI paginated reports did not render through the interactive embedding path and needed a separate one. Logging silently failed on IIS until the application pool was given write permissions to the log directory. The frontend constraints rendered correctly only in Chromium and required extra work for Safari and Firefox. None of these is dramatic on its own; together they are most of the distance between a prototype and a tool people use daily.

Tab. 1 Qualitative comparison of the three assessed commercial vendors against the client's key require-ments (adopted from Pavlík, 2025)

Criterion	Sophia Solutions	TARGETTY	GIST
On-premise operation	Partly supported	Possible, mostly not	Yes
Customization flexibility	Low	Medium	Very high
Integration with existing systems	Standard connectors + API	Strong MS-ecosystem integration	Any custom interface
Implementation speed	Medium	Fast	Long
Initial price	Higher	Medium	High
Cloud dependence	Optional	Yes	Optional



Tab. 2 Annual cost-planning screen of the deployed application, showing planned expense items per accounting period (adopted from Pavlík, 2025)



Tab. 3 Investment-creation form with depreciation parameters and reinvestment milestones (adopted from Pavlík, 2025)

4 DISCUSSION

The decisive constraint was not technical but organizational: the company wanted to keep its processes and its infrastructure. That single requirement explains most of the design. It ruled out the cloud-leaning vendors in Table 1, it made integrated Windows authentication the natural sign-in mechanism, and it made Power BI the obvious reporting tool, since the licences and the staff's familiarity with it were already there. A greenfield project would likely have chosen differently; here, fitting the existing environment mattered more than picking the theoretically best component.

The monolith deserves a second look, because microservices are often treated as the default for new systems. For one internal application maintained by one team, the usual reasons to split (independent scaling, separate release cycles, multiple teams) were absent, and a split would have added deployment and operational cost for no return. The Power BI choice is more debatable. It was right given the existing licences, but it ties the reporting to a paid product and to an embedding model that, as the paginated-report problem showed, is not uniform across report types. A company without those licences might reasonably build its own reporting instead.

We should be honest about what this work can and cannot claim. It shows that a tailored on-premise application can replace a manual budgeting process and run in production, and it documents the decisions and the obstacles in enough detail to be reusable. It does not show by how much the new process is faster or less error-prone, because those quantities were never measured. The qualitative statement that time savings exceeded expectations is the client's, not a result of this study, and treating it as a number would overstate the evidence. The most useful part for a reader facing a similar problem is probably the make-or-buy reasoning together with the list of five concrete obstacles, which are easy to underestimate at the planning stage.

5 CONCLUSION

A web application for corporate budgeting and investment planning was designed, implemented, and deployed in a company's own infrastructure, replacing a manual, spreadsheet-based consolidation that no longer scaled with the firm's growth. The on-premise requirement, combined with the need for customization and ERP integration, was the reason no commercial product fit and a tailored solution on ASP.NET Core, SQL Server, and Power BI was built instead. For practitioners weighing the same decision, the lesson is that infrastructure fit and account integration can outweigh feature checklists, and that the unglamorous problems (form-field limits, indexing, report embedding, IIS permissions, cross-browser rendering) are where much of the implementation effort actually goes. The work was not evaluated with quantitative metrics; that is its main limitation. Useful next steps are clear and were identified during the project: ADFS account federation for single sign-on to Power BI, a possible in-house reporting layer to drop the Power BI dependency, form-based or OAuth 2.0 authentication for access from outside the local network, a React frontend to lift the interactivity limits of Razor Pages, and a CI/CD pipeline, for instance on Azure DevOps (Manolov *et al.*, 2025), to make releases routine.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Re-search Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Pavlik, 2025) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

BASS, L., CLEMENTS, P., KAZMAN, R. 2012. *Software Architecture in Practice*. 3rd ed. SEI Series in Software Engineering. Upper Saddle River, NJ: Addison-Wesley Pro-fessional. ISBN 978-0-321-81573-6

FREEMAN, A. 2023. *Pro ASP.NET Core 7*. 10th ed. Manning Publications. ISBN 1-63343-782-5

GORBACHENKO, P. 2021. What are Functional and Non-Functional Requirements and How to Document These. *Enkonix*. https://enkonix.com/blog/functional-requirements-vs-non-functional/

KARPENKO, M. 2025. Innovative Solutions for Enhancing System Performance on the .NET Platform. *International Journal of Computer*. 54(1), 21-28. https://ijcjournal.org/index.php/InternationalJournalOfComputer/article/view/2354

LAUDON, K., LAUDON, J. 2020. *Management Information Systems: Managing the Digital Firm*. 16th ed. Harlow: Pearson. ISBN 978-0-13-519179-8

MANOLOV, V. *et al*. 2025. Practical Comparison Between the CI/CD Platforms Azure DevOps and GitHub. *Future Internet*. 17(4), 153. https://doi.org/10.3390/fi17040153

MUKHERJEE, S. 2019. *Indexes in Microsoft SQL Server*. ResearchGate. https://www.researchgate.net/publication/331889027_Indexes_in_Microsoft_SQL_Server

NEWMAN, S. 2021. *Building Microservices*. 2nd ed. Sebastopol, CA: O'Reilly Media. ISBN 978-1-492-03402-5

PATTON, R. 2002. *Software Testing (in Czech)*. Brno: Computer Press. ISBN 80-7226-636-5

PAVLÍK, O. 2025. *Web application for financial planning* [in Czech]. Master's thesis, head: TRENZ, O. Brno: Mendel University in Brno.

REDMAN, T. C. 2008. *Data Driven: Profiting from Your Most Important Business Asset*. Boston, MA: Harvard Business Review Press. ISBN 978-1-4221-1912-9

RICHARDS, M., FORD, N. 2020. *Fundamentals of Software Architecture: An Engineering Approach.* 1st ed. Sebastopol, CA: O'Reilly Media. ISBN 978-1-4920-4345-4.

ROMERO, C. *et al*. 2021. Web Application Commercial Design for Financial Entities Based on Business Intelligence. *Computers, Materials & Continua*. 67(3), 3177-3188 https://doi.org/10.32604/cmc.2021.014738

SOMMERVILLE, I. 2015. *Software Engineering*. 10th ed. Harlow: Pearson. ISBN 978-0-13-394303-0

Contact information

Ondrej Pavlik: e-mail: 77449@node.mendelu.cz

Oldrich Trenz: e-mail: oldrich.trenz@mendelu.cz

AUTHORSHIP VERIFICATION OF ENGLISH TEXT: TRADITIONAL STYLOMETRY, TRANSFORMERS, AND LARGE LANGUAGE MODELS COMPARED AND DEPLOYED

Jiří Pokorný[1], Michaela Veselá[1], Oldřich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_027

ABSTRACT

This article compares three families of methods for authorship verification, the binary decision whether two texts share one author. We frame the task as pairwise classification and test traditional stylometric pipelines (bag-of-words and TF-IDF with scikit-learn classifiers), fine-tuned transformers (Longformer, BigBird-RoBERTa, ModernBERT), and large language models used through prompting. Training used the PAN 2020 corpus, testing the disjoint PAN 2021 corpus, together 295,800 English text pairs. A fine-tuned Longformer classification head reached the highest F1 of 92.38%, but a much cheaper TF-IDF pipeline with a linear classifier trailed it by less than one percentage point at 91.50%. The way two text vectors are combined into one feature vector mattered more than the classifier itself: absolute difference beat element-wise multiplication everywhere. The three open large language models we tried produced unusable output. Because hosting was constrained, the deployed web application runs TF-IDF with logistic regression (F1 91.28%), which returns the probability the interface needs.

Keywords: authorship verification, stylometry, TF-IDF, transformers, Longformer, large language models
JEL Code: C45

1 INTRODUCTION

Authorship verification asks a narrow question with wide consequences: were these two texts written by the same person? In this article we treat the task as an instance-based pairwise binary classification problem, without constructing au-thor profiles. The setting is relevant to forensics, law, cybersecurity, and identity-theft detection (Stamatatos, 2008; Koppel *et al.*, 2004).

The practical question behind this work is which method a developer should actually pick. Traditional stylometry counts words and syntactic patterns and feeds them to a classifier; it is cheap and well understood (Savoy, 2020). Transformer language models promise stronger text

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

representations but cost far more to train and serve (Vaswani *et al.*, 2017; Devlin *et al.*, 2018). Large language models can be prompted to answer directly, with no training at all (Huang *et al.*, 2024). This work therefore compares these three families on the same PAN-based corpus and then takes the selected model one step further, into a running web application.

1.1 Related work

Stylometric verification has a long record. Koppel *et al.* (2004) cast it as one-class classification; later surveys traced the field as it moved from hand-crafted features toward learned representations (Stamatatos *et al.*, 2016; Tyo *et al.*, 2022). The PAN shared tasks have set the common benchmark for cross-domain verification, including the 2020 and 2021 editions on fanfiction text whose corpora we use here (Kestemont *et al.*, 2020; Kestemont *et al.*, 2021). Ensembles of traditional classifiers have stayed competitive: Abbasi *et al.* (2022) report strong results from ensemble learning, and the best PAN 2021 entry was itself an ensemble.

Neural approaches have grown alongside. Siamese networks learn a similarity space for author identification at scale (Saedi *et al.*, 2021), and graph-based features encode syntax explicitly through integrated syntactic graphs (Gómez-Adorno *et al.*, 2016). The most recent thread asks whether large language models can judge authorship by prompting alone; Huang *et al.* (2024) report over 88 % F1 on an attribution variant using a proprietary model. Whether smaller open models reach that bar, and whether a fine-tuned encoder beats classical stylometry by enough to justify its cost, is what this work tests.

2 METHODOLOGY

We frame verification as pairwise binary classification. Each text becomes a feature vector; the two vectors of a pair are merged into one, and a classifier labels the pair as same-author or different-author. We compared two merge operators throughout: the absolute difference of the two vectors and their element-wise multiplication. The intuition is that multiplication highlights features the texts share, while absolute difference captures where they diverge.

Four experimental directions were run on identical data. The first used bag-of-words and TF-IDF representations (unigrams) with classifiers from scikit-learn: Multinomial Naive Bayes, Logistic Regression, LinearSVC, k-Nearest Neighbors, Random Forest, and XGBoost. The second replaced those representations with integrated syntactic graphs built per text by the text2graphapi library, turning each text into a frequency vector over a node dictionary and reusing the same classifiers. The third fine-tuned a classification head on three long-context transformers: Longformer-base-4096, BigBird-RoBERTa-large, and ModernBERT-large. The fourth prompted three open large language models (Llama-2-70B-chat-GPTQ, Gemma-3-4b-it, and Qwen3-4B-Instruct-2507) with no fine-tuning, using conservative decoding (low temperature, low top-p).

Hyperparameters were tuned with Optuna, 30 trials per representation-and-classifier combination, optimizing F1. Transformer trials trained on a reduced set of 19,999 pairs and validated on 2,500, with input capped at 4096 tokens; BigBird-RoBERTa and ModernBERT can take 8192 tokens, but we kept 4096 to hold training time down. We report accuracy, precision, recall, and F1, with F1 as the deciding metric, plus normalized confusion matrices.

The corpus is the PAN fanfiction dataset: the larger PAN 2020 variant for training and the disjoint PAN 2021 set for testing (Kestemont *et al.*, 2020; Kestemont *et al.*, 2021), 295,800 English pairs in

total. The test set holds 19,999 pairs (10,000 same-author, 9,999 different-author), and its authors do not appear in training. Texts are long, with mean and median length around 21,000 characters, which is why long-context transformers were a natural fit. Experiments ran on e-INFRA CZ / MetaCentrum (CESNET): NVIDIA H100 GPUs for transformers and language models, A10 and A40 cards for the stylometric and graph runs.

The verification model also had to run, not just score well offline. The best practical model was wrapped in a two-tier web application: a React 18.3.1 frontend in TypeScript and a FastAPI 0.128.0 backend exposing a /predict endpoint with Pydantic validation, deployed on the Railway cloud. The architecture is shown in Figure 1.

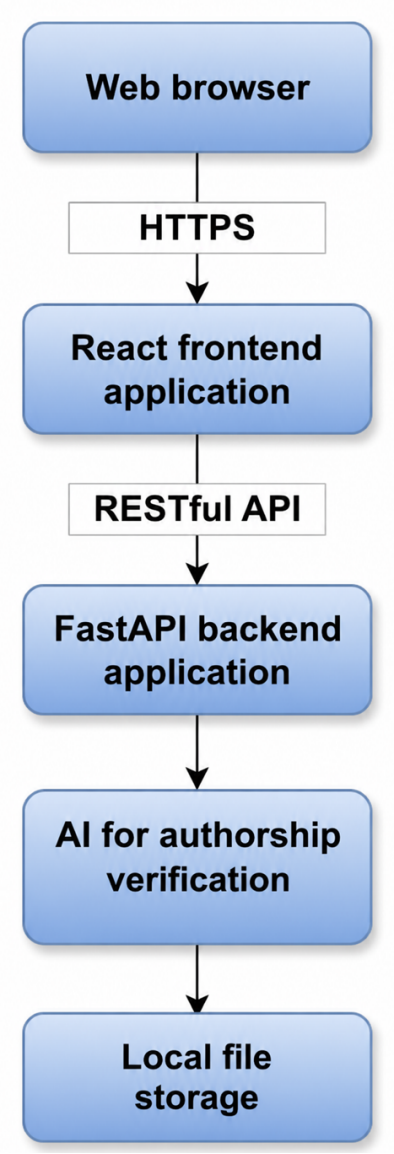


Fig. 1: System architecture of the authorship-verification web application, showing the layered flow from the web browser through the React frontend (HTTPS) and the FastAPI backend (RESTful API) to the AI authorship-verification module and the local file storage (adopted from Pokorny, 2026).

3 RESULTS

Table 1 lists the strongest run in each of the four directions, ranked by F1; the best overall row is in bold. The winner is a fine-tuned Longformer classification head at F1 92.38%. Second is the classical TF-IDF pipeline with LinearSVC at 91.50%, less than a percentage point behind. Two TF-IDF variants even posted higher accuracy (93.33% and 93.25%) than the Longformer (92.48%), though lower F1 (Pokorny, 2026).

Inside the stylometric direction (Table 2) the merge operator decided more than the classifier. With absolute difference, TF-IDF and LinearSVC reached F1 91.50%; the best multiplication run (BoW with LinearSVC) fell to 86.68%, and k-Nearest Neighbors collapsed to 63.75% with BoW and multiplication. TF-IDF beat plain bag-of-words across the board, since it weights author-specific words more heavily.

Among transformers (Table 3) Longformer led, BigBird-RoBERTa followed closely at 91.48%, and ModernBERT came last at 90.56% despite more than three times the training time of the others. The graph direction was weakest: its best result was Logistic Regression with absolute difference at F1 86.69%, and LinearSVC with multiplication dropped to 50.00% accuracy, no discriminative power at all. The three large language models all failed; their outputs were inconsistent or merely echoed the prompt, so they carry no numeric table.

Tab. 1 Best experiment per direction, ranked by F1

Representation/model	Combination	Classifier	Accuracy	Precision	Recall	F1
Longformer-base-4096	—	Classification head	92.48%	93.58%	91.22%	92.38%
TF-IDF	Absolute difference	LinearSVC	93.33%	89.39%	91.32%	91.50%
BigBird-RoBERTa-large	—	Classification head	91.64%	93.28%	89.76%	91.48%
TF-IDF	Absolute difference	Logistic Regression	93.25%	89.00%	91.08%	91.28%
ModernBERT-large	—	Classification head	90.72%	92.18%	88.99%	90.56%

Tab. 2 Best BoW / TF-IDF results, absolute-difference combination

Representation	Combination	Classifier	Accuracy	Precision	Recall	F1
TF-IDF	**Absolute difference**	**LinearSVC**	**93.33%**	**89.39%**	**91.32%**	**91.50%**
TF-IDF	Absolute difference	Logistic Regression	93.25%	89.00%	91.08%	91.28%
TF-IDF	Absolute difference	XGBoost	91.18%	88.96%	90.05%	90.17%
BoW	Absolute difference	Logistic Regression	90.21%	87.29%	88.73%	88.91%
BoW	Absolute difference	LinearSVC	90.72%	83.88%	87.17%	87.65%

Tab. 3 Best transformer results (classification head)

Representation	Classifier	Accuracy	Precision	Recall	F1
Longformer-base-4096	**Classification head**	**92.48 %**	**93.58 %**	**91.22 %**	**92.38 %**
BigBird-RoBERTa-large	Classification head	91.64 %	93.28 %	89.76 %	91.48 %
ModernBERT-large	Classification head	90.72 %	92.18 %	88.99 %	90.56 %
BoW	Logistic Regression	90.21%	87.29%	88.73%	88.91%
BoW	LinearSVC	90.72%	83.88%	87.17%	87.65%

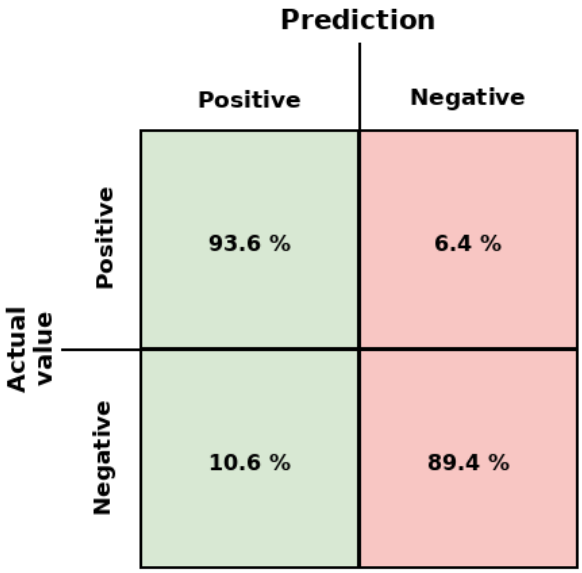


Fig. 2: Normalised confusion matrix of the best TF-IDF model (TF-IDF representation with LinearSVC and absolute-difference vector combination), reaching 93.6 % true positives and 89.4% true negatives (adopted from Pokorny, 2026)

4 DISCUSSION

The headline number favours the transformer, but the gap is too small to be the whole story. Less than one percentage point of F1 separates a fine-tuned Longformer from a TF-IDF pipeline that needs a fraction of the compute and no GPU to serve. If the only criterion is maximum F1 and sufficient GPU resources are available, the Longformer is the strongest tested option. For a service that must fit a modest cloud plan, the classical TF-IDF pipeline offers a better cost-performance trade-off without giving up much accuracy. That is why the deployed model is neither of the two top scorers but TF-IDF with logistic regression at F1 91.28 %: LinearSVC scores marginally higher yet returns only a hard label, while the interface needs a probability to show how confident the verdict is.

Why did the merge operator outweigh the classifier? Absolute difference keeps the dimensions where two texts disagree, which is exactly the signal verification needs: two different authors should differ somewhere. Multiplication rewards shared features, and shared vocabulary says little about identity when both texts come from the same genre. The same logic explains why TF-IDF beat bag-of-words: down-weighting common words sharpens the author-specific tail that carries the style.

The two disappointments are worth naming. The integrated syntactic graph never matched the stylometric baselines; we attribute this to not reproducing the exact construction of Gómez-Adorno *et al.* (2016), and the representation was left untuned. The large language models failed outright. Huang *et al.* (2024) got past 88 % F1, but they used a strong proprietary model on an attribution task; our open models, chosen partly to avoid sending text to a third-party service, were not up to pairwise verification by prompting alone. We are not sure how much of that gap is model size and how much is the prompting format itself.

Set against the field, the best result here would rank near fifth among PAN 2021 entries, whose top score was 95.45 % from a five-classifier ensemble (Kestemont *et al.*, 2021). The comparison is loose, since that contest let systems leave hard cases undecided and we did not, but it places the work in a believable range. One detail stands out: no PAN 2021 team used a fine-tuned transformer head, so the Longformer result adds a data point the contest did not have.

The limits are real. Everything was measured on long English fanfiction (~21,000 characters), so short messages and other languages remain open, and there was no systematic out-of-distribution test. The transformer head also offers little interpretability, which matters in forensic and legal use where a verdict may need to be explained.

5 CONCLUSION

Across one corpus of 295,800 English pairs, a fine-tuned Longformer gave the best authorship-verification F1 at 92.38%, but classical TF-IDF stylometry came within a percentage point at a small fraction of the cost. For most deployments that trade-off points to the classical pipeline, which is what we shipped: TF-IDF with logistic regression, serving a probability through a React and FastAPI application on the Railway cloud. The choice of how to combine the two text vectors, absolute difference over multiplication, turned out to matter more than the choice of classifier. Open large language models prompted without training were not usable for this task, and the syntactic-graph route underperformed. Worthwhile next steps include pairing transformer embeddings with traditional classifiers, testing short-text and multilingual robustness, building a Czech corpus, and adding retrieval-grounded explanations so a verdict can be justified, not just stated.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Pokorný, 2026) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

ABBASI, A. *et al*. 2022. Authorship identification using ensemble learning. *Scientific Reports*. 12, 9537. https://www.nature.com/articles/s41598-022-13690-4

DEVLIN, J. *et al*. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. CoRR. *arXiv*:1810.04805 http://arxiv.org/abs/1810.04805

GÓMEZ-ADORNO, H. *et al*. 2016. Automatic Authorship Detection Using Textual Patterns Extracted from Integrated Syntactic Graphs. *Sensors*. 16(9), 1374. https://doi.org/10.3390/s16091374

HUANG, B. *et al*. 2024. Can Large Language Models Identify Authorship? *EMNLP*, 445–460. https://aclanthology.org/2024.findings-emnlp.26.pdf

KESTEMONT, M. *et al*. 2020. Overview of the Cross-Domain Authorship Verification Task at PAN 2020. CLEF. *PAN*. https://pan.webis.de/clef20/pan20-web/author-identification.html

KESTEMONT, M. *et al*. 2021. Overview of the Authorship Verification Task at PAN 2021. CLEF. *PAN*. https://pan.webis.de/clef21/pan21-web/author-identification.html

KOPPEL, M. *et al*. 2004. Authorship Verification as a One-Class Classification Problem. In: *ICML 04*, pp. 62. https://doi.org/10.1145/1015330.1015448

POKORNY, J. 2026. *Authorship verification of text using selected artificial intelligence methods* [in Czech]. Master's thesis, head: Trenz, O. Brno: Mendel University in Brno.

SAEDI, C. *et al*. 2021. Siamese networks for large-scale author identification. *Computer Speech & Language*. 70, 101241. https://doi.org/10.1016/j.csl.2021.101241

SAVOY, J. 2020. *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*. Cham: Springer Nature. ISBN 978-3-030-53359-5.

STAMATATOS, E. 2008. A Survey of Modern Authorship Attribution Methods. *Journal of the American Society for Information Science and Technology*. 538–556. https://doi.org/10.1002/asi.21001

STAMATATOS, E. *et al*. 2016. Authorship Verification: A Review of Recent Advances. *Research in Computing Science*. 9–25. https://doi.org/10.13053/rcs-123-1-1

TYO, J. *et al*. 2022. On the State of the Art in Authorship Attribution and Authorship Verification. *arXiv:*2209.06869. http://dx.doi.org/10.48550/arXiv.2209.06869

VASWANI, A. *et al*. 2017. Attention Is All You Need. *arXiv*:1706.03762 https://arxiv.org/abs/1706.03762

Contact information

Jiri Pokorny: e-mail: xpokor17@node.mendelu.cz
Michaela Veselá: e-mail: michaela.vesela.ui@mendelu.cz
Oldrich Trenz: e-mail: oldrich.trenz@mendelu.cz

PRODUCT NAME MATCHING

Jan Přichystal[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_028

ABSTRACT

The problem of product name matching belongs to the area of entity matching which is one of the data integration tasks. It plays an important role in the analysis of data available on the internet and it combines various techniques and data preprocessing. Gathering and analyzing customer reviews about products published on the internet is challenging task, because particular e-shops offer the products under a bit different names and it could be complicated to match identical products. The article aims at selecting the identical products identified by their names and introduces methods for product name matching based on machine learning using the combination of token-based and character-based functions.

Keywords: customer reviews, product name, entity matching, machine learning
JEL: L86

1 INTRODUCTION

The string-matching problem appears in many applications, e.g. systems for the automated detection of possible plagiarism, data cleaning, information retrieval or comparison of DNA/protein sequences. Such systems describe Wise (1993), Pandit *et al.* (2020) or Moussiades (2005). In this article we focus on finding similar strings in set of product names offered by e-shops. The motivation is we need to union the reviews of products published by the customers. There we try to identify discussed topics and opinion and help other customers to select optimal product without the need to read all customers reviews on the internet. Similar problem is discussed by Yao (2016). Methods of finding string similarities are listed in Deng (2013). Bergsma and Kondrak (2007) propose an alignment-based discriminative framework for string similarity. They gather features from substring pairs consistent with a character-based alignment of the two strings. Yu (2016) present an extensive set of algorithms for string similarity search and join. Wang (2011) propose a new similarity metrics, called "fuzzy token matching based similarity", which extends token-based similarity functions (e.g., Jaccard similarity and Cosine similarity) by allowing fuzzy match between two tokens. Kopcke *et al.* (2010) investigate the degree the machine learning helps semi-automatically determine suitable match strategies with a limited amount of manual training effort. The product pairing problem on web portal Zboží.cz discusses Bumba (2021). Their approach is based on usage of the CatBoost library developed by the machine learning team at Yandex. This library deals with the generation of

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

so-called gradient boosted decision forests. Generating decision forests is supervised machine learning method, so they need to have enough annotated data to train the model. Then they generate signals (features) from this data. The trained model is using 77 different signals, which arise from various comparisons of the attributes of the offer and the product candidate, such as title, description, product number, price, EAN, ISBN, manufacturer and others. Although the approach looks promising, the disadvantage is obvious – it is not automatic but supervised.

2 DATA

The data was downloaded from various web portals including Czech and international shops. We work with two types of collections. First collection contains the complete list of products in certain segment – the reference set. The second collection is the set of downloaded reviews for each product in the segment identified by the product name. To get the widest possible data set of products and reviews we focused on the segment of cell phones, digital cameras and lenses. The primary reference set containing the list of all introduced products was obtained from mobilmania.cz (cell phones), digicamdb.com (digital cameras) and dpreview.com (lenses). The reviews of products were downloaded from czc.cz, alza.cz, heureka.cz, fotoguru.cz, mobilmania.cz, michalkupsa.cz. The data was downloaded and saved to NoSQL document-based database MongoDB[1] because of different structures of web pages. Together with the product names and reviews we obtain the product properties such as weight, dimensions and others as described in example below. The idea was to use such properties to simplify the process of matching both sets. Unfortunately, we found the properties are not filled properly on many e-shops, we cannot rely on their accuracy and as a result they rather complicate the pairing process.

An example of record we save into the MongoDB database could be seen below:

```
{
        “name”: “Samsung Galaxy A10s”,
        “brand”: “Samsung”,
        “dimensions”: “157×76×7.8”,
        “weight”: “168”,
        “battery_capacity”: “4000”,
        “screen_size”: “6.2”,
        “screen_resolution”: “720×1520”,
        “introduced”: “2019”
 }
```

Problems with matching product names arise mainly because different sellers offer products under slightly different names. It is always a single product, but is identified by different names that omit or add certain product features. This problem could be expressed by the examples listed below. The first set shows the product “Samsung Galaxy S10+” different variants that could be found in various e-shops. The last two items show different products, but with similar product name, they can cause the confusion and wrong assignment. For example, the similarity of “Samsung Galaxy S10 Plus black” could be higher to “Samsung Galaxy S10 128 GB black” than “Samsung Galaxy S10+” which is in the reference set.

[1]https://www.mongodb.com/

- Samsung Galaxy S10+
- Samsung Galaxy S10 Plus
- Samsung Galaxy S10 Plus black
- Samsung Galaxy S10+ 128GB
- Samsung Galaxy S10+ Dual SIM černá
- Samsung Galaxy S10+, 8GB/128GB, černá
- Samsung Galaxy S10+ 128GB+8GB RAM SM-G975F/DS Dual Sim 6.4" LTE
- Samsung Galaxy S10 128 GB black
- Samsung Galaxy S10e

As can be seen in list above the product name consists of separate parts. If we focus on category of mobile phones the product name usually starts with the name of the manufacturer. In our example it is "Samsung". Then follows the product line – "Galaxy", the product itself – S10. Then there could be product model, here it is symbolized by "Plus" or "+". The vendors sometimes add more information about color – "black", memory size – "128GB" or "8GB/128GB", "Dual SIM", product code – "SM-G975F/DS", display size – "6.4"" or network capability – "LTE".

It is obvious the product name could vary and the same product may have significantly different product name and on the other hand the different products may have almost identical names. For example, if we use simple character-based function to compare such strings it finds higher similarity between "Samsung Galaxy S10+" and "Samsung Galaxy S10e" than between "Samsung Galaxy S10+" and "Samsung Galaxy S10+, 8GB/128GB, černá". This causes problems in identifying and matching products.

The example below shows the same problem again. The similarity of "Apple iPhone Xs 64 GB gold" will be higher to "Apple iPhone X 64GB gold" than to "Apple iPhone Xs".

- Apple iPhone Xs
- Apple iPhone Xs 64 GB gold
- Apple iPhone X 64GB gold

Sometimes we face the problem of identifying different product versions expressed by the year. The example below shows the different products (different versions for two years) but with the same name. The problem could appear when the e-shop doesn't reflect the strategy of Samsung company and uses only the name Samsung Galaxy A3. Then identifying of right customers reviews could be problematic.

- Samsung Galaxy A3 (2016)
- Samsung Galaxy A3 (2017)

Such problems are more emphasized in the digital camera lenses as could be seen in the example below where each item is different product. The last item is the general product name sometimes used in e-shops. Then the assignment is also complicated.

- Canon EF-S 18-55mm F4-5.6 IS STM
- Canon EF-S 18-55mm f/3.5-5.6 IS STM
- Canon EF-M 18-55mm f/3.5-5.6 IS STM
- Canon EF-S 18-55mm f/3.5-5.6 III
- Canon EF-S 18-55mm f/3.5-5.6 IS II
- Canon EF-S 18-55mm f/3.5-5.6 IS
- Canon EF-S 18-55mm f/3.5-5.6 II
- Canon EF-S 18-55mm f/3.5-5.6

The last example shows the list of digital cameras where the first item is particular product but sometimes used for the rest of products listed below. The names distinguish the evolution of product during the years.

- Sony Cyber-shot DSC-RX100
- Sony Cyber-shot DSC-RX100 II
- Sony Cyber-shot DSC-RX100 III
- Sony Cyber-shot DSC-RX100 VA
- Sony Cyber-shot DSC-RX10

3 METHODS AND RESULTS

Existing studies mainly use the following functions to quantify similarity of two strings (Wang, 2011):

- **Token-based similarity functions:** They first tokenize strings as token sets ("bag of words"), and then quantify the similarity based on the token sets, such as Jaccard similarity and Cosine similarity. Usually if two strings are similar, their token sets should have a large overlap. Token-based similarity functions have a limitation that they only consider exact match of two tokens and neglect fuzzy match of two tokens.
- **Character-based similarity functions:** They use characters in the two strings to quantify the similarity, such as edit distance which is the minimum number of single-character edit operations (i.e., insertion, deletion, and substitution) needed to transform one to another. In comparison with token-based similarity, edit distance is sensitive to the positions of the tokens in a string.

The similarity of strings could be also measured as the longest similar substring – **sequence--based similarity**. Such methods are longest common subsequence similarity, longest common substring similarity or Ratcliff-Obershelp similarity. (Mayank, 2019)

The algorithms and approaches mentioned in the introduction doesn't provide sufficient results. The problem appears in minor differences of distinct product names and major differences in names of the same product. Also, the constantly changing set of reference products plays important role and prevents the creation of training set.

We decided to propose new algorithm avoiding listed problems. The algorithm first identifies suitable candidates and then chooses the right one.

Before matching the product names, it is necessary to preprocess the data to increase the chance of match. We apply removing of the special characters e.g. commas or apostrophes, replacing the slash with space and lowering the case of text. The "Samsung Galaxy S10+, 8GB/128GB" is transformed to "samsung galaxy s10+ 8gb 128gb".

An algorithm itself consists of these steps:

(1) Finding 100% match – if the product name from review set is completely equal the algorithm assigns it directly.

(2) Tokenization and Sorensen-Dice (Sørensen, 1948) – the product names of both sets are tokenized by the spaces and the Sorensen-Dice score is calculated for each product from reference set. Only the products of certain company (e.g. Samsung) are selected. Top 5 products with the highest score are selected.

(3) Longest common substring similarity (LCSTR) – for the purpose of score determination the spaces in the string are removed to avoid false similarity. The longest match of the product

from review set and products selected in previous step is calculated. The product with the longest match is picked. If there are two or more products with the same longest match value, the product with higher Sorensen-Dice score is selected.

3.1 Results

The table below shows the partial results of process of identification of similar products. In the first column there is a title of product from the reference set saved in the database we use to match product name from downloaded review. The second column shows the similarity score of reference title and the name of product being compared calculated by Sorensen-Dice in 2. step. The third column shows the longest substring of product name from downloaded review matched to reference title in 3. step. The last column presents the score calculated in 3. step using LCSTR function. The product name from downloaded review we try to match is "Xiaomi Mi 8 Pro 128GB transparentní".

Reference title	Sorensen-Dice score	Result substring	Substring score
xiaomi mi 8 pro	0.80	xiaomimi8pro	12
xiaomi mi 8 128gb	0.80	xiaomimi8	09
xiaomi mi 8 lite 128gb	0.73	xiaomimi8	09
xiaomi mi 8	0.67	xiaomimi8	09
xiaomi mi 8 lite	0.60	xiaomimi8	09

The matched product title from reference set is therefore "Xiaomi Mi 8 Pro" – the first row in table with the highest substring score.

The second example shows matching of the product name downloaded from review "Tamron AF 18-200mm f3.5-6.3 Di II VC pro Nikon".

Reference title	Sorensen-Dice score	Result substring	Substring score
tamron 18-200mm f3.5-6.3 di ii vc	0.800	18-200mmf3.5-6.3diiivc	22
tamron 18-200mm f3.5-6.3 di iii vc	0.660	18-200mmf3.5-6.3diii	20
tamron 18-270mm f3.5-6.3 di ii vc pzd	0.625	0mmf3.5-6.3diiivcp	18
tamron 18-400mm f3.5-6.3 di ii vc hld	0.625	00mmf3.5-6.3diiivc	18
tamron af 18-200mm f3.5-6.3 xr di ii ld aspherical (if) macro	0.600	af18-200mmf3.5-6.3	18

The matched product title from reference set is therefore "Tamron 18-200mm f3.5-6.3 Di II VC" with the highest substring score.

4 DISCUSSION

The proposed algorithm has more than 90% efficiency when all items in reviews set are also in reference product set. If there appear the products that are not listed in reference set we face the problem with false positives. It means when the product is not listed another (with similar name) product is assigned. For example, to the "Aligator S6000 Duo" the "Aligator s5070 Duo" is assigned.

The algorithm was implemented using the Python programming language using the textdistance[2] library.

We still make experiments with the combination of methods and their settings. Now we focus to third step of our algorithm, where the longest common substring does not provide sufficient results. The aim is to have different weights for different parts of string, because similarity of certain parts of string is less important (e.g. "mm" or "GB").

The problem also appears in situation where the product specification influences the tokenization of product name e.g. "Xiaomi Mi9" vs "Xiaomi Mi 9".

We also experiment with the information gathered from e-shop aggregators like Heureka[3], where the list of different product names for each product can be found on individual e-shops.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education and the project EG16_084/0010315 ¬– Multicriterial Text Analysis Software (2017¬–2020, MPO/EG).

REFERENCES

BERGSMA, S., KONDRAK, G. 2007. Alignment-based discriminative string similarity. In: *Proceedings of the 45th annual meeting of the association of computational linguistics*, pp. 656–663.

BUMBA, M. 2021. Párování zboží pomocí CatBoost a praktické zkušeností ze Zboží.cz. *Root.cz*. [Accessed: 14 January 2021] https://www.root.cz/clanky/parovani-zbozi-pomoci-catboost-a-prakticke-zkusenosti-ze-zbozi-cz/. 2021

DENG, D. *et al*. 2013. Top-k string similarity search with edit-distance constraints. In: *2013 IEEE 29th International Conference on Data Engineering (ICDE)*. IEEE, pp 925–936.

KOPCKE, H., THOR, A., RAHM, E. 2010. Learning-based approaches for matching web data entities. *IEEE Internet Computing*. 14(4), 23–31.

MAYANK, M. 2019. *String similarity — the basic know your algorithms guide!* [Accessed: 11 November 2020] https://itnext.io/string-similarity-the-basic-know-your-algorithms-guide-3de3d7346227. 2019

MOUSSIADES, L., VAKALI, A. 2005. PDetect: A clustering approach for detecting plagiarism in source code datasets. *The computer journal*. 48(6), 651–661.

PANDIT, A. A., TOKSHA, G. 2020. Review of Plagiarism Detection Technique in Source Code. In: *International Conference on Intelligent Computing and Smart Communication 2019*. Springer, Singapore, pp. 393–405.

[2]https://pypi.org/project/textdistance/

[3]https://www.heureka.cz

SØRENSEN, T. *et al*. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *K. Dan. Vidensk. Selsk. Biol. Skr*. 5, 1–34

YAO, Y., SUN, A. 2016. Mobile phone name extraction from internet forums: a semi-supervised approach. *World Wide Web*. 19(5), 783–805.

YU, M. *et al*. 2016. String similarity search and join: a survey. *Frontiers of Computer Science*. 10(3), 399–417.

WANG, J., LI, G., FE, J. 2011. Fast-join: An efficient method for fuzzy token matching based string similarity join. In: *2011 IEEE 27th International Conference on Data Engineering*. IEEE, pp. 458–469.

WISE, M. J. 1993. *String similarity via greedy string tiling and running Karp-Rabin matching*. Online Preprint, Dec, 1993, 119(1), 1–17.

Contact information

Jan Přichystal: e-mail: jan.prichystal@mendelu.cz

SUMMARY OF FINANCIAL ARTICLES

Jan Přichystal[1], Tomáš Jakúbek[1], Roman Valovič[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_029

ABSTRACT

The Summary of Financial Articles (SFA) application is intended for obtaining articles in the field of financial markets published by various online media and their subsequent analysis. The aim is to facilitate the work of financial analysts, who have to process a large amount of information on a daily basis, use it to create a picture of the state of financial markets and individual securities, and then issue recommendations to customers on their trading strategy. Gathering the articles and news from online sources is very demanding process due to the technologies used there and ever-changing structure of these web sites. Not only the web application but also the API for integration into customer applications is available.

Keywords: natural language processing, machine learning, sentiment analysis, text summarization, SFA application, web application
JEL Code: L86

1 INTRODUCTION

Every day, financial analysts and advisors process a large amount of information from the financial markets and related areas, so that they can make timely and correct decisions, adjust their trading strategies or advise clients. Their daily activities include the analysis of published articles and reports. However, a large number of them are currently published on the Internet and it is practically impossible to read and analyze them all in a short time.

Summary of Financial Articles is a system designed to speed up the process of obtaining and processing information for these experts and thus facilitate their work. This system obtains articles, news or tweets from selected sources, especially in the field of financial markets, and analyzes their content. The aim is to offer a summarized view of individual documents in the form of several sentences summarizing the main ideas, to determine their sentiment and to provide aggregated data for specified time periods.

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

2 SYSTEM DESIGN

The application creates (articles and short reports) abbreviated summaries describing the essence of the text from the obtained documents and analyzes their sentiment. It regularly downloads documents and displays them in the web application in the form of a clear list. Users are not presented with full articles, but only a summary. It is available in several forms. If the author of the article summarizes the topic in the introduction, this section is displayed (Author tab). It is similar with a possible conclusion at the end of the article, which also often contains important information (Conclusion tab). However, the most important is the summary generated by the SFA application (Generated tab). It is created whenever an article contains enough text to create a summary. For example, this is not always necessary for short messages. The individual summaries are available clearly on separate tabs, so the user has the opportunity to easily choose what interests him.

For each document, there is also information about the date of publication, keywords characterizing the area covered by the report, and a link to the full article available through its title. It is also indicated whether it is an article (Analysis) or a short report (News). Individual documents can be filtered in the application according to:

- the date they were published (Date range),
- ticker – the names of the securities to which they relate (Tickers),
- tags – the area to which they belong (Tags),
- category – article or news (Category),
- sentiment – positive, negative or neutral evaluation (Sentiment),
- domains – web resources from which it is downloaded (Domains),

and sort by date ascending or descending.

The application also displays an overview of the sentiment of the obtained documents for a certain period in the header. In the left graph, it is possible to analyze the ratio of positive and negative messages. The right graph contains information on the most discussed tickers for the selected period, or those for which there was the largest change in the number of comments. In the initial settings, the most discussed are displayed regardless of sentiment. Using the Change button, it is possible to specify a sentiment and an overview of the most discussed with the given sentiment is displayed. For each ticker, a value is also displayed indicating whether the number has increased or decreased (positive/negative value).

2.1 Frontend implementation

The frontend and backend of the system are two standalone applications that communicate via thin REST API on the web server. Heavy frontend is delivered to client browser just once and the content of the website is then partially updated, in places that require change, directly in the browser. This concept is called Single Page Application (SPA) and can be implemented in various JavaScript libraries or frameworks. In our case, to deal with view layer in MVC software architecture, we chose ReactJS library.

2.1.1 Used technologies

ReactJS is still a popular open-source library that is used for building web (and mobile) composable user interfaces. In ReactJS, the user interface is broken down in a large number of reusable components. These components manage their states to handle both the logic and the view itself. Their hierarchical structure clearly defines unidirectional flow of data in the application and their APIs and lifecycle hooks describe the behavior.

The state of the component is changed after data mutation or by executing a user action and this update triggers a smart and controlled rendering process. In the background, ReactJS engine tries to improve performance by minimizing expensive manipulation with real DOM as much as possible, so it first renders the changes to the virtual DOM represented in JavaScript. Then its job is to reconcile the recent changes against the old one and only the necessary ones applied to the real DOM.

To get the best of ReactJS world, we used JSX (syntax extensions to JavaScript), TypeScript, functional components and React hooks. To manage application-level state we injected Redux into application. Redux is a predictable state container which plays a helping hand while developing large scale JavaScript application. To keep a similar look and feel to our previous apps, we selected Ant Design as the UI framework. This framework is known mostly in the East and does not follow ubiquitous and shabby material design.

2.1.2 User interface

We decided to render the individual article records into easily expandable card components. The grid design seems to be the best choice for large amounts of purely textual data (without any graphics). Each article can have up to three types of summaries (author, generated and author's conclusion), which are displayed as quickly switchable tabs component. The coloring of the cards at first glance corresponds to the sentiment of the article.

To enhance the user experience, article feed is updated automatically by infinite vertical scrolling and a potentially huge number of tags and tickers associated with the article are rendered in a horizontal virtual scrollbar. The application is completely responsive, including specific elements rendered only on smaller devices, and user settings are remembered in the browser storage.

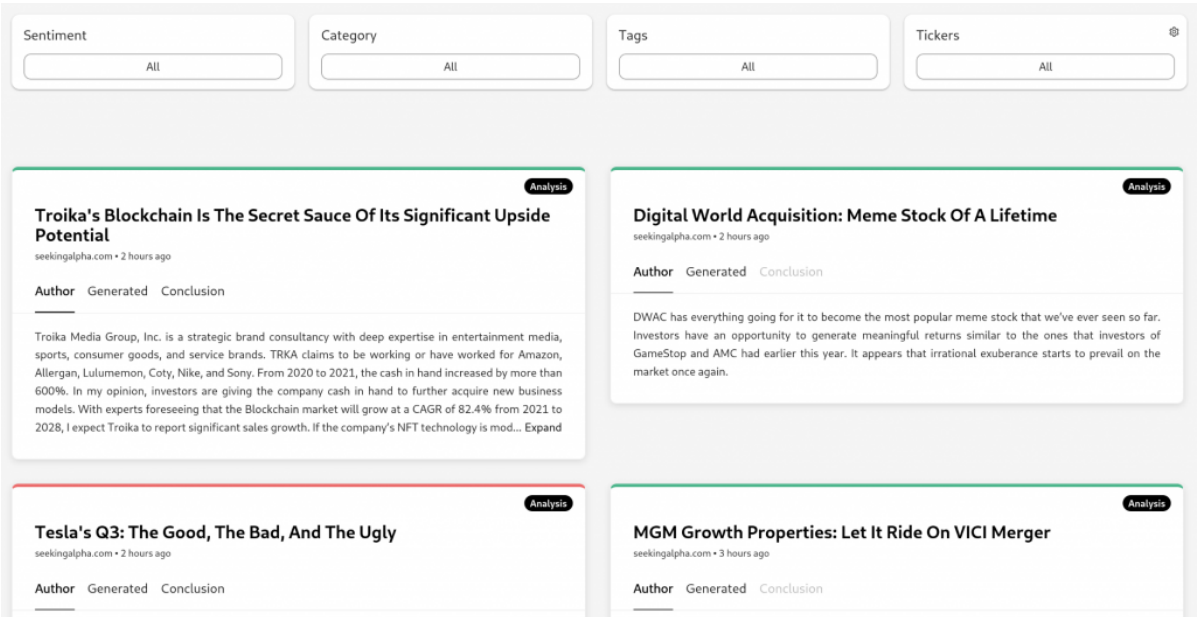


Fig. 1: Article feed with filters

2.2 Data scraping

Other parts of the system are web scrapers. Their task is to automatically retrieve documents from various internet sources. Specifically, these are newspaper articles and short reports from the financial markets.

The articles are fetched at regular intervals, which are appropriately set according to the frequency of publication. Around these scraping modules are therefore created complex dev-ops mechanisms in the form of various pipeline jobs and schedulers that combine the practices of continuous integration and continuous delivery. Scrapers store data in a well-defined structured format in our databases, from where it is consumed by text analysis in the following step.

Scrapers were created in two different fashions. One group of crawlers is implemented in the Apify framework and interpreted in the NodeJS runtime environment, the other makes use of the Scrapy library and the Python programming language. It is also possible to identify two different web scraping approaches prevalent in our solution. In the first one, the scraper starts on a certain web page and, in addition to data scraping, simultaneously collects new links to other documents or subpages. These are recursively visited later and the whole process is repeated. The second approach assumes that a list of URLs to be crawled is available in advance. In this case, the scraper does not look for subsequent links but just crawls the list. This approach was preferred when suitable sitemaps were available. Its significant advantage is the possibility of parallelism and concurrency. To provide incremental crawling functionality and ensure that the already crawled page won't be visited again we implemented middleware for URL filtering.

One of the most efficient ways of preventing access to a public website is IP address blocking. It is very common to encounter during web scraping. To solve this problem, we implemented a custom module to rotate the proxies. Proxy manager runs in parallel with scrapers, searches the internet for freely available proxies, automatically tests them and filters out dead ones. After the crawler detects that the page is blocking access or loading significantly slower, it automatically replaces the proxy with fresh one and repeats the action.

2.3 Application backend

Server-side applications provide several services such as web-scrape (see above), text summarization (SFA Summarization), sentiment classification (SFA Sentiment) and, most importantly for the user, provide a communication interface to the results of the data analysis (SFA Rest API).

The technological complexity and computing power requirements vary dramatically between applications; for example, the SFA Rest API, which responds to a client request with a preprocessed result stored in a database, has only minimal performance requirements on the machine on which the application runs. SFA Sentiment Analysis, on the other hand, provides outputs based on the computations of a complex BERT neural network.

However, the workload of each application varies throughout the day. While the Rest API is queried virtually constantly during normal business hours (8AM–4PM), Sentiment Analysis classifies a new batch of articles every hour. Not only from the point of view of application sustainability and robustness, but also from the point of view of cost-effectiveness, we did not implement the SFA system as a single monolith but opted for a "micro services" architecture. Although there is no strict definition of the term micro service, but we can talk about the framework that this architecture should fulfill:

- services communicate via HTTP (Fowler, 2015),
- services are independent of each other (Nadareishvili, 2016),
- services can be implemented using different programming languages, databases, environments and technologies, depending on the task they are supposed to solve (Chen, 2018),
- services are small, communicate only in a bounded context, autonomous, built and deployed by an automated process (Nadareishvili, 2016).

We implemented the fulfillment of these requirements using Docker technology. In Docker jargon, each service represents a so-called container. The container is a small simple standalone software package and as a monolith contains everything needed to run the application – runtime environment code, system libraries and configuration. Containers isolate software and the environment, ensuring platform independence. Containers can be shut down, restarted and changed independently without directly affecting the running of any other service.

It also allows us to run each service individually and thus efficiently optimize costs – e.g. by scaling performance or shutting down servers completely when the service is not in use. This would not be possible with a monolith architecture.

We provided the automated build requirement for the deployment using Github CI/CD, which provides tools to automate this process in addition to versioning the code. Thus, the developer's only concern is to upload code changes to the main repository. To minimize errors in the code, automated tests are run before each deployment to prevent changes from being introduced to the code if they fail.

2.4 Communication interface

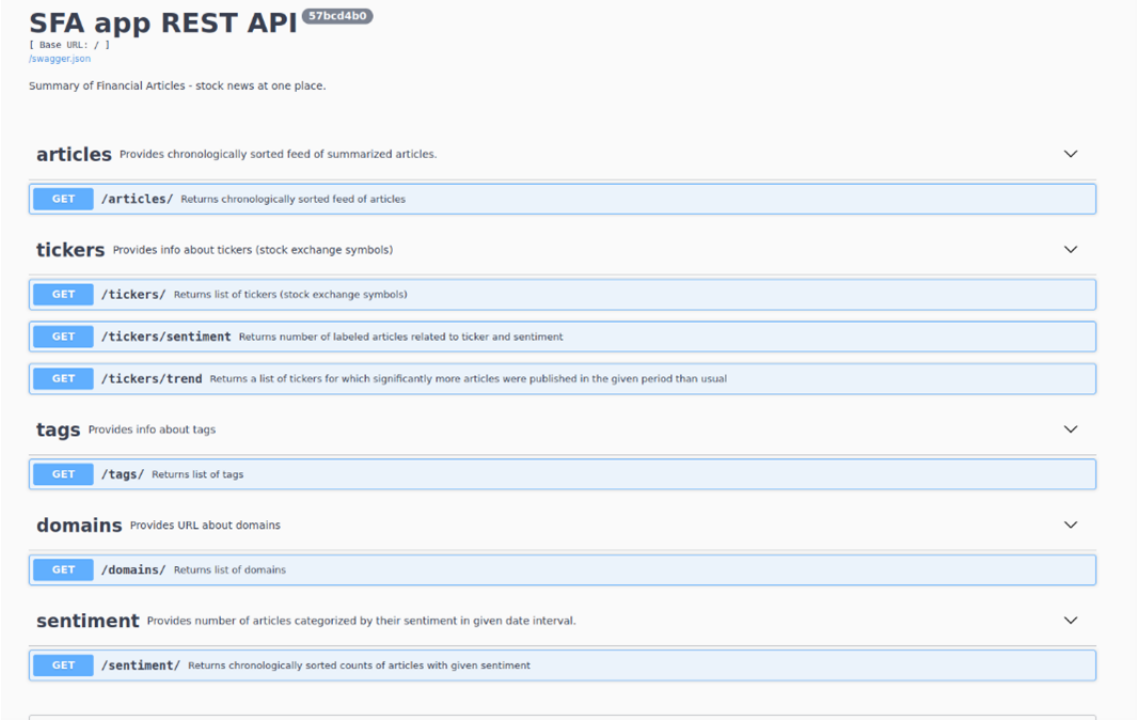


Fig. 2: Auto-generated documentation

The Flask framework[1] was used for development, which implements the usual best-practices in the field of web application development. It also enables fast deployment thanks to the WSGI communication interface, which is supported by an Apache web server. The framework works as a REST-API, i.e. that, based on the URI and HTTP method of the request performs action and returns the appropriate response, usually structured in JSON format. As a result, a single backend implementation can handle both web and mobile client requests.

During the development we put emphasis on the creation of documentation, which is often neglected resulting in a mismatch between the actual and documented functionality. To avoid this, we used tools that document individual API endpoints automatically. The result is the Swagger framework user interface (Fig. 2), which clearly specifies how one can interact with our REST API.

2.5 Data storage

Two different repositories are used for the storage of persistent data. Articles are automatically retrieved by scraping from various websites. Unfortunately, each site has a different structure, which also changes over time, so we cannot rely on the consistency of the data obtained. Therefore, in the first phase, the data is stored in the non-relational database MongoDB. Non-relational databases do not store data in tables, but as so-called documents. Each document then stores the data in key-value form, but the internal structure of the documents may not be uniform (i.e., each document may contain different keys).

In the second phase, the data is normalized and stored in the traditional PostgreSQL relational database. In this database, we store structured information about each article, such as the title, date of publication, author abstract and, of course, the textual content of the article.

To avoid having to modify the scraper every time the site structure changes, we implemented an intermediate step where we transform data from a non-relational MongoDB database into a relational PostgreSQL database. In addition to changing the structure, this intermediate step allows us to keep track of changes to the article content itself that may have occurred over time.

2.6 Text processing

For each article, the system performs 2 operations:

- • Summarization: the original long text is shortened by the extractive summarization technique to a paragraph containing a few sentences. The technique is based on identifying and selecting the most important parts from the original text.
- • Sentiment classification: for the summarized content of the article, the sentiment is evaluated into one of 3 classes: positive, negative or neutral. The sentiment of each news can be evaluated from different perspectives, for example, information about the decline in electric car sales will not please Tesla shareholders, but on the other hand, oil company shares will probably rise.

Our model classifies sentiment from the perspective of the investor who owns the stock, so the example above would be evaluated as negative.

[1] https://flask.palletsprojects.com/

2.6.1 Text summarization

We analyzed several different approaches from the point of view of both abstractive and extractive summary creation. We compared outputs of our own solutions and commonly available libraries (e.g. Sumy[2], Python Text Summarizer[3]) in Chochula (2021).

The summary, which contains essential information from the whole article, is created in an extractive way. The Gensim library from the Python programming language is used to create the summary. The extraction is based on the TextRank algorithm (Mihalcea and Tarau, 2004), which is a variant of the PageRank (Brin and Page, 1998) algorithm used by Google to evaluate the relevance of websites.

Algorithm process:

- preprocessing (stop words elimination, stemming),
- creating a graph, sentences forming vertices,
- edges indicating the similarity between two sentences in vertices,
- run the PageRank algorithm on this weighted graph,
- the highest score is selected and added to the summary,
- the number of vertices to be selected is determined based on the ratio or number of words.

Although we identified more successful algorithms in our research, unfortunately it is not possible to use them in the application. Their deployment prevents technical issues related to the outdated processor architecture on the dedicated computer.

We are currently working on our own solution that combines both approaches (abstractive and extractive) and takes advantage of both. The summary includes unchanged key information, which is very important for financial analysts and at the same time the sentences follow each other smoothly. We will offer this solution in the web application as an alternative summary.

2.6.2 Sentiment analysis

For sentiment analysis, we used several deep learning algorithms according to the state of the art in this field, such as BERT (and its derivatives) or LSTM (Šťastná, 2022). In general, the drawback of machine learning is the need for a large amount of labeled data to train and build the model. Sentiment analysis is a domain-oriented task; a model trained with data from a certain domain (e.g., movie reviews) usually performs poorly in another domain (e.g., financial news). On the other hand, data from the financial domain is limited and manual labeling is expensive and time-consuming.

Therefore, we collect data from available sources on the Internet. Specifically, we use the web scraper tool (described in Section 2.2) to obtain data from seekingalpha.com The articles on this site are categorized into “Bearish” (market declines) and “Bullish” (market rising) (see Fig. 3). In this terminology, we consider bullish news to be positive and bearish news to be negative. In total, we have collected more than 8.5 thousand articles that we can use to train the model.

We compared several algorithms and inputs for this task and tried to uncover how article length affects the accuracy of sentiment prediction. We attempted to classify sentiment for three types of inputs – article title only, article summary, and full article text. As shown in Fig. 4, all methods achieved the highest F1 score for article summaries, which suggests that the article title does not contain enough information; on the other hand, the raw article text is likely to contain unusable features that are not relevant for sentiment analysis.

[2]https://pypi.org/project/sumy/

[3]https://www.geeksforgeeks.org/python-text-summarizer/

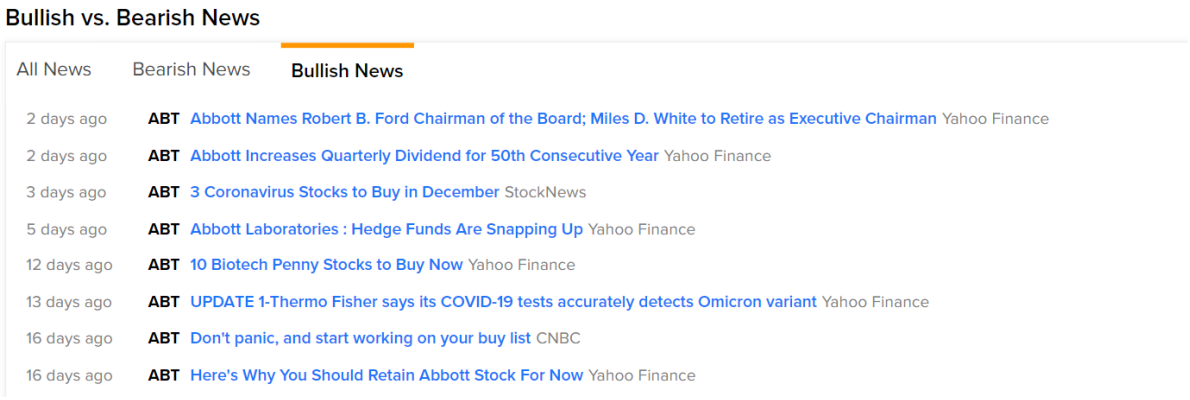


Fig. 3: Bearish and bullish news on Seeking alpha website

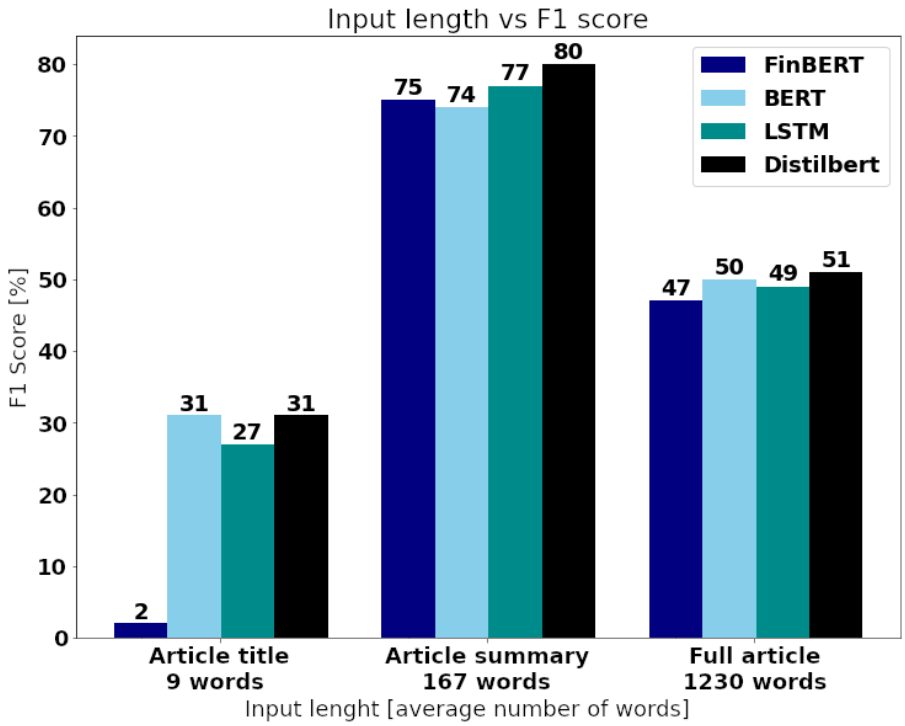


Fig. 4: Performance of methods depending on input length

2.6.3 Process pipeline

The whole process can be summarized as follows. Scrapers get a new batch of articles from selected sources. These are then stored in a non-relational MongoDB database since we cannot guarantee a fixed data structure at the time of download. Some aspects of the article are downloaded afterwards (list of related assets, tags etc.). Once the article is complete, the data is transformed into a relational Postgres database with a fixed structure.

The new records from the PostgreSQL database are taken as input by the summarization method, then the sentiment is classified. The results of both the processes are again stored in the Postgres database. At the last level, a presentation application figures, which clearly displays the analysis results in a user-friendly web browser environment.

Graphically, the pipeline is illustrated in Fig. 5.

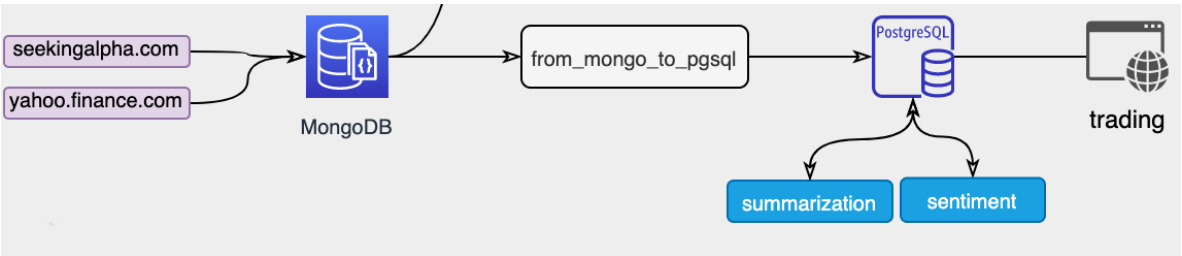


Fig. 5: Illustration of process pipeline

3 DISCUSSION AND CONCLUSIONS

The presented system is currently used by Cyrrus Corporate Finance, a.s. for their daily usage. Based on client requirements, additional information sources are regularly added and the portfolio of analyzed articles is still expanding. The list of long-term analyzes presented in the form of graphs is also under development. The goal is also to integrate the created API into a complex application. This will allow regular users to create their own business strategies, where one of the key attributes may be the sentiment of published articles

There are a number of similar projects available on the Internet that offer analysis of articles in the area of financial markets and securities trading. Sentimentrader provides daily sentiment report and ad hoc reports if there is anything especially timely or unusual. The Daily Sentiment Report includes an overview of where short- and intermediate-term sentiment is each day, along with updates on indicator extremes or studies focused primarily on sentiment, breadth and price action (Sentimentrader, 2021).

Sentdex (2021) is an application providing the sentiment analysis in Finance, Politics and Geographical area. The Financial analysis shows the sentiment for selected companies and long--term analysis od S&P 500 Index. The analysis is available for 7 days, 30 days, 6 months and 1 year interval. The stock price is available in presented chart too.

Stockgeist is an interactive platform for monitoring the current popularity of 2200+ publicly traded companies. It provides ranking based on the number of published messages about companies, stock overview, aggregated news from various websites, stock sentiment analysis based upon social media messages, wordcloud with the most popular topics the crowd is talking about, fundamentals such as market capitalization and P/E ratio, short story on the selected company to give a general idea of what they do and charts that display historical data about social media messages and positivity. Stockgeist (2021)

Stocktwits is a platform focusing on the same problematic. It shows what actual investors and traders are saying in real time about the stocks, crypto, futures and forex. (Stocktwits, 2021)

We can see from this overview the sentiment analysis is very important task in natural language processing and the financial analysts are really interested in information published in articles and news. We know from a discussion with our customer that even a summarized view of published texts is very important. However, other applications do not provide this functionality, which we consider an important advantage of our solution. The prototype of our application is available at https://mta.pef.mendelu.cz/trading/#/feed.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education and project EG17_176/0015722 – Data processing software (2020–2023, MPO/EG).

REFERENCES

BRIN, S., PAGE, L. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*. 30(1-7), 107–117.

CHEN, L. 2018, April. Microservices: architecting for continuous delivery and DevOps. In: *2018 IEEE International conference on software architecture (ICSA).* pp. 39–397. IEEE.

CHOCHULA, P. 2021. *Sumarizace dokumentů z oblasti finančních trhů*. Final thesis, Mendel University in Brno. https://is.mendelu.cz/lide/clovek.pl?id=71901;zalozka=7;zp=69217;studium=107183.

FOWLER, M., LEWIS, J. 2015. Microservices: Nur ein weiteres Konzept in der Softwarearchitektur oder mehr. *Objektspektrum*. 1(2015), 14–20.

MIHALCEA, R., TARAU, P. 2004, July. Textrank: Bringing order into text. In: P*roceedings of the 2004 conference on empirical methods in natural language processing.* pp. 404–411.

NADAREISHVILI, I., MITRA, R., MCLARTY, M., AMUNDSEN, M. 2016. M*icroservice architecture: aligning principles, practices, and culture*. O'Reilly Media, Inc.

Sentdex. 2021. http://sentdex.com/.

Sentimentrader. 2021. https://sentimentrader.com/.

Stockgeist. 2021. https://www.stockgeist.ai/.

Stocktwits. 2021. https://stocktwits.com/.

ŠŤASTNÁ, P. 2020. *Analýza sentimentu v oblasti financí a obchodování s cennými papíry*. Final thesis. Brno: Mendel University in Brno.

Contact information

Jan Přichystal: e-mail: jan.prichystal@mendelu.cz

NPL STUDIO

Jan Přichystal[1], Roman Valovič[1], Tomáš Jakúbek[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_030

ABSTRACT

This article deals with the description of NLP Studio software. This web application allows users to prepare their own textmining experiments for the natural language analysis without the need to use the programming language. There is a development environment in which the user can work in the form of visual programming. Preprocessing, training or visualization components are available. The software is designed primarily for teaching natural language processing and aims to help students navigate this area even if they do not have a deep knowledge of programming.

Keywords: natural language processing, web application, textmining, visual programming
JEL: L86

1 INTRODUCTION

The field of natural language processing is very much discussed nowadays and text analysis tools are used in many fields of human activity. However, there are not many experts who are able to perform text analysis using commonly available tools, and often they do not have sufficient domain knowledge related to the area in which the texts are published. On the other hand, domain experts often do not have knowledge of the necessary tools (currently often programming languages) and therefore find in-depth analysis of the texts under study difficult. Thus, the gaze of the professional community is currently turned towards tools that would facilitate the process of analyzing textual data in a simple way. One way may be to provide an accessible environment that allows experiments to be carried out at a higher level of abstraction without knowledge of programming languages. This may be a development environment using visual programming principles.

The idea of visual programming is not new, it was first used in programming teaching. Charntaweekhun and Wangsiripitak (2006) found that most students were more willing to accept flowcharts as the carrier of programming instead of writing codes directly. There are tools using the same concept mentioned in Boshernitsan (2004). The closest approach that is proposed in this study was proposed by VisComposer (Mei, 2018) which uses the visual programming interface for data transformations and layout construction. At the same time authors inherited the popular visualization tools disadvantages user interface heterogeneity, big number of additional panels with different settings and interaction models.

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

Ivanov (2020) presented the concept of visual programming data visualization authoring environment that completely relies on a node-based visual editor and allows the separation or reuse of any part of the workflow graph inside a single or multiple environments.

The main goal of this application is to simplify the process of creating textmining experiments especially for users not familiar with traditional programming languages but wanting to analyze the textual data.

2 SYSTEM DESIGN

The application is designed for creating experiments in natural language processing. A development environment using the concept of visual programming is available. This approach is intended to make it easier for users without knowledge of common programming languages to create their own tasks for analyzing text documents.

The application is divided into several parts. The main part contains the development environment, where the user places individual components (tasks) to solve specific tasks. These tasks are subsequently interconnected with each other through links to create a complex dataflow. Each component has defined inputs, outputs or both and performs a specific activity. From the broadest perspective, components can be divided into three groups:

- input component,
- transformation component,
- output (visualization) component.

Input components allow the user to specify the data source to work with. Examples of these sources include JSON, CSV, or plain text file format. Input data can be imported from existing files, from a document database located on the server where the application runs or defined manually. Currently, the set of input formats is limited, but in the future, it will be possible to connect to, for example, a relational database, a web resource containing text data, etc.

The core of the application is of course the transformation components, which further differ in their purpose depending on the natural language processing tasks they solve. Thus, these are primarily components providing preprocessing, classification, cluster analysis, etc. Each of these components allows the user to configure the parameters of a given task in the form of a clear menu. Each task also has specific requirements for the format of input and output data.

The task of the output components is to clearly display the results of analyses in the form of tables, graphs, maps or other formats. These components are usually the final element in the dataflow and do not allow the linking of other components.

The components are available in the left part of the application in a separate menu from where the user can select them as needed.

In the right part of the application there is a menu allowing modification of user data in the basic settings and in case a specific component is selected there is an option to make its settings. As such, the application allows you to manage the user accounts of registered users. The main point is that the user can save the experiments in progress and their results. A registered user has the possibility to work on multiple experiments and save datasets at will. An unregistered user is only allowed to create one experiment and is not allowed to save any data on the server where the application runs.

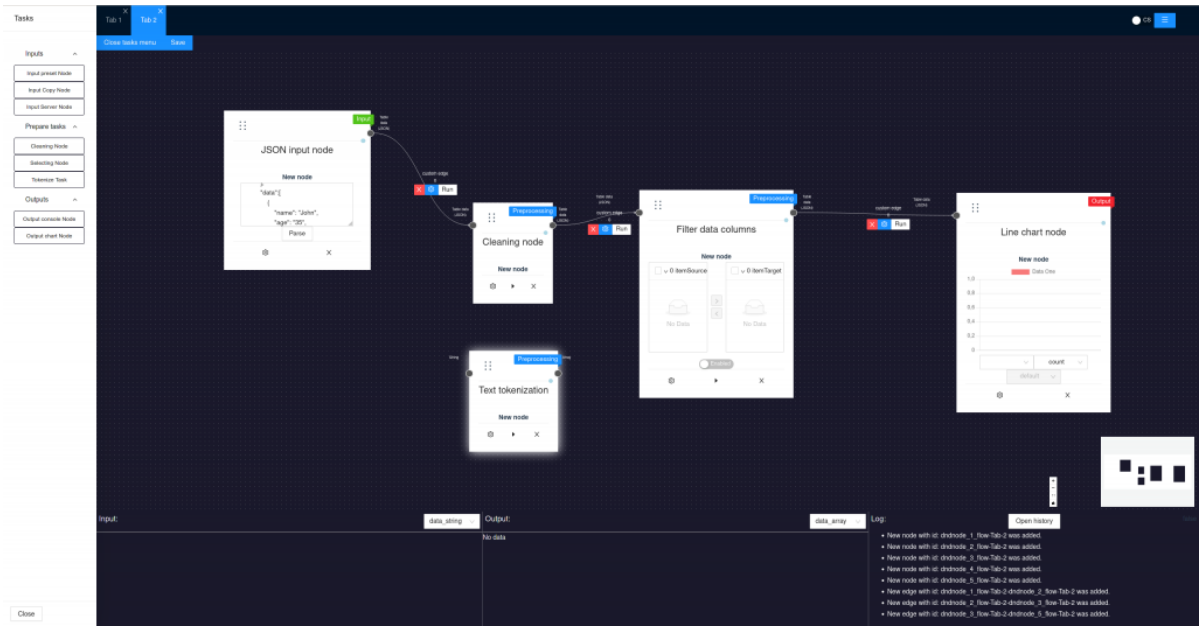


Fig. 1: Development environment
Source: own elaboration

2.1 Frontend implementation

The frontend is implemented as a standard single-page application. For this reason, a whole range of interactions and application logic is performed directly in the browser and the server faces significantly less load. Conversely, a significant load is transferred to the heterogeneously powerful end-user computers, which leads to a visible slowdown and decrease in the user experience if optimization is neglected. In applications with a one multi-interactive page (e.g., photo or text editor), this applies many times more. Therefore, with such an application concept, memoization, optimization of calculations, or avoidance of unnecessary renders play an important role.

This client application belongs to the category of so-called visual programming. The focus is oriented on visual objects and especially various relationships between them. Other features are concreteness, directness, explicitness and immediate visual feedback.

2.1.1 Used technologies

The implementation of the user interface is built on the JavaScript framework VueJS 3, which is used to create web UIs. This is the most up-to-date version that provides improved reactivity and change detection, a typed and better optimized core and a new Composition API syntax.

The key library is the open source vue-flow, which provides an engine for building interactive diagrams and complex node-based editors. It includes a workspace, graph controls and a dynamic viewport with custom elements. In this case, the elements are not rendered into the HTML canvas, but are normal DOM elements positioned dynamically using traditional CSS styles.

Communication between remote parts of the application is managed by the Pinia library, an alternative to the traditional Vuex. It is a global state manager that centralizes data, serves as the main source of truth and solves a programming anti-pattern called *props drilling*. The Ant Design framework provides the initial design of the component. Type checking from TypeScript makes the source code of higher quality.

2.1.2 User interface

The user interface is based on the *Canvas plus* palette design pattern. Its principle consists of a canvas on which nodes and edges are added, manipulated and their state changed in real time. A similar Microsoft Azure Machine Learning Designer application was designed with this idea in mind.

Nodes represent an input dataset, a processing task, or an output point. The user can define his own dataset or use one of the prepared ones. The growing list of supported tasks includes e.g. tokenization, lemmatization, POS tagger or stop-words removal. Edge between nodes means a connection enabling mutual communication and data transfer through the universal interface of input and output points.

The user can create multiple canvases within one workspace and solve different tasks at the same time. In the lower panel there is a visualization of the data of the active node and a log with last events. The user interface is currently localized into English and Czech languages and the new visitor is guided through it using quick onboarding. The layout of the entire flow is continuously saved in the *localStorage* of the browser, so the user does not lose his progress even after closing.

2.2 Application backend

The server part of the system is responsible for solving various tasks required by the user. From this perspective, the backend architecture can be divided into two levels.

The first level is the communication interface with the frontend and the database. Basically, it is a conventional web server that communicates using the REST API. At this level, the application receives and responds to requests, authenticates the user but also queues required tasks.

The second level is made up of a more complex, analytical part. The system is used for solving various NLP tasks and it is oriented towards processing large amounts of data, which is usually both time-consuming and computationally complex. Therefore, we implemented this part of the application as independent software. The analytic part picks up tasks from a queue and solves them based on the specified parameters in the background. It stores the intermediate state as well as the final result in a database. Thus, the first level is practically not involved in the actual task solving. If it did, the API would respond to requests with high latency. Additionally, such an arrangement does not allow for efficient task management or running in the background.

2.3 Communication interface

The Rest API is implemented in Python language in the modern FastAPI framework. One of the main reasons for choosing FastAPI is that it allows asynchronous handling of HTTP requests. Normally, Python executes code sequentially. For example, in the case of HTTP, this means that the server will only accept a new request when the previous request has been completely handled. However, asynchronous methods can be temporarily suspended (e.g. while waiting for data from the database), thus they do not block other, so-called competing requests waiting in the queue to run (Hattingh, 2020). Thus, the asynchronous paradigm significantly reduces the time required to handle a request and is well suited to a concurrent task environment such as a web server. The tasks are handled by the analytical part of the application at the second level, as we mentioned in the introduction of this chapter.

The communication is initiated by the client, by sending a POST request to the API endpoint/task. The endpoint accepts requests for all task types. The specific specification and parameters are sent in the request body in JSON format. For example, the following request illustrated in Figure 2 invokes the “Word tokenize” task over the dataset ID 622, which is stored in the database. A request body can contain any number of parameters depending on the type of task.

```
{
    "input_data_id": "622",
    "task": "word_tokenizer",
    "lang": "english"
}
```


Fig. 2: Example of task request data
Source: own elaboration

Instead of executing the request directly, i.e., running the corresponding method, the API creates a new object for the task and inserts a new record in the Mongo database containing the necessary metadata. The API returns the ID of the created task to the user, but the result is processed asynchronously in the background.

The running of the task itself is handled by Celery, an implementation of a queue of tasks which asynchronously executed in the background. However, Celery requires a message transport system to send and receive messages between the client and the so-called worker (the thread that executes the code itself). For these purposes, we used Redis, in which the API stores the job record. This is an important of system since the API and the analytics layer works independently. Redis, thus, acts as a communication interface between them. The data flow is illustrated in Figure 3.

Once the job is completed, Celery writes the result to MongoDB. The client can retrieve the result by addressing the API at the appropriate endpoint, as it is shown in Figure 3. In this case, we no longer require Celery since it is a simple process, which only reads records from the database.

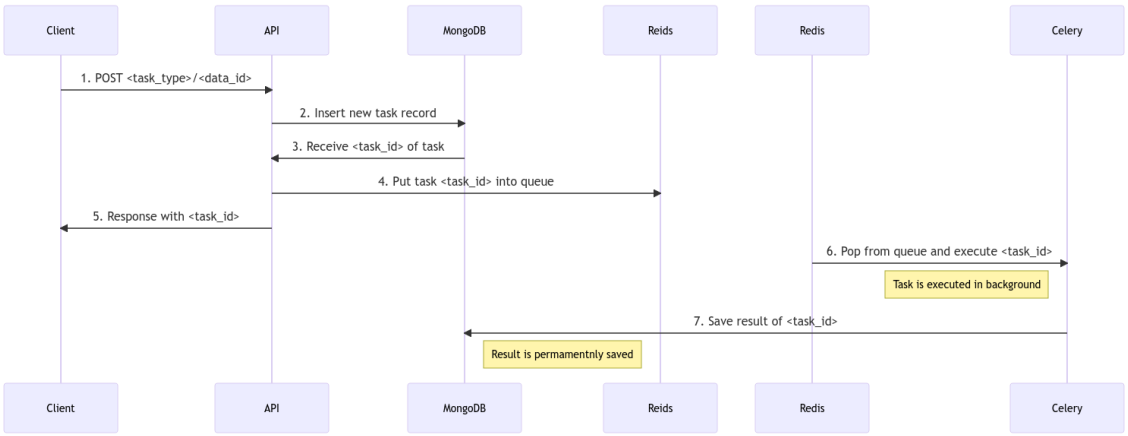


Fig. 3: Flowchart for processing a task request
Source: own elaboration

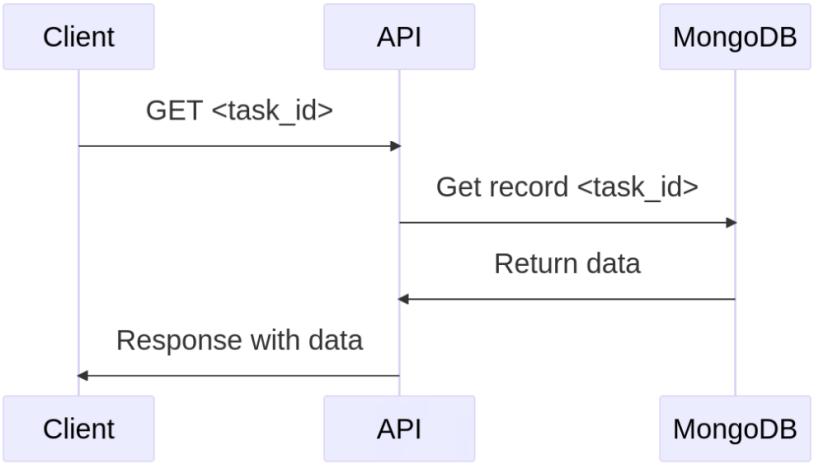


Fig. 4: Flowchart for processing a task request
Source: own elaboration

2.4 Data storage

Similarly, to the application layer, the data layer can be divided into two levels. Permanent data storage is provided by the non-relational Mongo database.

Temporary data, related to the task passing between the API and Celery is handled by Redis technology.

One of the key aspects of the system is the processing of data based on the user's needs. In order to provide the best user experience, we do not require any data structure that the user must follow in order to interact with the system. The user can upload their data stored in a text file in txt (plain text), JSON, or CSV format. The API recognizes the uploaded file's format and automatically executes the appropriate parser. To be able to work with the data regardless of its input format, the parser converts content of each file into a uniform internal format that the various system components are prepared for. An example of converting a CSV file to an internal format is illustrated in Figure 5. Content of each uploaded file is labeled by an ID and name as user's reference.

The internal format is designed so that the data is structured and stored in a table. This makes it possible to use popular libraries for working with tabular data such as Pandas or Numpy. Moreover, it gives the user the ability to use features that he is used to from common office applications – column removal, numeric operations on column values, row filtering, transformation of data types, etc.

The API always returns the processed data in JSON format, however, in the future we plan to extend the output to other formats.

```
Survived,Pclass,Name,Sex,Age,Siblings/Spouses Aboard,Parents/Children Aboard,Fare
0,3,Mr. Owen Harris Braund,male,22,1,0,7.25
1,1,Mrs. John Bradley (Florence Briggs Thayer) Cumings,female,38,1,0,71.2833
1,3,Miss. Laina Heikkinen,female,26,0,0,7.925
1,1,Mrs. Jacques Heath (Lily May Peel) Futrelle,female,35,1,0,53.1
```

```
{
 "_id": "63d7d2960b49cf2b9925e446",
 "name": "titanic_passengers",
 "payload": [
  {
   "Survived": 0,
   "Pclass": 3,
   "Name": "Mr. Owen Harris Braund",
   "Sex": "male",
   "Age": 22,
   "Siblings/Spouses Aboard": 1,
   "Parents/Children Aboard": 0,
   "Fare": 7.25
  },
  {
   "Survived": 1,
   "Pclass": 1,
   "Name": "Mrs. John Bradley (Florence Briggs Thayer) Cumings",
   "Sex": "female",
   "Age": 38,
   "Siblings/Spouses Aboard": 1,
   "Parents/Children Aboard": 0,
   "Fare": 71.2833
  },
```

Fig. 5: Transformations of CSV (upper) to internal format (below)
Source: own elaboration

3 DISCUSSION AND CONCLUSIONS

Of course, the presented application is not the only one currently available. There are similar projects available on the Internet that offer visual programming environments for machine learning or similar projects. The closest representatives of this type of applications are e.g. Azure ML Studio (2023), BotX (2023) or Datablocks (2023).

Azure Machine Learning Studio (ML Studio) is a graphical, collaborative, drag-and-drop web interface from Microsoft for building machine learning models. It does not require coding knowledge of R or Python, and the cloud interface can be used to build models through ML Studio. ML Studio lets user to train machine learning models, evaluate trained model performance, launch new experiment and view the results of past experiments, deploy trained models, monitor deployed models, track datasets and models, provision and manage compute resources, conduct data science experiments in online notebooks. The most important part of ML Studio is the Designer a drag and drop tool to create ML workflow from the input of a dataset to the evaluation. It is easy to use and has many features which can be searched via text or found by clicking on the arrows and captions on the left. For the Designer, there are currently more than 60 Modules that can be divided into data

transformation, data input and output, feature selection, statistical functions, regression, clustering, classification, model training, and model evaluation/scoring. The whole machine learning set is covered by the designer modules, so the user can build models without coding. He can simply choose the modules and connect them and then get the pipeline ready to be tested and interfered. The designer is attractive to people who would like to think at a higher level and don't want to worry about specific Python or R code for executing standard tasks. However, users still need to know a lot of machine learning concepts as far as normalizing data, feature engineering, splitting data for training and validation, and various machine learning algorithms. Additionally, it's worth noting that there are code nodes for both Python and R code that users can include into a machine learning pipeline. These nodes allow user to execute custom code in a machine learning pipeline for certain steps while still taking advantage of the drag and drop and visual nature of the designer (Hucker, 2021), (Eland, 2022).

The Datablocks is a node based editor with several blocks for processing data. It's based on our open source library React Flow. The editor is divided into four parts. On the left user has the sidebar with the different blocks. In the center there is the editor pane and at the bottom there is the data output view and a terminal.

The blocks are divided into the groups Input, Transform, Geo, Visualization and Misc. The input group gives user blocks to read in data. Either by pasting text content or by file upload. Currently, datablocks can handle json, csv, topojson and geojson files. There is also an example data block for testing and playing around. The transform blocks handle tabular data. Users can filter data entries or merge different data sets for example. The geo blocks help them to simplify data or convert geo formats. With the visualization blocks you can create simple visualizations like a histogram or a bar chart. The misc section contains markdown and a simple statistics block for now.

The editor pane is the place where users are working most of the time. On the editor pane user can connect different blocks to create a data flow and you can configure the blocks. In the current version all configuration happens in the blocks. At the bottom the user can find the data output view. This view shows the output of the selected block. It can display tabular data, JSON objects and geojson data. Users can also export data as csv, json or geojson. In the right bottom corner is the terminal. It is used to display application errors or logs from the Javascript block (Datablocks, 2023).

This overview reveals the development of experiments in the area of machine learning and AI without coding is a very important challenge. This and similar areas related to data analysis are being addressed by several commercial companies that see an opportunity to offer solutions for ordinary users on the market. As we know from discussions with our students, the small knowledge of programming languages limits their ability to analyze the various types of data including the textual. The applications discussed above offer only limited functionalities for the free users. This is an opportunity to apply our application, which is free and offers all functions without limitations. Now, there is a noticeable difference in the number of functions offered between the application we have developed and commercial solutions. However, intensive development and communication with potential users will allow us to offer a competitive solution that will stand up to comparison with others in the foreseeable future.

The prototype of our application is available at https://mta.pef.mendelu.cz/lab/.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education and the project EG17_176/0015722 – Data processing software (2020–2023, MPO/EG).

REFERENCES

Azure ML Studio. 2023. https://ml.azure.com

BOSHERNITSAN M., DOWNES M. S. 2024. *Visual programming languages: A survey*. Computer Science Division, University of California.

BotX. 2023. https://www.botx.cloud/.

CHARNTAWEEKHUN, K., WANGSIRIPITAK, S. 2006. Visual programming using flowchart. In: *2006 International Symposium on Communications and Information Technologies*, IEEE, pp. 1062-1065.

Datablocks. 2023. https://datablocks.pro/.

ELAND, M. 2022. *Azure ML Studio: Getting Oriented with Azure Machine Learning*. https://dev.to/integerman/azure-ml-studio-getting-oriented-with-azure-machine-learning-5cog

Hattingh, C. (2020). Using Asyncio in Python: Understanding Python's Asynchronous Programming Features. O'Reilly Media Inc., 2020, ISBN 978-1492075332.

HUCKER, M. 2021. *A Brief Introduction to Azure Machine Learning Studio*. https://towardsdatascience.com/a-brief-introduction-to-azure-machine-learning-studio-9bbf41800a60.

IVANOV, E., KHOROSHAVIN, A., KARSAKOV, A. 2020. Visual programming environment based on data visualization grammar specification. *Procedia Computer Science*. 178, 434–439.

MEI, H., CHEN, W., MA, Y., GUAN, H., HU, W. 2018. Viscomposer: A visual programmable composition environment for information visualization. *Visual Informatics*. 2(1), 71–81.

RAY, P. P. 2017. *A survey on visual programming languages in internet of things/Scientific Programming*.

SAABITH, A. S., FAREEZ, M. M. M., VINOTHRAJ, T. 2019. Python current trend applications an-overview. *International Journal of Advance Engineering and Research Development*. 6(10).

Contact information

Jan Přichystal: e-mail: jan.prichystal@mendelu.cz

FACTORS INFLUENCING THE SELECTION OF CRAFT BEER: A FACTOR ANALYSIS AND GENERATIONAL SEGMENTATION IN THE CZECH MARKET

Lenka Procházková[1], Marek Záboj[1], Matěj Sysel[1]

[1]Department of Marketing and Trade, Faculty of Business and Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_031

ABSTRACT

This article presents the results of an empirical study of consumer behavior in the Czech craft beer market. The primary objective was to identify the key factors influencing the selection of craft beer and, based on these factors, to perform a market segmentation analysis. The research was conducted through a quantitative survey combining online distribution via social media platforms and offline distribution using QR codes placed in craft beer pubs in Brno. After the application of screening criteria, a total of 458 respondents aged 18 to 57 years were included in the statistical analysis. Ten formulated hypotheses were tested using non-parametric statisti-cal methods (Pearson's chi-square test, the Mann–Whitney U test, the Wilcoxon signed-rank test, and the Kruskal–Wallis test). The principal analytical method was exploratory factor analysis with Varimax rotation, which extracted four factors explaining 43.45% of the total variance: price sensitivity and conservatism, taste and sensory attributes of beer, authenticity and exclusivity, and social and visual influences. Taste and aroma emerged as the most influ-ential individual determinants of craft beer selection, whereas recommendations from influ-encers were found to be the least influential. Generational segmentation revealed that Genera-tion Y constitutes the most active market segment, while Generation Z consumes craft beer significantly less frequently. The findings suggest that value-based attributes, such as authen-ticity and regional origin, outweigh price sensitivity and traditional forms of advertising among craft beer consumers in the Czech Republic.

Keywords: consumer behavior, craft beer, factor analysis, market segmentation, authenticity, storytelling, generational differences, Czech Republic
JEL Code: M31, D12, L66

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

In recent years, the craft beer market has become one of the most dynamically growing segments of the European brewing industry. In the Czech Republic, the number of craft breweries has nearly tripled since 2013, reaching a total of 508 microbreweries by the end of 2023 (Czech-Moravian Association of Microbreweries, 2025). This growth reflects a broader global phenomenon known as the craft beer revolution, which originated in transformations within the American brewing industry in the mid-1960s and subsequently spread to other regions of the world. The defining characteristic of this transformation is not production volume but rather an emphasis on quality, authenticity, and regional identity of production (Wojtyra *et al.*, 2020).

The growing popularity of craft beer is accompanied by a paradox of globalization. Although consumers have easy access to beer brands from around the world through global distribution networks, demand is increasingly directed toward regional products (Villacreces *et al.*, 2022). This shift reflects changing consumer preferences, particularly among younger generations, who seek authentic, distinctive, and high-quality products. At the same time, industrially produced beer maintains a strong market position primarily due to its lower price and widespread availability, placing small and independent breweries in a competitive environment where they must differentiate themselves from mass-produced alternatives through value-based attributes.

A key distinguishing feature between craft and industrial beer lies in the production process. Craft breweries are typically small and independent enterprises that experiment with ingredients and brewing techniques to achieve unique flavor profiles, whereas industrial breweries focus on standardization and large-scale production (Rosales *et al.*, 2021). Sensory, chemical, and process-related differences between the two categories are comprehensively documented in the review by Baiano (2021).

Within the Czech Republic, the beer market has long represented an important component of both the national economy and cultural heritage. In a study focused on the establishment of craft breweries, Březinová (2021) identified the motivations of brewery founders as well as the structural characteristics of this market segment. Svatošová *et al.* (2021) examined broader factors influencing consumer behavior in the Czech beer market and highlighted, among other findings, reduced price sensitivity among craft beer enthusiasts. International studies focusing specifically on craft beer attributes, such as Lerro *et al.* (2020), employing the Best–Worst Scaling method, identified taste, bitterness, ingredient composition, and regional origin as the most important product attributes. In the Italian context, Pilone *et al.* (2023) demonstrated that consumer preferences vary according to the level of market knowledge and familiarity with the craft beer segment.

From the perspective of psychological mechanisms underlying consumer decision-making, the craft beer segment is strongly influenced by the principles of authenticity (Newman, 2019; Bartsch *et al.*, 2022), the storytelling effect (Ahmed *et al.*, 2024), and the scarcity principle (Goldsmith *et al.*, 2020). Simultaneously, the demographic composition of beer consumers is changing. Livingston *et al.* (2024) and Caluzzi *et al.* (2024) document a gradual decline in alcohol consumption among the youngest generations, a trend that may lead to long-term structural changes within the market.

Although individual aspects of the craft beer market have been examined in both Czech and international literature, a comprehensive study integrating the identification of key selection factors, the testing of psychological effects, and generational market segmentation has thus far been absent in the Czech context. The primary objective of the present research was to identify the factors influencing consumer behavior in the selection of craft beer in the Czech Republic and, based on these factors, to perform a segmentation of this market. Secondary objectives included describing consumer attitudes and preferences and formulating recommendations for business entities operating in the craft beer sector.

2 METHODOLOGY AND DATA

The empirical study was conducted as a quantitative survey targeting beer and craft beer consumers in the Czech Republic. The questionnaire consisted of 28 questions organized into four sections: consumption frequency and purchasing locations, factors influencing craft beer selection, perceptions of promotional forms and marketing elements, and demographic characteristics. Seventeen questions were closed-ended, two were open-ended, and six item batteries comprised a total of 40 statements evaluated using a seven-point Likert scale (1 = completely unimportant, 7 = very important). Prior to the official launch, a pilot study was conducted, and questions that respondents found difficult to understand were reformulated.

The questionnaire was distributed using a mixed-mode approach. The online version was disseminated through Google Forms via social media platforms, particularly Facebook groups focused on brewing and beer-related interests. An additional offline distribution channel consisted of flyers containing QR codes placed in craft beer pubs located in the center of Brno. Before completing the questionnaire, respondents were informed about the purpose of the study, its anonymous nature, and the distinction between industrially produced and craft beer.

A total of 506 completed questionnaires were collected. To be included in the main analytical sample, respondents had to meet three criteria: be at least 18 years of age, consume beer at least occasionally, and consume craft beer at least occasionally. Thirty-three respondents failed to meet these criteria. An additional 15 respondents aged 58 years and older were excluded due to the non-representatively small size of this age group. Consequently, the final analytical sample consisted of 458 respondents. Generational segmentation was based on the classification developed by the research agency STEM/MARK (2024), distinguishing Generation X (43–57 years), Generation Y (28–42 years), and Generation Z (18–27 years). The quota sampling procedure was designed to achieve a relatively balanced distribution across generations.

Data analysis was performed using Statistica 14 and Microsoft Excel. Ten research hypotheses were formulated and tested using four non-parametric statistical methods selected according to the nature of the variables: Pearson's chi-square test, with the strength of association assessed using Cramér's V; the Mann–Whitney U test for comparisons between two independent groups; the Wilcoxon signed-rank test for paired observations; and the Kruskal–Wallis test for comparisons across multiple groups, followed by post hoc analysis with Bonferroni correction. The hypotheses focused on relationships between beer-type preference and income, purchasing location, consumption frequency, generational cohort, and selected value-related factors (regional origin, environmentally friendly packaging, brewery history, and limited-edition products). The significance level was set at $\alpha = 0.05$. Prior to applying Pearson's chi-square test, assumptions regarding expected cell frequencies were verified according to the recommendations of Bolboacă *et al.* (2011); when these assumptions were not met, adjacent categories were merged. The use of non-parametric methods was motivated by the ordinal nature of the data and the violation of the normality assumption (Field, 2013).

To achieve the primary research objective—namely, identifying the factors influencing consumer behavior—an exploratory factor analysis (EFA) was conducted. The analysis included 31 variables comprising both factor ratings obtained from Likert-scale items and attitudinal statements. Principal component analysis (PCA) with normalized Varimax rotation was employed. The number of extracted factors was determined using the scree plot criterion. Variables with factor loadings of at least 0.60, or those falling slightly below this threshold, were considered relevant indicators of a given factor.

Market segmentation was originally intended to be conducted using cluster analysis. However, due to the insufficient distinctiveness and substantial similarity of the resulting clusters, this approach was rejected. Instead, segmentation was performed based on generational cohorts, which proved to be more interpretable and more accurately reflected differences in consumer behavior.

3 RESULTS

The structure of the research sample corresponded to the demographic profile of craft beer consumers in the Czech Republic. Of the total 458 respondents, 73.9% were men and 26.1% were women. The generational distribution was relatively balanced: Generation Y (28–42 years) was represented by 175 respondents, Generation Z (18–27 years) by 166 respondents, and Generation X (43–57 years) by 153 respondents. Nearly half of the respondents (47.3%) had attained a university degree, while employment was the predominant economic status (67%). In terms of net monthly income, the largest proportion of respondents (25.5%) belonged to the CZK 30,001–40,000 income category.

The survey revealed that 95.5% of respondents consume beer and nearly 93% consume craft beer. When both industrial and craft beer are available in a hospitality venue, 83.4% of respondents choose craft beer, although only 54.8% identify it as their primary preference. The most common place of purchase for craft beer is restaurants and pubs (36.6%), followed by breweries (22.4%) and specialized beer stores (pivotéky). The most frequently reported consumption occasion is informal socializing with friends, cited by 51.7% of respondents.

3.1 Relationship Between Income and Beer Preference

The hypothesis concerning the association between beer-type preference (craft vs. industrial beer) and net personal income was tested using Pearson's chi-square test. The analysis revealed a statistically significant relationship ($\chi^2 = 19.902$; $df = 6$; $p = 0.003$), with a weak-to-moderate effect size (Cramér's $V = 0.220$). Respondents with lower incomes (up to CZK 30,000) were more likely to prefer industrial beer, whereas respondents with incomes exceeding CZK 40,000 exhibited a substantially stronger preference for craft beer.

3.2 Generational Differences in Consumption Frequency

The Kruskal–Wallis test demonstrated statistically significant differences in median craft beer consumption frequency across generations ($H(2, n = 469) = 66.120$; $p < 0.001$). Post hoc analysis with Bonferroni correction confirmed that significant differences existed between Generation Z and both older generations, whereas no significant difference was found between Generations X and Y. Generation Z consumes craft beer significantly less frequently than older generational cohorts.

3.3 The Storytelling Effect

The Mann–Whitney U test was used to compare the perceived importance of brewery history between two groups of respondents: those who identified an engaging brewery story as the most persuasive promotional element ($n = 200$) and all remaining respondents ($n = 258$). The analysis revealed a statistically significant difference ($Z = 5.731$; $p < 0.001$). This finding may be interpreted

as empirical support for the storytelling effect within the craft beer market. Respondents who are more receptive to narrative-based marketing also attribute greater importance to brewery history in their purchasing decisions.

3.4 Evaluation of Selection Factors

The assessment of 40 selection factors using a seven-point Likert scale indicated that taste and aroma constitute the most important factor influencing craft beer choice (mean = 6.11; median = 7; SD = 1.43). Other highly rated factors included originality and uniqueness (mean = 4.98), traditional recipes, ingredient composition, regional origin, and support for local producers. In contrast, recommendations from influencers and celebrities (mean = 1.94), the availability of a non-alcoholic craft beer version, and environmentally friendly packaging were identified as the least important factors.

3.5 Factor Analysis

The scree plot of eigenvalues (see Figure 1) supported the extraction of four factors, which together explained 43.45% of the total variance. The structure of the extracted factors and their principal items are summarized in Table 1.

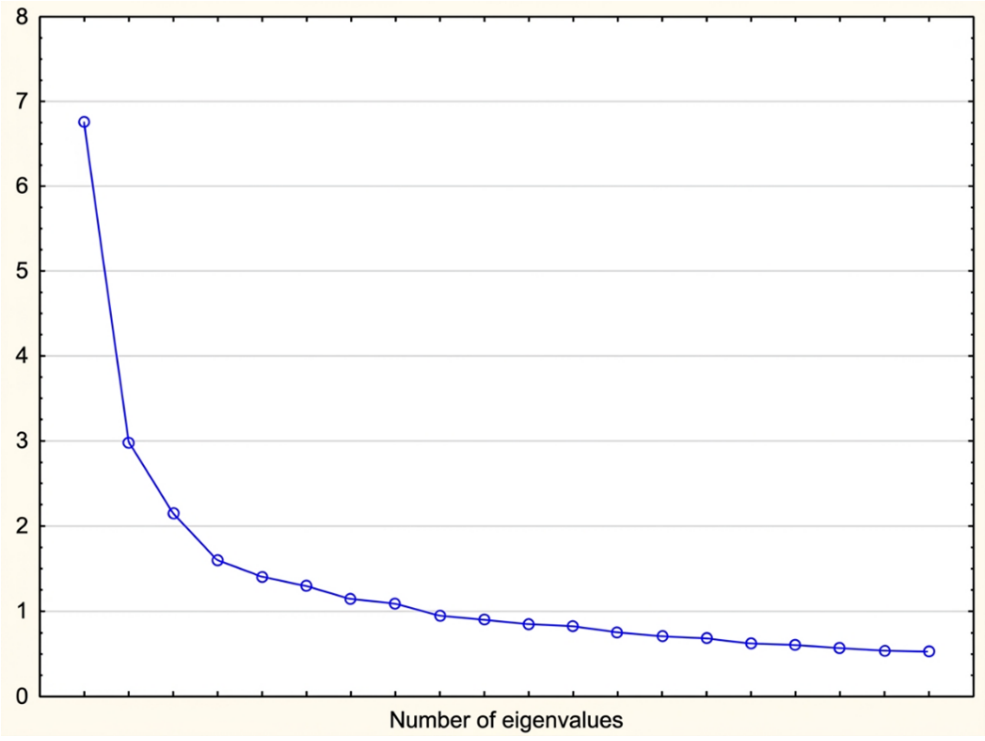


Fig. 1: Scree Plot of Eigenvalues Obtained from the Exploratory Factor Analysis
Source: Questionnaire survey data (n = 458).

Tab. 1 Extracted Factors and Their Key Items with Factor Loadings

Factor	Key Items (Factor Loadings)	Explained Variance
F1 – Price Sensitivity and Conservatism	price (0.672); I choose beer based on price (0.703); I prefer well-known and easily available beers (0.649)	21.77%
F2 – Taste and Sensory Attributes	taste and aroma (0.639); I select beer carefully (0.711); I prefer bitter beers (0.649); I prefer a strong and full-bodied flavour (0.712)	9.62%
F3 – Authenticity and Exclusivity	regional origin (0.709); support for local producers (0.638); seasonal offerings (0.626); limited editions (0.590)	6.90%
F4 – Social and Visual Influences	packaging design (0.605); recommendations from friends/ family (0.575); recommendations from influencers (0.657)	5.16%

Source: Authors' own calculations based on questionnaire survey data (n = 458).

3.6 Generational Segmentation

Generation Y emerged as the most active market segment. Among respondents in this age group, 37.8% consume craft beer several times per week, and more than half spend over CZK 500 per month on craft beer purchases. This generation places particular emphasis on sensory characteristics, authenticity, and small independent breweries. Generation X exhibits stable consumption habits and a strong preference for regional origin and support for local producers. Generation Z represents the least active segment; its members consume craft beer irregularly, demonstrate greater price sensitivity, and more frequently prefer well-known and easily accessible brands. This tendency may be related to their lower average income, as more than half of Generation Z respondents reported a monthly income below CZK 20,000.

4 DISCUSSION AND CONCLUSIONS

The identification of key factors influencing craft beer selection confirmed the dominant role of sensory attributes. The finding that taste and aroma constitute the most influential factor is consistent with the results of Lerro *et al.* (2020), who, using the Best–Worst Scaling method, demonstrated that flavor-related characteristics form the core of consumer evaluations. Likewise, the importance of traditional recipes, ingredient composition, and the production process aligns with the characterization of the craft beer segment as a quality-oriented form of production (Villacreces *et al.*, 2022; Baiano, 2021).

The statistically significant relationship between beer-type preference and income level complements and confirms the findings of Steinbach *et al.* (2023), according to which price is one of the principal reasons why consumers who primarily prefer industrial beer consume craft beer only occasionally. Furthermore, a statistical test (Wilcoxon signed-rank test) demonstrated that respondents who prefer craft beer attach greater importance to regional origin than to price. This finding is consistent with the conclusions of Svatošová *et al.* (2021), who reported that price is not a decisive criterion for craft beer enthusiasts. Thus, value-based attributes—namely authenticity and local provenance—outweigh price sensitivity among consumers who are oriented toward quality and product storytelling.

The finding of significantly lower craft beer consumption among Generation Z corresponds with international studies documenting declining alcohol consumption among the youngest cohorts (Livingston *et al.*, 2024; Caluzzi *et al.*, 2024). This trend represents a structural challenge for the entire segment, as Generation Z will constitute the core of the beer market in the coming decades.

In contrast, the hypothesis derived from the study by Nieto-Villegas *et al.* (2024), which suggested that consumers with positive attitudes toward sustainability would prefer both craft beer and environmentally friendly packaging, was not supported in the Czech context. Respondents who preferred craft beer did not assign significantly greater importance to environmentally friendly packaging than consumers who preferred industrial beer. This finding suggests that conclusions drawn from sustainability-focused studies are not automatically transferable across markets and require contextual verification.

Among the most important psychological mechanisms identified in the study were the storytelling effect and the principle of authenticity. The storytelling effect was supported by a statistically significant relationship between preference for narrative-based promotion and the importance assigned to brewery history as a selection criterion, which is consistent with theoretical assumptions regarding the influence of storytelling on brand relationships (Ahmed *et al.*, 2024). The principle of authenticity, as conceptualized by Newman (2019) and Bartsch *et al.* (2022), emerged in the analysis as an independent extracted factor (F3 – Authenticity and Exclusivity), encompassing regional origin, support for local producers, and exclusive product offerings. The scarcity principle (Goldsmith *et al.*, 2020) was supported by the finding that regular consumers attributed greater importance to limited-edition products than occasional consumers.

With regard to promotional activities, the findings clearly favored experiential forms of communication. The most persuasive promotional elements identified by respondents were collaborations between craft breweries and local businesses (mean = 4.94), brewery visits accompanied by beer

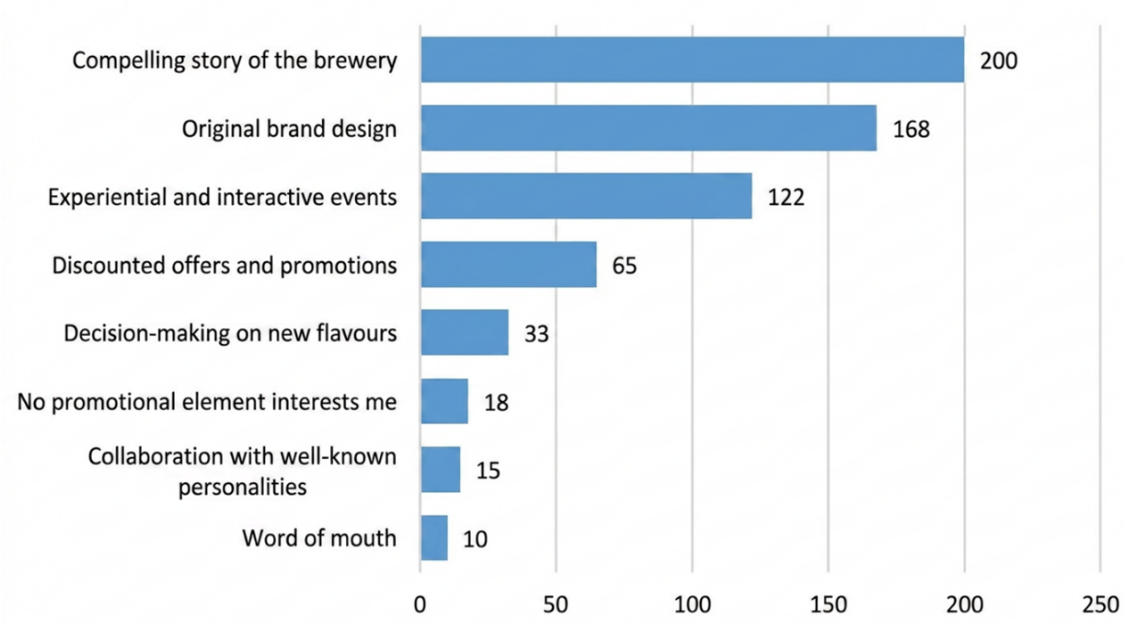


Fig. 2: Most Preferred Promotional Elements of Craft Beer (n = 458)
Source: Questionnaire survey data (n = 458).

tastings (4.87), and participation in beer festivals (4.78). In contrast, recommendations from influencers, newspaper advertising, and billboards were evaluated as relatively unpersuasive. These findings suggest that effective craft beer marketing relies primarily on authentic experiences and direct interaction with the brand rather than on traditional mass advertising.

The study is subject to several limitations. The purposive sampling strategy does not permit generalization of the findings to the entire population of beer consumers in the Czech Republic. The online mode of data collection may have underrepresented older generations with lower levels of technological familiarity, which ultimately led to the exclusion of respondents aged 58 years and older. The substantial representation of respondents from the South Moravian Region and the Vysočina Region may have influenced certain results, such as brand preferences. Furthermore, the research employed an exclusively quantitative design, which may have overlooked deeper consumer motivations and emotional dimensions of decision-making. Future research should therefore incorporate qualitative interviews and behavioral experiments to provide more robust testing of psychological effects, particularly the availability heuristic (Shah *et al.*, 2012), which received only partial support in the present study.

The principal managerial implication for businesses operating in the craft beer market is the need to focus promotional activities on experiential marketing, authentic brand storytelling, and the sensory quality of the product. Generation Y represents the key target segment, while Generation X constitutes a stable segment characterized by a preference for tradition and local provenance. In contrast, attracting Generation Z requires a combination of affordable offerings and inclusive experiential formats. Craft beer consumers do not perceive the product merely as a beverage but rather as an experience characterized by quality, narrative, and authenticity.

In discussion, please provide a confrontation of the achieved results with previously published papers, author's opinion of established differences, his/her attitude to the results. The discussion section also provides a space to outline the need of further potential solution or importance for the development of science, society or practice.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

AHMED, S. *et al*. 2024. Crafting emotional engagement and immersive experiences: Comprehensive scale development for and validation of hospitality marketing storytelling involvement. *Psychology & Marketing*. 41(7), 1514–1529.

BAIANO, A. 2021. Craft beer: An overview. *Comprehensive Reviews in Food Science and Food Safety*. 20(2), 1829–1856.

BARTSCH, F., ZEUGNER-ROTH, K. P., KATSIKEAS, C. S. 2022. Consumer authenticity seeking: conceptualization, measurement, and contingent effects. *Journal of the Academy of Marketing Science*. 1–28.

BOLBOACĂ, S.D. *et al*. 2011. Pearson-Fisher chi-square statistic revisited. *Information*. 2(3), 528–545.

BŘEZINOVÁ, M. 2021. Beer industry in the Czech Republic: Reasons for founding a craft brewery. *Sustainability*. 13(17), 9680.

CALUZZI, G. *et al*. 2024. ‚Ninety per cent of the time when I have had a drink in my hand I'm on my phone as well': a cross-national analysis of communications technologies and drinking practices among young people. *New Media & Society*. 26(10), 6120–6140.

ČESKOMORAVSKÝ SVAZ MINIPIVOVARŮ. 2025. O nás. *Českomoravský svaz minipivovarů*. [Accessed: 27. 4. 2025] https://www.minipivo.cz/cz/o-nas/

FIELD, A. 2013. *Discovering Statistics Using IBM SPSS Statistics*. 4th Edition. London: SAGE Publications.

GOLDSMITH, K., GRISKEVICIUS, V., HAMILTON, R. 2020. Scarcity and consumer decision making: Is scarcity a mindset, a threat, a reference point, or a journey? *Journal of the Association for Consumer Research*. 5(4), 358–364.

LERRO, M., MAROTTA, G., NAZZARO, C. 2020. Measuring consumers' preferences for craft beer attributes through Best-Worst Scaling. *Agricultural and Food Economics*. 8, 1–13.

LIVINGSTON, M. *et al*. 2024. Youth drinking in decline: Recent developments and future priorities. *Drug & Alcohol Review*. 43(3).

NEWMAN, G. E. 2019. The psychology of authenticity. *Review of General Psychology*. 23(1), 8–18.

NIETO-VILLEGAS, R., BERNABÉU, R., RABADÁN, A. 2024. Effects on beer attribute preferences of consumers' attitudes towards sustainability: The case of craft beer and beer packaging. *Journal of Agriculture and Food Research*. 15, 101050.

PILONE, V., DI PASQUALE, A., STASI, A. 2023. Consumer preferences for craft beer by means of artificial intelligence: are Italian producers doing well? *Beverages*. 9(1), 26.

ROSALES, A. *et al*. 2021. Craft beer vs industrial beer: chemical and sensory differences. *British Food Journal*. 123(12), 4332–4346.

SHAH, A.K., MULLAINATHAN, S. a SHAFIR, E. 2012. Some consequences of having too little. *Science*. 338(6107), 682–685.

STEINBACH, J., BURGARDT, V. C. F., MACHADO-LUNKES, A. 2023. Perceptions, attitudes, and motivational factors for consumers and nonconsumers of traditional and craft beers. *Journal of Sensory Studies*. 38(2), e12813.

STEM/MARK. 2024. Generace Y je velkorysá vůči starším, generace Z přináší naději pro planetu . *Stenmark.cz*. Praha. [Accessed: 27. 4. 2025] https://stemmark.cz/generace-y-je-velkorysa-vuci-starsim-generace-z-prinasi-nadeji-pro-planetu

SVATOŠOVÁ, V., KOSOVÁ, P. a SVOBODOVÁ, Z. 2021. Factors influencing consumer behaviour in the beer market in the Czech Republic. *Czech Journal of Food Sciences*. 39(4).

VILLACRECES, S., BLANCO, C. A., CABALLERO, I. 2022. Developments and characteristics of craft beer production processes. *Food Bioscience*. 45, 101495.

WOJTYRA, B. *et al*. 2020. Geography of craft breweries in Central Europe: Location factors and the spatial dependence effect. *Applied Geography*. 124, 102325.

Contact information

Lenka Procházková: lenka.prochazkova@mendelu.cz
Marek Záboj: e-mail: marek.zaboj@mendelu.cz
Matěj Sysel: e-mail: matejsysel@seznam.cz

CLOSING THE GREEN GAP: HARMONIZING PROJECT AND PROCESS APPROACHES IN HIGHER EDUCATION VIA SUSTAINABLE LEADERSHIP AND DIGITAL INTERNAL MARKETING

Petra Skořepová[1], Petra Marešová[1]

[1]Department of management, University of Hradec Králové, Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_032

ABSTRACT

Higher Education Institutions (HEIs) are currently navigating the complexities of the global "Twin Transition" (green and digital). However, a significant "implementation gap" persists between strategic sustainability declarations and operational reality. This paper scrutinizes this phenomenon within the University of Hradec Králové (UHK), analyzing the misalignment between the university's "Sustainable Development Strategy 2030" and actual workforce perception. Drawing on the duality of project and process approaches in institutional adaptability, this study utilizes qualitative data from 15 in-depth semi-structured interviews. The findings delineate a critical asymmetry: while respondents widely recognize tangible infrastructure improvements ("hardware") such as waste sorting stations and energy-saving technologies, they report a profound lack of awareness regarding the strategic vision. The analysis identifies key barriers preventing behavioral adoption, specifically "bureaucratic inertia" where digital goals clash with persistent analog administrative processes, and a "skill gap" where staff lack the know-how to operate new technologies (e.g., smart thermostats), leading to frustration and disengagement. The study argues that without addressing the behavioral dimension through specific internal marketing strategies, the potential of green infrastructure remains untapped, resulting in institutional decoupling. The paper concludes that leadership must pivot towards an Industry 5.0 mindset, leveraging digital tools not just for management, but as internal marketing instruments to "nudge" the sustainable workforce towards active participation.

Keywords: change management, institutional adaptability, Industry 5.0, internal marketing, organizational efficiency, green nudging, sustainable workforce
JEL Code: I23, M14, M31, Q01

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

In the contemporary landscape of tertiary education, Higher Education Institutions (HEIs) are increasingly recognized not merely as centers of knowledge creation but as critical drivers of the global sustainability transition. Following the framework of the UN Agenda 2030, specifically Sustainable Development Goals 4 (Quality Education) and 12 (Responsible Consumption and Production), universities are expected to transcend their traditional pedagogical roles to become living laboratories of sustainable practice. As highlighted by Purcell *et al.* (2019), universities act as the "engine of transformational sustainability", bearing the responsibility to model best practices for the wider society.

However, the adoption of ambitious sustainability declarations does not automatically translate into institutional reality. A growing body of literature suggests that HEIs are particularly prone to a "rhetoric-reality gap", where high-level policy commitments fail to permeate the daily operational fabric of the institution (Aleixo *et al.*, 2018). Recent reports highlight that a significant challenge facing universities in the post-pandemic era is the "implementation gap"—the profound disconnect between policy intentions and practical execution. This gap represents a failure of institutional adaptability—the capacity of the university to adjust its internal operations, culture, and governance structures to meet external strategic goals.

While university leadership may formulate ambitious policies aligned with the European Green Deal, these strategies often face substantial barriers during the deployment phase, leading to stalled initiatives and wasted resources. this paper approaches sustainability through the dual lens of change management and internal marketing, specifically examining the dichotomy between the *project approach* (technical execution) and the *process approach* (people-side adoption).

The research focuses on the University of Hradec Králové (UHK) as a representative case study of a Central European public university facing these transition challenges. The central research question driving this study is:

> ***Does the current internal marketing strategy effectively support the process side of change, ensuring the adaptability of the sustainable workforce, or does it contribute to institutional decoupling and sub-optimal asset utilization?***

2 THEORETICAL BACKGROUND

To understand the mechanics of the implementation gap, this study synthesizes theories from change management, sustainable leadership, and the emerging paradigm of Industry 5.0. The proposed theoretical framework posits that institutional adaptability is contingent upon the harmonization of project and process management styles. Higher Education Institutions (HEIs) are characterized by complex governance structures, often described as "loosely coupled systems", making them historically resistant to rapid transformation (Cai & Mehari, 2015). Within this context, a core theoretical distinction exists between project and process approaches to change. The Project Approach is primarily focused on technical deliverables, rigid timelines, budgets, and the installation of physical "hardware", such as thermal insulation, solar panels, or new waste containers. As noted by Hornstein (2015), project management provides the structure and discipline required for

execution. In the context of university sustainability, this relies heavily on "Hard" change management tools and is often perceived as the straightforward component of transformation.

In contrast, the Process Approach focuses on the "people side of change", addressing the intangible but critical elements of organizational culture, resistance management, and behavioral adoption. As Al-Haddad and Kotnour (2015) argue, effective change requires the seamless integration of these two approaches; without the process dimension, sustainability initiatives remain superficial "bolt-on" projects rather than "built-in" adaptive capabilities. Furthermore, an exclusive focus on project management risks creating a ‚compliance culture,' whereas a process approach aims for a ‚commitment culture' where sustainable behavior is internalized. Stouten *et al.* (2018) emphasize that while project management ensures the installation of change, process management ensures the realization of benefits. The disconnect between these two approaches is often where the implementation gap widens.

Transitioning towards a sustainable university requires leadership that acts as a "change agent" to bridge this dichotomy. Sustainable leadership focuses on developing a "sustainable workforce", where employee competencies, well-being, and engagement are directly correlated with organizational performance. According to Di Fabio and Peiró (2018), "Human Capital Sustainability Leadership" is essential for promoting healthy and resilient organizations, fostering an environment of trust rather than relying solely on transactional models. In the context of HEIs, where the workforce possesses a high degree of autonomy, leaders must facilitate rather than command. However, significant barriers remain; Blanco-Portela *et al.* (2017) identify that the lack of clear communication channels and resistance to change are primary barriers to institutionalizing sustainability. This resistance often stems from a lack of understanding of why the change is necessary, pointing to a failure in internal marketing.

To bridge the gap between vision (project) and engagement (process) and overcome these barriers, effective communication is essential. Internal marketing views employees and students as "internal customers" whose buy in is a prerequisite for external success. If internal marketing fails, the institution suffers from institutional decoupling, a state where the formal structure (strategy) is detached from daily work activities. Papadas *et al.* (2019) introduce the concept of "Internal Green Marketing Orientation", suggesting that for an organization to achieve competitive advantage through sustainability, it must first successfully market these values internally. However, traditional internal marketing is frequently insufficient in the digital age. This is where the concept of Industry 5.0 becomes relevant. Unlike Industry 4.0, which focused heavily on automation, Industry 5.0 places the human being back at the center of the production process, emphasizing well-being and resilience (Breque *et al.*, 2021). In an HEI context, this means leveraging technology not to replace human decision-making, but to "nudge" it. While current literature defines Digital Twins primarily as models of physical assets, this paper proposes a theoretical extension: the use of digital feedback loops as internal marketing tools. By moving from passive dissemination to active, data-driven behavioral influence (digital nudging), universities can effectively close the process gap.

3 METHODOLOGY

To deeply investigate the "implementation gap" and the misalignment between strategic intent (Project Approach) and operational reality (Process Approach), a qualitative research strategy was deemed most appropriate. Unlike quantitative methods, which might measure the volume of energy saved, this qualitative inquiry seeks to understand the underlying reasons behind workforce behavioral adoption or resistance. Consequently, this research employs a qualitative single-case study design focusing on the University of Hradec Králové (UHK). The choice of a single-case study is justified by the need for a "thick description" of the organizational context (Rashid *et al.*, 2019), allowing for an in-depth exploration of the complex social dynamics that quantitative methods might miss. This qualitative inquiry serves as a necessary exploratory phase to identify latent behavioral variables, such as cognitive dissonance in bureaucracy. UHK, as a dynamic mid-sized public university in the Czech Republic, serves as a representative example of an institution undergoing rapid modernization while grappling with historical administrative legacies. The university has recently adopted a rigorous "Sustainable Development Strategy 2030", making it an ideal environment to test the alignment between strategic intent and operational reality.

The data collection was conducted between September and November 2025 using a design that consisted of two distinct phases to ensure methodological triangulation. First, a comprehensive document analysis was performed, reviewing the *Sustainable Development Strategy of UHK until 2030*. This analysis aimed to identify key strategic goals across the university's strategic pillars and establish the "Project" baseline - essentially mapping what the university claims it is achieving regarding green infrastructure and digitalization. The second phase involved collecting primary data through semi-structured interviews with a purposive sample of 15 internal stakeholders (N=15). The sample was constructed to ensure maximum heterogeneity and avoid echo chambers, consisting of three distinct cohorts: Academic Staff (R1–R5), whose perspective was crucial for understanding the "frontline" implementation of sustainability in education and research; Administrative Staff (R6–R10) from the Rectorate and Faculty offices, who provided insights into the operational and bureaucratic bottlenecks; and Student Representatives (R11–R15), who represented the voice of the primary "customer" of the university's services.

To ensure the data collected was sufficiently granular, the interview protocol was designed based on the ADKAR model (Awareness, Desire, Knowledge, Ability, Reinforcement) as suggested by Hiatt (2006), utilizing six thematic blocks to assess the maturity of the process approach. The specific inquiries included questions on *Awareness & Leadership* (e.g., "How well were you acquainted with the Sustainable Development Strategy?" and "Do you perceive that the university management actively supports and communicates this strategy?"); *Alignment & Communication* to identify gaps between rhetoric and reality (e.g., "Do the marketing campaigns correspond to the strategic goals?" and "Does the campaign appear authentic to you, or do you perceive it as greenwashing?"); *Engagement & Participation* to assess the "people side" of change (e.g., "Do you feel, as an employee/student, part of the change towards sustainability?" and "Was there space to participate in the strategy or activities?") ; and finally, I*mpact & Reinforcement* to evaluate tangible results (e.g., "Have you noticed specific changes in operations, communication, or infrastructure?" and "What should the university change in communicating the topic of sustainability?").

The interview transcripts were subsequently analyzed using Reflexive Thematic Analysis following the updated guidelines by Braun and Clarke (2022). This involved a recursive six-phase process: familiarization with the data, generating initial codes, searching for themes (such as "bureaucratic frustration" or "visible hardware"), reviewing themes, defining and naming themes,

and producing the report. Data were manually coded using a semantic approach to identify latent themes regarding organizational behavior. To ensure the trustworthiness of the findings, the study adopted the criteria proposed by Nowell *et al.* (2017) . Credibility was established through prolonged engagement with the field and investigator triangulation, ensuring that the findings reflected the participants' reality, while transferability is supported by the detailed description of the organizational context. Ethical considerations were strictly adhered to; participation was voluntary, and all data were anonymized. Finally, regarding the use of technology, generative AI tools were utilized during the preparation of this manuscript exclusively for linguistic refinement and stylistic editing to enhance the clarity, flow, and readability of the text. The author verified the AI-generated output to ensure it accurately reflected the original research findings, and the study design, data collection, analysis, and final scientific conclusions remain the original work of the author, who assumes full responsibility for the content.

4 RESULTS

The thematic analysis revealed a stark dichotomy between the physical transformation of the university and the behavioral adaptation of its workforce. The findings are summarized in Table 1 and categorize the current state into two main themes: the success of the "Hardware" (Project Approach) and the failure of the "Software" (Process Approach).

In terms of the Project Approach, respondents universally recognized the tangible, physical changes occurring on campus. Investments in infrastructure—such as thermal insulation, photovoltaic panels, and visible waste sorting stations—were widely cited as evidence of the university's commitment. One participant remarked, "*When I see the new solar panels on the roof, I feel the university is taking it seriously*" (R3). This sentiment was echoed across all respondent groups, indicating that the university excels in the visual legitimacy of sustainability (Miotto *et al.*,

Tab. 1 The Asymmetry between Project and Process Approaches at UHK

Dimension	Project Approach (Hardware & Strategy)	Process Approach (Software & Behavior)	Representative Evidence / Quote
Visibility & Perception	**High Visibility:** Tangible assets are prominent (solar panels, insulation, waste stations).	**Low Awareness:** Strategic goals are abstract; internal marketing is viewed as "spam".	*"When I see the new solar panels, I feel the university is serious... but I don't know the plan."*
Administrative Workflow	**Digital Goals:** "Digital Strategy 2030" and ERP systems are technically implemented.	**Analog Reality:** Persistent bureaucracy requiring paper forms and physical stamps.	*"It is schizophrenic... We have smart lights, but I need a physical stamp for a digital ticket."*
User Interaction	**Automated:** Smart technologies (blinds, thermostats) operate on algorithms.	**Resistant:** Users lack training and override systems due to frustration.	*"The automatic blinds go down... so I turn on artificial lights. It's absurd."*
Change Management	**Delivery-Focused:** Projects are completed on time and within budget (CAPEX).	**Adoption-Lag:** Behavioral change is neglected, leading to low utilization (ROI).	*"Tech is ready, people are not." (Observed gap in skills)*

Source: Author's own processing

2020). The physical assets serve as powerful symbols. This observation aligns with recent research by Vaverková *et al.* (2023), which suggests that visible waste management infrastructure acts not only as a functional tool but as a primary pedagogical instrument for environmental awareness on campuses. Consequently, UHK demonstrates high competence in delivering technical solutions on time and within budget; the "Hardware" is effectively installed.

Conversely, the "human side" of the transition revealed significant deficiencies, hindering true institutional adaptability. A primary driver of this failure is bureaucratic inertia, which creates a state of *cognitive dissonance* for employees who are asked to be sustainable but are forced by internal processes to be wasteful. Despite strategic goals emphasizing digitalization, administrative processes remain rigid and paper-heavy, a phenomenon partly attributed to the path dependency often observed in post-socialist Central European institutions. The most frequently cited example was the travel order process. "I *fill out the digital request in the ERP system, but then I must print it, sign it, and carry it to three different offices for physical stamps. It makes the strategy feel like a marketing façade*", noted one administrative staff member (R6). Another respondent (R8) highlighted the absurdity of the situation: "*We have smart lights in the hallway, but I need a physical stamp to approve a digital conference ticket. It is schizophrenic.*" This paradox undermines the credibility of the internal marketing narrative, causing employees to disengage and view sustainability as a management buzzword rather than a core organizational value.

Furthermore, beyond administrative hurdles, the study identified a critical lack of behavioral adoption regarding green technologies. Interviews revealed that employees often lack the specific know-how to interact with new energy-saving systems, leading to a loss of control and frustration. For instance, staff members reported fighting with automated systems: "*The automatic blinds go down when the sun shines to save cooling costs, but I need light to work. So I turn on the artificial lights. It's absurd*", noted a researcher (R4). Similarly, a skill gap was observed in waste management; while sorting stations are present, unclear rules for specific materials (e.g., coffee cups with plastic lining) lead to the contamination of recycling bins. This evidence confirms that while the "project" (delivery of the new thermostat or blinds) was completed, the "process" (training, adoption, and habit formation) was neglected. Without the "soft skills" to operate the "hard tech", the technology remains underutilized, indicating a clear implementation gap where top-down goals are met, but bottom-up adoption lags significantly.

5 DISCUSSION AND CONCLUSIONS

The findings of this study provide critical insights into the dynamics of institutional adaptability and the role of internal marketing in the sustainability transition. The results confirm that UHK faces an "implementation gap" caused by a significant imbalance between project and process management. While the university demonstrates high *technical adaptability*, evidenced by the successful installation of physical infrastructure, it exhibits low *social adaptability*. This misalignment supports the theory of institutional decoupling, where strategic ambitions remain detached from daily operational reality. More importantly, this gap poses a severe threat to the strategic effectiveness of the institution. Unlike passive infrastructure, such as thermal wall insulation which delivers benefits independently of user behavior, active technologies—like smart thermostats, sensor-based lighting, and advanced waste sorting—depend entirely on workforce interaction to function correctly.

As noted by organizational theorists, "culture eats strategy for breakfast". In the context of university sustainability, we can add that culture also determines the utility of technical innovation. If internal marketing fails to secure buy in and behavioral change, the actual utility of green investments remains unrealized (Bromley & Powell, 2012). A smart building filled with users who behave "dumbly" (due to lack of training or motivation) is not a smart building; it represents a failure of resource utilization. This emphasizes that physical assets cannot become a source of advantage unless they are complemented by the necessary organizational capabilities and "human software".

To address this operational disconnect, the identified failure of traditional internal marketing points to a need for a shift in leadership style and tools. Leadership must move from passive dissemination of information to active behavioral influence. This is where the transition from Industry 4.0 to Industry 5.0 offers a viable solution. While Industry 4.0 focused on digitizing processes, Industry 5.0 focuses on the collaboration between humans and machines to enhance human well-being and sustainability (Psarommatis *et al.*, 2020). In this context, the study proposes the implementation of a "Behavioral Green Digital Twin". Just as digital twins in industry optimize machinery by using real-time data, a digital twin of employee behavior allows for personalized, non-intrusive feedback. Unlike industrial systems that monitor machine health, a ‚Green Employee Digital Twin' would aggregate anonymized behavioral data to provide gamified feedback. Instead of a reprimand, the system might suggest: "*You utilized 10% less paper than last month – you saved 2 kg of* CO_2."

The psychological impact of such a system is profound. It shifts the locus of control from external regulations (compliance) to internal motivation (commitment), fostering a sense of ownership over the university's carbon footprint. This concept relies on Nudging theory, altering the "choice architecture" to guide behavior without forbidding options (UNEP, 2020). As demonstrated by Hummel and Maedche (2019), digital nudging can significantly enhance user compliance and adoption rates. Consequently, to close the "Green Gap", university leadership must adopt a multi-faceted strategy. Internal marketing should be viewed strictly as a crucial management tool to prevent asset underutilization. Furthermore, resources must be harmonized; for every unit of effort spent on "Hardware", a proportional investment must be made in the "Software" of training and adoption, as the "Sustainable Workforce" must be cultivated, not just bought. Finally, the current state of "bureaucratic schizophrenia", where digital goals clash with analog processes, must be resolved by leveraging these Industry 5.0 principles to "nudge" employees towards sustainable habits without intrusive surveillance (Adam *et al.*, 2021).

6 DISCUSSION AND CONCLUSIONS

This study evaluated the alignment between UHK's sustainability strategy and workforce perception, contributing to the broader discourse on institutional adaptability in Higher Education. The research concludes that while the project approach has successfully delivered green infrastructure, the process approach lags behind due to ineffective internal marketing and insufficient focus on behavioral change. By harmonizing these approaches and leveraging digital internal marketing, universities can transform from mere purchasers of green technology into truly adaptive, sustainable organizations, thereby securing the ROI of their green investments.

REFERENCES

ADAM, M., WESSEL, M., BENLIAN, A. 2021. AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*. 31, 427–445. https://doi.org/10.1007/s12525-020-00414-7

ALEIXO, A. M., LEAL, S., AZEITEIRO, U. M. 2018. Conceptualization of a sustainable university by institutional leaders: A systematic review. *International Journal of Sustainability in Higher Education*. 19(6), 1192 1216. https://doi.org/10.1016/j.jclepro.2016.11.010

AL-HADDAD, S., KOTNOUR, T. 2015. Integrating the organizational change literature: A model for successful change. *Journal of Organizational Change Management*. 28(2), 234–262. https://doi.org/10.1108/JOCM-11 2013-0215

BLANCO-PORTELA, N., BENAYAS, J., PERTIERRA, L. R., LOZANO, R. 2017. Towards the integration of sustainability in Higher Education Institutions: A review of drivers of and barriers to organisational change and their comparison against found of companies. *Journal of Cleaner Production*. 166, 578. https://doi.org/10.1016/j.jclepro.2017.07.252

BRAUN, V., CLARKE, V. 2022. *Thematic analysis: A practical guide*. SAGE Publications.

BREQUE, M., DE NUL, L., PETRIDIS, A. 2021. *Industry 5.0: Towards a sustainable, human-centric and resilient European industry*. European Commission, Directorate-General for Research and Innovation. https://doi.org/10.2777/308407

BROMLEY, P., POWELL, W. W. 2012. From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *Academy of Management Annals*. 6(1), 483–530. https://doi.org/10.5465/19416520.2012.684462

Cai, Y., Mehari, Y. 2015. The use of institutional theory in higher education research. In: HUISMAN, J., TIGHT, M. (Eds.). *Theory and Method in Higher Education Research*. Emerald Group Publishing, pp. 1–25. https://doi.org/10.1108/S2056-375220150000001001

DI FABIO, A., PEIRÓ, J. M. 2018. Human Capital Sustainability Leadership to promote sustainable development and healthy organizations: A new scale. *Sustainability*. 10(7), 2413. https://doi.org/10.3390/su10072413

HIATT, J. M. 2006. *ADKAR: A model for change in business, government and our community*. Prosci Learning Center Publications.

HORNSTEIN, H. A. 2015. The integration of project management and organizational change management is now a necessity. *International Journal of Project Management*. 33(2), 291–298. https://doi.org/10.1016/j.ijproman.2014.08.005

HUMMEL, D., MAEDCHE, A. 2019. How effective is nudging? A quantitative review on the effect sizes of cognitive biases within e-commerce. *Journal of Decision Systems*. 28(2), 127–159. https://doi.org/10.1016/j.socec.2019.03.005

MIOTTO, G., DEL-CASTILLO-FEITO, C., BLANCO-GONZÁLEZ, A. 2020. Reputation and legitimacy: Key factors for Higher Education Institutions' sustained competitive advantage. *Journal of Business Research*. 112, 342–353. https://doi.org/10.1016/j.jbusres.2019.11.076

NOWELL, L. S., NORRIS, J. M., WHITE, D. E. 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*. 16(1). https://doi.org/10.1177/1609406917733847

PAPADAS, K. K., AVLONITIS, G. J., CARRIGAN, M. 2019. The interplay of strategic and internal green marketing orientation competitive advantage. *Journal of Business Research*. 104, 632–643. https://doi.org/10.1016/j.jbusres.2018.07.009

PSAROMMATIS, F., MAY, G., DREYFUS, P. A., KIRITSIS, D. 2020. Zero defect manufacturing: State-of-the-art review, shortcomings and future directions in research. *International Journal of Production Research*. 58(7), 1917–1953. https://doi.org/10.1080/00207543.2019.1605228

PURCELL, W. M., HENRIKSEN, H., SPENGLER, J. D. 2019. Universities as the engine of transformational sustainability toward delivering the sustainable development goals. *International Journal of Sustainability in Higher Education*. 20(8), 1343–1357. https://doi.org/10.1108/IJSHE-02-2019-0103

RASHID, Y., RASHID, A., WARRAICH, M. A., SABIR, S. S., WASEEM, A. 2019. Case study method: A step-by-step guide for business researchers. *International Journal of Qualitative Methods*. 18, 1–13. https://doi.org/10.1177/1609406919862424

STOUTEN, J., ROUSSEAU, D. M., DE CREMER, D. 2018. Successful organizational change: Integrating the management practice and scholarly literatures. *Academy of Management Annals*. 12(2), 752–788. https://doi.org/10.5465/annals.2016.0095

UNEP. 2020. *The Little Book of Green Nudges*. United Nations Environment Programme & The Behavioural Insights Team. https://www.unep.org/resources/publication/little-book-green-nudges

VAVERKOVÁ, M. D., ELBL, J., KODA, E., ADAMCOVÁ, D. 2023. Closing the loop: A case study on pathways for promoting sustainable waste management on university campuses. *Science of the Total Environment*. 892, 164349. https://doi.org/10.1016/j.scitotenv.2023.164349

Contact information

Petra Skořepová: e-mail: petra.skorepova.2@uhk.cz

CLASSIFYING SCANNING ELECTRON MICROSCOPE DEFECTS WITH A CNN AND A CUSTOM FOURIER-TRANSFORM LAYER

Magdaléna Skřičilová[1], Oldřich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_033

ABSTRACT

A scanning electron microscope has to be checked periodically, and two of its most common image faults, defocus and astigmatism, are tedious to judge by eye. This article presents a convolutional neural network that sorts pairs of microscope images into four focus and astigmatism classes so the check can run without an operator. Each sample is two shots of the same spot: one in the base position and one with a shifted working distance, because the shifted shot is what makes astigmatism visible. The decisive ingredient is a custom Fourier-transform preprocessing layer that converts each input to its magnitude spectrum before the convolutional part sees it. On a dataset of about three thousand image pairs, the base model variants with-out the Fourier layer reached 49–65% test accuracy, while the corresponding variants with the Fourier layer reached 89–94%; the best result was 94% with channel-axis concatenation. The same shallow model beat three pretrained ImageNet networks (MobileNetV2 87%, EfficientNetV2L 52%, ResNet50 44%), which we did not expect. The best fusion of the two images was concatenation along the channel axis together with the Fourier layer.

Keywords: convolutional neural network, scanning electron microscope, Fourier transform, image classification, astigmatism
JEL Code: C45

1 INTRODUCTION

In the development and testing of scanning electron microscopes, repeated regression tests are needed to check whether microscope functions continue to work correctly after software or parameter changes. Two image faults addressed in this work are defocus and astigmatism; both can occur for several reasons, including sample or detector contamination, incorrect electron-source settings, magnetic-lens hysteresis or unsuitable microscope parameters. Their visual

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

similarity makes manual assessment difficult, which motivates testing whether a neural network can classify the image states automatically and support an automated regression-testing framework for the microscope.

A single shot is ambiguous: a focused image with astigmatism and a defocused image without it can be visually identical. Software testing treats this kind of input-dependent behaviour as something to exercise systematically rather than check once (Ammann and Offutt, 2008), which is the spirit of running the classifier as a repeatable test. The way out is to take two pictures of the same position and let the network compare them. The aim is to build such a classifier, push its accuracy high enough to be useful, and find which design choices matter.

1.1 Focus and astigmatism correction in the SEM

Automatic focusing and astigmatism correction for electron microscopes is an old problem. Erasmus and Smith (1982) built an early hardware system for it, and Ogasawara *et al.* (1999) showed that the Fourier transform of an SEM image carries a clean signature of both faults, because defocus and astigmatism reshape the frequency spectrum in characteristic ways. Lu, Zhang and Li (2018) later turned this into a simplified single-image correction method based on the spectrum. The shared idea, that the frequency domain rather than the raw pixels is where focus information lives, is the starting point here.

Recent work has moved the same task onto deep learning. Na *et al.* (2021) used a network to discriminate refocused SEM images for materials science, Lee *et al.* (2021) built a dual deep-learning network for robust autofocusing, and Schubert, Saxena and Kornfeld (2024) reported fast focus and astigmatism correction. Most of these read a single image. The classifier described below instead reads a pair and asks a four-way question, focused or not and stigmatic or not, which is why two shots are needed.

1.2 Convolutional networks and blur detection

Convolutional neural networks (CNNs) are the standard tool for image classification (LeCun, Bengio and Hinton, 2015; Aggarwal, 2018). The deep ImageNet architectures, AlexNet (Krizhevsky, Sutskever and Hinton, 2012) and ResNet (He *et al.*, 2016) among them, win on natural-image benchmarks with tens to hundreds of layers. Detecting blur is narrower, and there the picture is less one-sided: Szandała (2020) showed that a CNN can replace the classical Laplacian sharpness measure. Whether a deep pretrained network or a small purpose-built one fits the SEM defect task is one of the questions answered here.

2 METHODOLOGY

The task was framed as a four-class classification of image pairs. Each sample is two pictures of the same position: a base (focused) shot and a second shot taken with the working distance shifted, which throws the image out of focus on purpose. The four classes are focused-stigmatic, focused-astigmatic, defocused-stigmatic and defocused-astigmatic. The second shot disambiguates the hard pair from the introduction, since astigmatism shows up as a sideways elongation once the image is defocused. A focused-astigmatic and a defocused-stigmatic sample, identical in the first shot, separate in the second.

Data were collected by a script hooked into the microscope-control software, which cycled through commands to take the four image types at various positions and magnifications. Defocus was induced by shifting the working distance by 0.05 to 0.9 mm depending on magnification. Images were then checked and sorted by hand, which was slow and not error-free, since the labelling task is hard for a person too. The final dataset held about 3 000 image pairs, balanced across classes, shuffled and split into training, validation and test parts, with 80% used for training. The standard tin-balls-on-carbon sample was used throughout, because its regular round shape makes defects easy to see.

Everything ran in Python with TensorFlow and its Keras API on a GPU. Each image was decoded, resized to 128×128 pixels, cast to float and normalized to the range [0, 1]. On the training set only, both images of a pair were augmented with the same random seed (horizontal and vertical flips, an occasional 90° rotation, a brightness change, a random crop) so the pairing stayed consistent.

The base network is deliberately small: three convolutional layers (3×3 filters, 32 then 64 then 96 of them) interleaved with 2×2 max-pooling, then four fully-connected layers ending in the four-class output. It uses ReLU activations, He-normal initialization, L2 weight regularization and a single dropout layer. The whole model is about 1.3 million parameters and 5.04 MB on disk. Training used the Adam optimizer (Kingma and Ba, 2014) with an exponentially decaying learning rate, sparse categorical cross-entropy loss, and roughly 50 epochs on average, with learning-rate reduction and early stopping on the validation loss.

Two design axes were varied. The first is how the two images are fused into one input: side by side (horizontal), stacked as extra channels (channel-axis), or processed in separate branches whose outputs are concatenated. The second is the Fourier-transform layer described below, tested on and off for every fusion type. For comparison, three pretrained ImageNet models from Keras Applications (MobileNetV2, EfficientNetV2L and ResNet50) were also tried as feature extractors. Fine-tuning them gave very poor results, probably because the dataset is small, so only feature extraction is reported.

2.1 The Fourier-transform layer

The key addition is a custom Keras preprocessing layer, with no trainable weights, placed right after the input. For each image it converts to grayscale, runs a two-dimensional FFT, shifts the zero frequency to the centre, takes the log-magnitude spectrum and normalizes it. For the fused inputs the combined image is first split back into its two images, each transformed on its own, then put back together. The point is what the spectrum looks like: a sharp focused image keeps its high frequencies, a defocused one loses them, and astigmatism turns the spectrum into an ellipse stretched to one side. The network therefore sees the defect almost drawn out for it, instead of learning the same cue from raw pixels. Figure 1 shows where the layer sits in the model, and Figure 2 gives a defocused-stigmatic example with its spectrum.

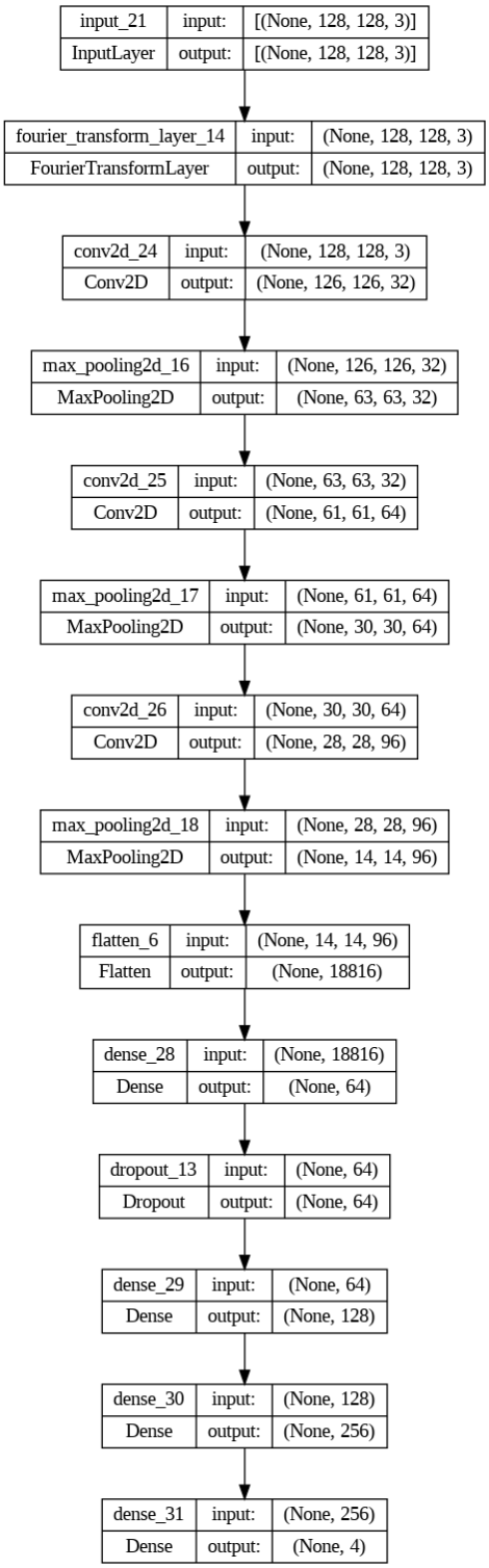


Fig. 1: Architecture of the proposed convolutional neural network with a custom Fourier-transform layer, classifying 128×128 electron microscope images into four focus and astigmatism classes (adopted from Skricilova, 2024)

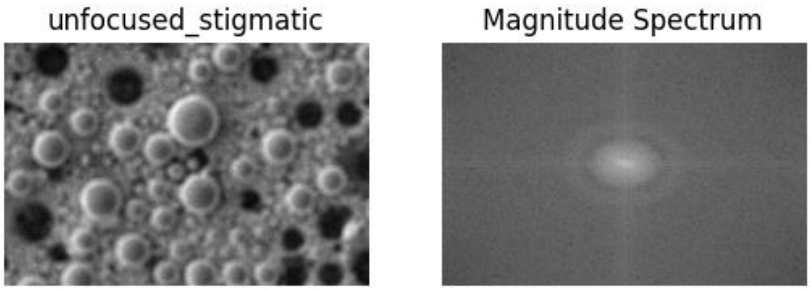


Fig. 2: Example electron microscope image of the defocused-stigmatic class with its corresponding Fourier magnitude spectrum used as model input (adopted from Skricilova, 2024)

3 RESULTS

All accuracies were measured on the held-out test set and are rounded to whole percent. Table 1 reports the base model across the three fusion types, with and without the Fourier layer; Table 2 sets the best base model against the three pretrained networks. The best row in each table is shown in bold.

The Fourier layer changed the outcome on every fusion type, and by a wide margin. Channel-axis fusion went from 55% to 94%, the output-layer variant from 49% to 93%, and horizontal fusion from 65% to 89%. Without the layer the ranking of the fusion types is reversed, with horizontal best at 65% and channel-axis near the bottom at 55%. There is no single best way to combine the two images; it depends on whether the Fourier layer is present (Skricilova, 2024).

The small custom model came out on top at 94%, ahead of all three pretrained networks. Among those, the lightest one, MobileNetV2, did best at 87%, while the two heavier networks landed near chance for a four-class problem (ResNet50 44%, EfficientNetV2L 52%). The Fourier layer did not help the pretrained models, since their best results all came with it switched off, the opposite of what it did for the shallow model. The chosen model for the framework is the custom network with channel-axis fusion and the Fourier layer.

Tab. 1 Base model accuracy by input fusion type, without and with the Fourier-transform layer

Fusion type	Accuracy without FFT	Accuracy with FFT
Horizontal	65%	89%
Channel-axis	**55%**	**94%**
Output-layer	49%	93%

Tab. 2 Best base model against pretrained ImageNet models (feature extraction, best parameter combination per model)

Model	Fusion	FFT layer	Accuracy
Custom base model	**channel-axis**	**yes**	**94%**
MobileNetV2	output-layer	no	87%
ResNet50	output-layer	no	44%
EfficientNetV2L	output-layer	no	52%

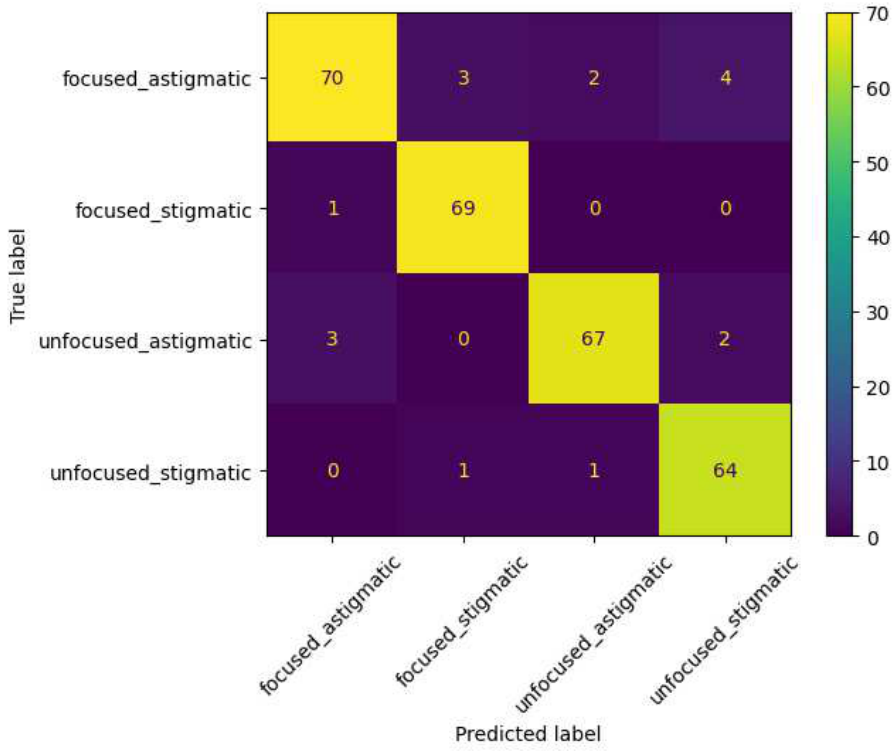


Fig. 3: Confusion matrix of the trained model on the test set across the four classes (focused or defocused × stigmatic or astigmatic) (adopted from Skricilova, 2024)

Where the errors fall (Figure 3) tells the rest of the story. Focused-stigmatic images, the easy case, were classified correctly in almost 100% of cases. Astigmatism was the hard part: images containing it accounted for more than 82% of all misclassifications. The single most common mistake was a focused-astigmatic image read as defocused-stigmatic, the pair that looks identical in the first shot and can only be told apart from the second.

4 DISCUSSION

The FFT-based preprocessing layer appears to help because it changes the representation of the input images before the convolutional layers process them. Defocus is associated with a loss of high--frequency information, while astigmatism can appear as an elongated pattern in the frequency spectrum. This is also why Fourier-based representations have been used in classical focusing and astigmatism-correction methods (Ogasawara *et al.*, 1999; Lu, Zhang and Li, 2018). In this study, adding the FFT layer substantially improved the base model across all tested fusion strategies: the best non-FFT variant reached 65%, whereas the best FFT variant reached 94%. This should be interpreted as an empirical result for the tested configurations, not as proof that the FFT layer alone would always produce the full difference.

A notable result was that the small custom model outperformed the tested pre-trained ImageNet models, and that the Fourier layer improved the custom model but not the best-performing pretrained configurations. A plausible reading is that ImageNet pretraining teaches a network to

chase fine natural-image detail that is not present in a frequency spectrum of a defect. The deep networks keep looking for structure the transformed input has thrown away, whereas a shallow model with few parameters takes the spectrum at face value. This fits the wider observation that a purpose-built CNN can match or beat heavier machinery on a narrow task like blur detection (Szandała, 2020). It is a useful caution against reaching for the largest pretrained model by default.

Two images are the cost of asking a richer question. Single-image spectrum methods (Lu, Zhang and Li, 2018) need only one shot but cannot pull apart a combined astigmatism-and-defocus state; the pair makes that separation possible at the price of a second exposure. Data scarcity was the other constraint: the model overfitted early, and only enlarging the dataset fivefold together with augmentation and L2 regularization brought it under control. L2 alone was the most effective single regularizer but not enough on its own.

The 94% should be read with one caveat. The study used a single sample type, tin balls on carbon, whose round regular shape makes defects stand out. On an irregular natural sample the Fourier layer might read the sample's own shape as astigmatism, so the figure holds for this well-behaved specimen rather than for arbitrary samples. The model is presented as suitable for the regression-testing framework; its behaviour once deployed there was not measured here.

5 CONCLUSION

A three-convolution CNN using image pairs, channel-axis fusion and a custom Fourier-transform layer achieved 94% test accuracy when classifying SEM images into four focus and astigmatism classes. This result indicates that the proposed approach is promising for integration into an automated regression-testing framework, although the study was limited to one sample type and the behaviour of the classifier inside a deployed testing workflow was not measured. The tested FFT variants substantially outperformed the corresponding variants without FFT, and the small custom model also outperformed the three tested pretrained ImageNet models. Useful next steps are training on more varied and irregular samples, adding further defect classes and evaluating the classifier in the target testing framework.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Skricilova, 2024) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

AGGARWAL, C. C. 2018. *Neural Networks and Deep Learning*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-94463-0

AMMANN, P., OFFUTT, J. 2008. *Introduction to Software Testing*. Cambridge: Cambridge University Press. ISBN 978-0-521-88038-1.

ERASMUS, S. J., SMITH, K. C. A. 1982. An automatic focusing and astigmatism correction system for the SEM and CTEM. *Journal of Microscopy*. 127(2), 185-199. ISSN 0022-2720. https://doi.org/10.1111/j.1365-2818.1982.tb00412.x

HE, K., ZHANG, X., REN, S., SUN, J. 2016. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770-778. ISBN 978-1-4673-8851-1. https://doi.org/10.1109/CVPR.2016.90

KINGMA, D. P., BA, J. 2014. Adam: A Method for Stochastic Optimization. arXiv preprint. *arXiv*:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*. 60(6), 84–90. https://doi.org/10.1145/3065386

LECUN, Y., BENGIO, Y., HINTON, G. 2015. Deep learning. *Nature*. 521(7553), 436-444. ISSN 0028-0836. https://doi.org/10.1038/nature14539

LEE, W., KIM, H., HAN, S., YOON, J. 2021. Robust autofocusing for scanning electron microscopy based on a dual deep learning network. *Scientific Reports*. 11(1), 20933. ISSN 2045-2322. https://doi.org/10.1038/s41598-021-00412-5

LU, Y., ZHANG, X., LI, H. 2018. A simplified focusing and astigmatism correction method for a scanning electron microscope. *AIP Advances*. 8, 015124. ISSN 2158-3226. https://doi.org/10.1063/1.5009683

NA, J., KIM, G., KANG, S.-H., KIM, S.-J., LEE, S. 2021. Deep learning-based discriminative refocusing of scanning electron microscopy images for materials science. *Acta Materialia*. 214, 116987. https://doi.org/10.1016/j.actamat.2021.116987

OGASAWARA, M., SATO, M., KOMURO, M., KAGA, R. 1999. Automatic Focusing and Astigmatism Correction Method Based on Fourier Transform of Scanning Electron Microscope Images. *Japanese Journal of Applied Physics*. 38(2R), 957. ISSN 0021-4922. https://doi.org/10.1143/JJAP.38.957

SCHUBERT, P. J., SAXENA, R., KORNFELD, J. 2024. DeepFocus: fast focus and astigmatism correction for electron microscopy. *Nature Communications*. 15, 948. https://doi.org/10.1038/s41467-024-45042-3

SKRICILOVA, M. 2024. *Classification of electron microscope images using an artificial neural network* [in Czech]. Master's thesis, head: Trenz, O. Brno: Mendel University in Brno. .

SZANDAŁA, T. 2020. Convolutional Neural Network for Blur Images Detection as an Alternative for Laplacian Method. In: *2020 IEEE Symposium Series on Computational Intelligence (SSCI).* pp. 2901-2904. ISBN 978-1-7281-2547-3. https://doi.org/10.1109/SSCI47803.2020.9308594

Contact information

Magdalena Skricilova: e-mail: Hugouse@seznam.cz
Oldrich Trenz: e-mail: oldrich.trenz@mendelu.cz

POST-QUANTUM CRYPTOGRAPHY: CURRENT STATE OF STANDARDIZATION AND DEPLOYMENT

Tomáš Sochor[1,2], Hana Sochorová[3]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

[2]European Research University, Sokolská 33, 702 00 Ostrava, Czech Republic

[3]Institute of Laboratory medicine, Institute of Medical Biophysics and Medical Information Technology, Faculty of Medicine, University of Ostrava, Syllabova 19, 703 00 Ostrava, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_034

ABSTRACT

The paper reviews the quantum computing advances and their impact on classical cryptography. the aspect of potential breaking of present (believably secure) cryptographic algorithms. Recent progress in the abilities of quantum computers as well as the standardization of new generation of cryptographic algorithms resistant to quantum-computing attacks is summarized. The classical cryptographic algorithms are reviewed and corresponding threats from quantum algorithms were summarized. In addition, existing symmetric and hash algorithms under quantum attacks is discussed and recommendations for improving security in long-term are formulated.

Keywords: quantum computing, cryptographic algorithm, post-quantum cryptography, symmetric cryptography, asymmetric cryptography, hashing, key exchange
JEL Code: O33, L86, C81

1 INTRODUCTION

Communication through the internet (as well as through of other computer networks) is often vulnerable to various types of attacks ranging from eavesdropping through blocking up to silent modifications. Original insecure application protocols belonging to TCP/IP protocol suite (like telnet, FTP, SMTP, later prevailed by HTTP) massively used in early stages of internet have been later gradually replaced by their more secure alternatives (e.g. SSH and HTTPS) so that the above risks were eliminated. However, even the design of such secure protocols has been proposed decades ago reflecting the 1990's knowledge in cryptography when potential breaking of cryptographic algorithms like RSA, SHA and DES (or later AES) seemed to be aeons away.



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

Nevertheless, recent progress in quantum computing (as described e.g. by Gill *et al.* 2025) has changed the perspective a lot. Now, the question does not seem to sound whether quantum computers will be able to perform computational operations for wide public but when they are going to do so. As a result, so far rather theoretical algorithms applicable to quantum computers able to solve some problems much quickly than classical computers (like factorization of large integers) will become widely available not only for researchers, academia and industry, but also for hackers. This could pose a significant threat to the security of information and communication systems relying on classical cryptography and therefore it is necessary to reflect this threat in near-future improvement of them, namely by implementation of quantum-attack-resistant algorithms.

2 RELEVANT EXISTING RESEARCH

Despite quantum computers of various types have been investigated for at for several decades, the real implementations started to emerge just in recent years. Although all present quantum computers are still rather experimental devices with very limited (inf any) practical impact, it is almost impossible when they overcome their experimental status and start their availability for real calculations.

Among the most promising applications, in addition to simulation of complex physical, chemical and other natural systems, there are calculations allowing significant simplifying (lowering the complexity) of computations problems having non-polynomial (exponential) computational complexity. Most of such non-polynomial-complex problems (e.g. problem of factorization of large numbers) have been and still are a fundament of various cryptographical systems, starting from RSA in 1970's (see Rivest, Shamir, Adleman 1977).

Such a threat (in the form of future potential availability of quantum computers being able to decrypt/break such cryptosystems quickly) poses a significant risk to all systems relying on cryptographic algorithms potentially prone to quantum-computing break. The reason is simple: many systems cannot change their cryptographic algorithms to more resistant quickly enough. Moreover, even if the decryption of vulnerable communication is not feasible now, still potentially sensitive communication can be captured by attackers now and decrypted later. This approach is called "Harvest now, decrypt later" (see Singh, 2024). For certain portion of data, it is still a significant risk, even though most of the encrypted communication would not feel compromised when decryption will happen in, say, 5 or 10 years from now, as demonstrated by Mascelli& Rodden M. 2025. Among the identified use-cases, bitcoin and its underlyying middleware blockchain could be at risk because of later unauthorized decryption of their communication.

There are numerous papers and articles devoted to the topic of post-quantum cryptography focusing on various aspects thereof. Among them, Kumar& Pattnaik (2020) provides a review of existing solutions, Joseph *et al.* (2022) focuses on PQC transition, and Lysenko *et al.* (2024) emphasizes general security aspects of post-quantum cryptography.

3 CRYPTOGRAPHIC ALGORITHMS REVIEW FROM QUANTUM COMPUTING PERSPECTIVE

3.1 Symmetric encryption algorithms

Currently, AES (as standardized by National Institute of Standards and Technology 2001) is the dominant symmetric encryption algorithm, definitely (see e.g. Ganes *et al.* 2025) and in fact, there is no real competition for it since 3DES as the last remaining alternative was deprecated by NIST in 2023.

Several recent studies (e.g. Hosseini, S.M. and Pilaram, H., 2024.) demonstrate that AES is relatively well prepared for quantum cumputing era. Despite the fact that Grover quantum algorithm reduces the time necessary for cracking AES encryption order of this type of search from $O(n)$ to $O(\sqrt{n})$, no other risks for AES have been presented. Therefore, one can conclude that for the near "quantum" future, it is sufficient to increase (ideally diouble) the key length to keep the same security level. Even though the current AES standard offers only three key lengths (128, 192 and 256 bits, respectively), with mostly used the maximum, it would not be difficult to extend the algorithms for its usage with longer keys.

3.2 Asymmetric encryption and digital signature algorithms

Asymmetric encry0ption applications do not date far in the history because RSA as the first asymmetric cryptographic algorithm was described in Rivest, Shamir, Adleman 1977. Since then, other asymmetric cryptosystems, namely Elliptic Curve Cryptography (ECC) and ElGamal, have been developed (see e.g. Zhang 2021), and asymmetric cryptosystems have spread rapidly because they have a significant advantage over symmetric algorithms, namely they can operate safely without having a covert channel for distributing a shared secret that is necessary for symmetric algorithms.

On the other hand, a significant drawback of asymmetric algorithm lies in their higher computational complexity reflecting in significantly higher number of operations necessary to encrypt or decrypt a chunk of data in comparison to adequately strong symmetric algorithm. Therefore, a combination of symmetric and asymmetric encryption is massively applied in practical implementations. The more detailed description of the combination is given in section 4.

There is a significant threat from a quantum-computer-run Shor algorithm because it converts the exponentially complex problem of finding of secret factors into a problem with polynomial complexity. Therefore, public availability of quantum computers that is expected in the near future, will pose a threat to the security of any cryptosystem that does not replace them with post--quantum alternatives that are summarized in Section 3.5.

To summarize those observations and findings, one must conclude that virtually all classical asymmetric encryption algorithms are not quantum-attack-resistant and therefore they must be replaced with post-quantum alternatives so that to keep the cryptosystem security in the long--term. The same is true for digital signature algorithms because they are mostly based on the same or very similar mathematical fundaments to asymmetric encryption.

Unlike for symmetric algorithms, here, replacement is necessary. Fortunately, despite of the fact that there is no direct replacement algorithm among those standardized as post-quantum (see section 3), we can still find a solution in combination of FIPS203 that can efficiently replace the asymmetric cryptography in the most frequent use-case, namely the key negotiation (key exchange) as a part of TLS handshaking between www server and a web browser.

3.3 Hashing algorithms

Together with encryption algorithms, hashing completes the cryptographic algorithmic triad necessary to perform a safe encryption. Similarly to encryption, hashing algorithms could be also potentially threatened by quantum simplification of certain operations necessary to break hashing algorithms, typically by simplifying finding collisions.

The most significant threat come from Grover quantum algorithm that can reduce the time necessary for performing pre-image or collision attacks to $O(\sqrt{n})$, or $O(\sqrt[3]{n})$, respectively (see Hosseini&Pilaram 2024). This means that a short-term solution to mitigate the threat could be an increase of the hash length (with existing hashing algorithm offering a reasonable reserve for that) while in long-term, replacing of existing hashing algorithms with quantum-resistant is the only feasible way.

3.4 Key exchange algorithms

Among cryptographic protocols, key exchange plays and important complementary role. Key exchange algorithms do not serve only for seldom operation of replacing the oldy keys with new ones, but they are employed every day, for instance at every key negotiation phase of TLS connection that is mentioned in section 4.

The most commonly key exchange algorithm class, Diffie-Hellman, is based on similar mathematical problem like RSA and therefore it is also threatened by Shor algorithm when run of a quantum computer. Therefore, the recommendation for Diffie-Hellman algorithm is almost the same as for RSA: the only way how to keep any cryptosystem long-term secure is the replacement of Diffie-Hellman algorithm with its post-quantum alternatives. Here, the new FIPS203 standard provides a solution.

3.5 Standardization of quantum-resistant algorithms

In response to rapidly emerging threat to classical algorithms from rapidly improving quantum computers, several new quantum-attack-resistant algorithms have been developed and some of them have been NIST-standardized recently. Despite the standardization process started already in 2016, first final PQC standards were presented in 2024. The following specifies the first generation of post-quantum cryptography:

- FIPS 203: Module-Lattice-Based Key-Encapsulation Mechanism (ML-KEM), derived from the CRYSTALS-KYBER submission. ML-KEM is a KEM (for key establishment) based on module--LWE. It includes parameter sets ML-KEM-512, ML-KEM-768, and ML-KEM-1024 (increasing security levels). This KEM can be used in hybrid TLS or other protocols for key exchange (e.g. combining with X25519). (National Institute of Standards and Technology 2024a)
- FIPS 204: Module-Lattice-Based Digital Signature Algorithm (ML-DSA), derived from CRYSTALS-Dilithium. ML-DSA provides three parameter sets (ML-DSA-44, -65, -87) corresponding to different security categories. It yields signatures comparable in size to existing ECC schemes and is intended for general-purpose signing (certificates, code signing, etc.) (SEE National Institute of Standards and Technology 2024B).
- FIPS 205: Stateless Hash-Based Digital Signature Algorithm (SLH-DSA), derived from SPHINCS+. SLH-DSA is a hash-based signature scheme with parameter variants (128s/128f, 192s/192f, 256s/256f) that trade off signature size and speed. It provides long-term security even against quantum adversaries, at the cost of larger signature sizes. Details are available in National Institute of Standards and Technology 2024c.

Tab. 1 Overview of NIST Post-quantum cryptographic standards

Algorithm name	Purpose	Security assumption	Parameter sets (FIPS)	Deployment
ML-KEM (FIPS 203)	Key-encapsulation (key exchange/ establishment)	Module-LWE (lattice-based)	ML-KEM-512, ML-KEM-768, ML-KEM-1024	TLS key exchange (incl. hybrid with X25519); larger key shares may impact bandwidth/ latency.
ML-DSA (FIPS 204)	Digital signatures (general-purpose)	Module-lattice (Module-LWE / Module-SIS)	ML-DSA-44, ML-DSA-65, ML-DSA-87	PKI, code-signing, firmware signing; larger signatures/public keys than ECDSA/ Ed25519.
SLH-DSA (FIPS 205)	Stateless hash-based digital signatures	Hash function properties (preimage/ collision resistance)	128s/128f, 192s/192f, 256s/256f (SHA-256 / SHAKE-256 variants)	Conservative choice; suitable for long-term signatures (e.g., firmware); larger signatures.

Sources: National Institute of Standards and Technology 2024a- 2024c

The description of new post-quantum standards by NIST is provided in Table 1.

As the research results mentioned in section 2, there are quantum-attack-resistant algorithms available. On the other hand, available implementations in existing systems are still quite rare – see details in section 5.

4 APPLICATION OF CURRENT CRYPTOGRAPHIC ALGORITHMS

Among the applications of current cryptographic algorithms, the Transport Layer Security protocol (TLS) plays a dominant role because it is a fundamental component of HTTPS (Secure HTTPS) that is massively used (nowadays at about 95% of HTTP traffic uses TLS as statistics show, e.g. Google, 2026, and Kerschbaumer *et al.* 2025). TLS was standardized by IETF in several versions starting with version 1.1 in 1999, followed by version 1.2 in RFC5246 in 2008. Currently last version 1.3 comes from 2018, 8446 (Rescorla E., 2018).

The operation of TLS combines multiple cryptographic techniques so that the encryption operates both securely and quickly. In the beginning, the TLS client should rely on the asymmetric key of the server from its certificate only. Here, the client can send the small data chunks to the server (encrypted using slower asymmetric algorithm) as a part of Diffie-Hellman key-exchange process in establishing the session key. After successful key exchange, the rest of the communication involving encryption of large data files is performed using a symmetric encryption algorithm that is significantly faster.

Other secure protocols (like SSH, and even IPSecure) follow a similar pattern: they use classical asymmetric cryptographic algorithms for distribution or exchange of temporary session key, and symmetric cryptography for payload encryption. All these classical asymmetric cryptographic components (certificates, key exchange etc.) need post-quantum cryptography replacements.

5 AVAILABLE IMPLEMENTATIONS

The above-mentioned post-quantum cryptographic algorithms have been implemented partly in some major software packages (see Demir 2025). Another recent study (Ahmed *et al* 2025) demonstrates that for some libraries, developers work hard on supporting post-quantum cryptography. A recent news from OpenSSL developers announced that starting April 2025, the above mentioned standard post-quantum cryptographic algorithms are implemented since version 3.5. It seems likely that implementations in other popular libraries will follow but it is a good idea to verify the status before use of a specific library.

6 CONCLUSIONS

As the above results demonstrate, all classes of cryptographic algorithms, with a small expectation of relatively few impacted symmetric ones, are threatened by potential availability of quantum computers in relatively near future. Therefore, it is critical so that the existing cryptosystems are modified so that potentially weak cryptographic algorithms are promptly replaced with post-quantum alternatives. These alternative cryptographic algorithms are already available thank to their recent standardization by NIST so there are no formal obstacles preventing information and communication system administrators and owners to start the process of replacement.

In summary, the transition to quantum-safe cryptography is underway. With standards in hand and growing support in cryptographic libraries, organizations can begin implementing post-quantum cryptography now to safeguard against the coming quantum era.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

REFERENCES

AHMED, N., ZHANG, L., GANGOPADHYAY, A. 2025. August. A survey of post-quantum cryptography support in cryptographic libraries. In: *2025 IEEE International Conference on Quantum Computing and Engineering (QCE).* Vol. 1, pp. 906-917. IEEE.

BARRETT-DANES, F., AHMAD, F. 2025. Quantum computing and cybersecurity: a rigorous systematic review of emerging threats, post-quantum solutions, and research directions (2019–2024). *Discover Applied Sciences*. 7(10), 1083.

DEMIR, E. D., BILGIN, B., ONBASLI, M. C. 2025. Performance analysis and industry deployment of post-quantum cryptography algorithms. *arXiv preprint arXiv*:2503.12952.

GILL, S. S., CETINKAYA, O., MARRONE, S., CLAUDINO, D., HAUNSCHILD, D., SCHLOTE, L., WU, H., OTTAVIANI, C., LIU, X., MACHUPALLI, S. P., KAUR, K. 2025. Quantum computing: Vision and challenges. In: *Quantum Computing*. pp. 19-42. Morgan Kaufmann.

GANESH, R., KHAN, B. U. I., KHAN, A. R., KAMSIN, A. B. 2025. A panoramic survey of the advanced encryption standard: from architecture to security analysis, key management, real-world applications, and post-quantum challenges. *International Journal of Information Security*. 24(5), 1-45.

GOOGLE. 2026. *HTTPS encryption on the web*. [Accessed: 2026-01-07] https://transparencyreport.google.com/https/overview?hl=en

HOSSEINI, S. M., PILARAM, H. 2024. *A Comprehensive Review of Post-Quantum Cryptography: Challenges and Advances*. Cryptology ePrint Archive.

JOSEPH, D., MISOCZKI, R., MANZANO, M., TRICOT, J., PINUAGA, F. D., LACOMBE, O., LEICHENAUER, S., HIDARY, J., VENABLES, P. AND HANSEN, R. 2022. Transitioning organizations to post-quantum cryptography. *Nature*. 605(7909), 237-243.

KERSCHBAUMER, C., BRAUN, F., FRIEDBERGER, S., JÜRGENS, M. 2025. The State of https Adoption on the Web . *Mozilla Research*. [Accessed: 2026-04-14] https://research.mozilla.org/files/2025/03/the_state_of_https_adoption_on_the_web.pdf

KUMAR, M., PATTNAIK, P. 2020. September. Post quantum cryptography (pqc)-an overview. In: *2020 IEEE High Performance Extreme Computing Conference (HPEC).* pp. 1-9. IEEE.

LYSENKO, S., BOKHONKO, O., VOROBIYOV, V., GAJ, P. 2024. Method for identifying cyberattacks based on the use of social engineering over the phone. In: *IntelITSIS*. pp. 318-329. [Accesssed: 2026-01-11] https://ceur-ws.org/Vol-3675/paper23.pdf

MASCELLI, J., RODDEN, M. 2025. *"Harvest Now Decrypt Later": Examining Post-Quantum Cryptography and the Data Privacy Risks for Distributed Ledger Networks*. FEDS Working Paper No. 2025-93. [Accessed: 2026-01-12] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5956313

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. 2001. FIPS 197: Advanced Encryption Standard (AES). *Nist.org* [Accessed: 2026-01-07] https://doi.org/10.6028/NIST.FIPS.197-upd1

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. 2024a. *FIPS 203. Module-Lattice-Based Key-Encapsulation Mechanism Standard*. [Accessed: 2026-01-12] https://csrc.nist.gov/pubs/fips/203/final

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. 2024b. *FIPS 204. Module-Lattice-Based Digital Signature Standard*. [Accessed: 2026-01-12] https://csrc.nist.gov/pubs/fips/204/final

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. 2024c. FIPS 205. Stateless Hash-Based Digital Signature Standard. [Accessed: 2026-01-12] https://csrc.nist.gov/pubs/fips/205/final

RESCORLA, E. 2018. The Transport Layer Security (TLS) Protocol Version 1.3. *IETF.* [Accessed: 2026-01-12] https://datatracker.ietf.org/doc/html/rfc8446

RIVEST, R. L., SHAMIR, A., ADLEMAN, L. 1977. *On Digital Signatures and Public-Key Cryptosystems* (No. MITLCSTM82).

SINGH, H. 2024. Managing the Quantum Cybersecurity Threat: Harvest Now, Decrypt Later. In: *Quantum Computing.* pp. 142-158. CRC Press.

ZHANG, Q. 2021. An overview and analysis of hybrid encryption: the combination of symmetric encryption and asymmetric encryption. In: *2021 2nd international conference on computing and data science (CDS).* pp. 616-622. IEEE.

Contact information

Tomáš Sochor: e-mail: tomas.sochor@mendelu.cz, e-mail: tomas.sochor@eruni.org
Hana Sochorová: email: hana.sochorova@osu.cz

A UNIFIED FORMAT FOR AGGREGATING PRE-TRAINED POSE ESTIMATORS INTO CUSTOM HUMAN-SKELETON DATASETS

Damian Sova[1], Oldrich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_035

ABSTRACT

Human pose estimation works well on benchmark datasets but loses accuracy once a model is deployed on real footage with varying distances, crowds and occlusion. Building a labelled dataset for each new deployment is expensive, so this work asks whether existing pre-trained detectors can be combined to generate such a dataset cheaply, without retraining any of them. The proposed Unified Format runs three public detectors (PoseNet, MoveNet, MMPose), maps their native keypoints into one common 27-keypoint skeleton, matches detected people across models and fuses them by weighted average. The method was evaluated on the COCO WholeBody validation subset (2,693 images with humans, 11,004 ground-truth instances) using OKS, average percentage error and mean squared error. The fused skeleton provides usable positional localisation (APE 3.3%, MSE 1,213.84) and predicts slightly more pose instances than MMPose alone (70.39% vs 69.84%), but it remains weaker than MMPose on the main accuracy-oriented metrics (OKS 0.23 vs 0.28; correct keypoints 37.93% vs 43.62%). The result is an honest, reproducible tool chain rather than a new state of the art.

Keywords: human pose estimation, neural networks, dataset generation, data fusion, COCO WholeBody, OKS
JEL Code: C45

1 INTRODUCTION

Human pose estimation has achieved strong benchmark results, but its deployment in real-world footage remains difficult. On curated benchmarks such as COCO the leading networks reach high scores, yet the same networks drop in accuracy when they meet conditions the benchmark never contained: people far from the camera, several bodies overlapping, a limb hidden behind furniture. Public datasets simply do not cover these cases (Alinezhad Noghre *et al.*, 2022).

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

The usual remedy is to collect and annotate a dataset for the specific site. That is slow and costly. Manual labelling of joint positions takes hours per hundred frames, and sensor-based motion capture needs laboratory hardware and controlled lighting (Zheng *et al.*, 2020). For a small company that wants to deploy a pose detector in, say, a care home, this cost is often the part of the project that does not get funded. The present work grew out of exactly such a setting: a healthcare start-up interested in detecting people who fall. The company later withdrew and supplied no production data, so the downstream retraining could not be carried out; the dataset-generation method described here was nonetheless fully built and tested on public data.

The idea is to avoid new annotation altogether. Several pre-trained detectors already exist, each with its own strengths and its own skeleton definition. If their outputs are standardised into one common format and fused, the combination can label real footage automatically, and the labelled frames then serve as a custom dataset. No detector is retrained. The aim of this paper is to describe that pipeline and to report how close the fused output gets to ground truth, including where it falls short of a strong single detector.

1.1 Pose detectors and their skeletons

Modern pose estimators are convolutional networks, typically following one of two designs: regress joint coordinates directly (Toshev and Szegedy, 2014), or predict heatmaps and read off the peaks (Sun *et al.*, 2019). Top-down variants first detect a person with a region-proposal network such as Faster R-CNN (Ren *et al.*, 2015) and then locate joints inside each box. The three detectors used here span this range. The MediaPipe Pose landmark model, referred to in the source thesis as PoseNet, is a lightweight single-person detector producing 33 body landmarks (BlazePose, 2020). MoveNet, also from Google, is a fast single-person model with 17 keypoints (Movenet, 2024). MMPose from OpenMMLab is heavier and multi-person, with 133 keypoints covering body, face, hands and feet, built on the HRNet backbone (Sun *et al.*, 2019; MMPose Contributors, 2020).

1.2 Combining detector outputs

Fusing the predictions of several models is not a new idea in pose estimation. Zhang and Chen (2022) report decision-level fusion of detector outputs, and automated dataset generation for real-world pose estimation has been proposed before (Alinezhad Noghre *et al.*, 2022). What differs here is the framing: the goal is not a higher benchmark score but a cheap, reusable labelling tool, so the design favours simplicity (a fixed common skeleton, a weighted average) over a learned fusion model that would itself need training data.

2 METHODOLOGY

The pipeline has four stages, shown in Figure 1: run each detector on the input frames, map every model's native keypoints into the Unified Format, match the people detected by different models against each other, and fuse the matched detections into one skeleton per person. A final stage scores the fused output against ground-truth annotations.

The Unified Format is a 27-keypoint skeleton primarily derived from the 17 MoveNet body joints and extended with selected hand, foot and face points, so that the richer MMPose output can be reduced to a common representation while the sparser models remain compatible. Each keypoint

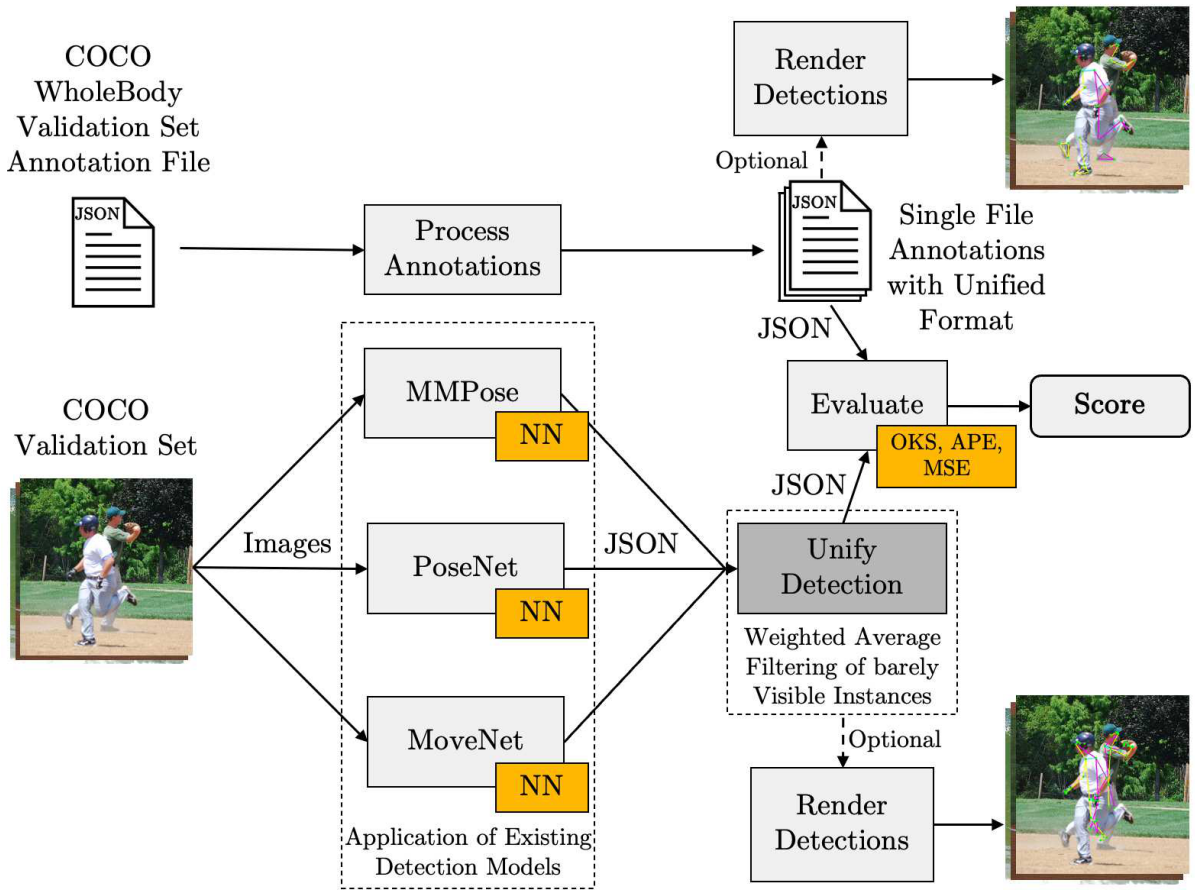


Fig. 1: Overall processing pipeline of the proposed evaluation framework, from existing detection models (MMPose, PoseNet, MoveNet) through the Unified Format and weighted-average instance unification to OKS / APE / MSE scoring against COCO WholeBody annotations (adopted from Sova 2024).

is stored as a triple [x in pixels, y in pixels, visibility]. The skeleton is multi-person. Its layout, with the ID assigned to every joint, is given in Figure 2.

Matching people across models is the awkward part. MMPose returns a bounding box for every detected person, so its instances can be compared by box overlap (intersection over union). PoseNet and MoveNet return no box and only one person, which makes the correspondence harder and is, as the results show, a source of error. Once instances are matched, the fused keypoint is a weighted average of the per-model coordinates and visibility. The weights reflect trust in each model: MMPose 0.5, PoseNet 0.3, MoveNet 0.2 (Table 1). MMPose carries the highest weight because the thesis evaluation and manual inspection identified it as the most reliable of the three input detectors for the processed data. The weighted average is used only where all three models supply a keypoint; for hand and foot points, which MoveNet does not produce, a plain average is taken.

Evaluation used the COCO WholeBody validation 2017 subset (Xu *et al.* 2022): 5,000 images, of which 2,693 contain people, giving 11,004 ground-truth pose instances. A predicted instance is matched to a ground-truth instance when their bounding boxes overlap with IoU above 65%. Three metrics are then computed per matched pair. Object Keypoint Similarity (OKS) is the COCO standard, normalised by joint-specific constants (COCO Consortium 2024); because COCO ships these constants only for 17 keypoints, they were recomputed for all 27 from an initial run, and

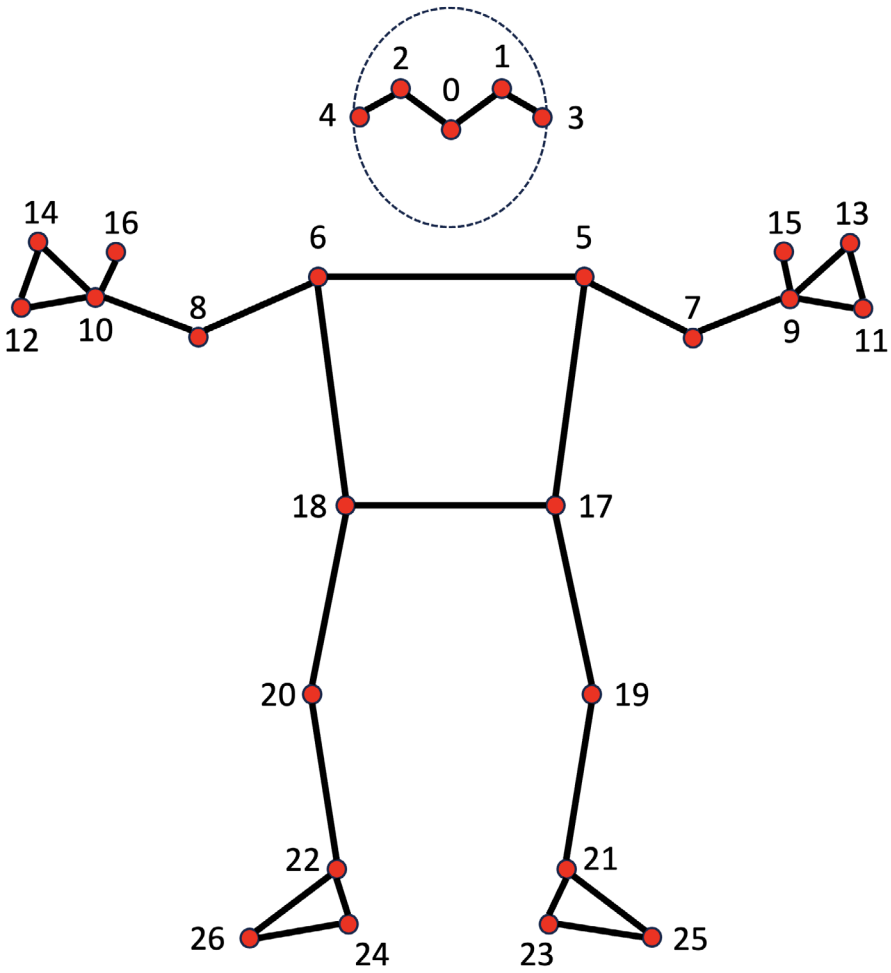


Fig. 2: Structure of the proposed Unified Format skeleton with 27 keypoints and their IDs, providing a common pose representation across heterogeneous detection models (adopted from Sova 2024)

a keypoint counts as correct when its OKS exceeds 0.45. Average percentage error (APE) and mean squared error (MSE) measure positional distance in pixels. Invisible ground-truth keypoints are skipped. All experiments ran on a MacBook Pro M1 Max (64 GB RAM); no NVIDIA GPU was available, which is why NVIDIA BodyPoseNet from the original assignment could not be included.

Tab. 1 Table 1: Input detectors and their weights in the Unified Format fusion

Model	Keypoints	Multi-person	Weight
PoseNet	33	No	0.3
MoveNet	17	No	0.2
MMPose	**133**	**Yes**	**0.5**
Unified Format	27	Yes	–

3 RESULTS

Table 2 compares the Unified Format against MMPose alone on the full evaluation set. MMPose is the reference here, not a competitor: it is the strongest of the three inputs, so the question is how much the fusion costs or gains relative to its best single member. The accuracy-oriented columns favour MMPose, and the table marks them in bold.

Two readings come out of this. The fused skeleton reaches acceptable positional localisation in the context of dataset-generation support (APE 3.3%, MSE 1,213.84), but it is clearly less pre-cise than MMPose alone (APE 0.95%, MSE 123.73). On the strict similarity metric, it is behind: OKS 0.23 against 0.28, and 37.93% correct keypoints against 43.62%. The one place where fusion helps the count is coverage. The Unified Format predicted 7,746 instances (70.39%) against 7,685 (69.84%) for MMPose alone, so adding the lighter detectors recovered a few people that MMPose missed (Sova, 2024). The number of instances that actually matched ground truth, 5,833, is the same for both, which already hints that the extra detections were not the well-localised ones.

Tab. 2 Overall evaluation on COCO WholeBody (2,693 images, 11,004 ground-truth instances) — Unified Format vs MMPose alone

Statistic	Unified Format	MMPose
Object Keypoint Similarity (OKS)	0.230	0.280
Average percentage error (APE)	3.30%	0.95%
Mean squared error (MSE)	1,213.84	123.73
Correct keypoints detected	37.93%	43.62%
Matching poses (instances)	5,833 (53.01%)	5,833 (53.01%)
Predicted poses (instances)	7,746 (70.39%)	7,685 (69.84%)
Ground-truth poses (instances)	11,004	11,004



Fig. 3: Example of multi-person skeleton detection in a cluttered real-world scene, illustrating estimation under realistic conditions (adopted from Sova, 2024)

4 DISCUSSION

Why does a fusion of three detectors land below its own best member on OKS? The averaging is the cause, and it works against the method in two ways. First, the input detectors were picked for speed and small size, not accuracy: the lightning versions of PoseNet and MoveNet run at 10 to 30 FPS on a CPU but were never the most precise option the vendor-reported figures are not directly comparable across these models and datasets; in the thesis evaluation on COCO WholeBody, MMPose was the strongest component and therefore received the highest fusion weight. Feeding a fast-but-rough estimate into a weighted mean pulls the result away from the accurate MMPose detection rather than toward it. Second, OKS is unforgiving. A pose that looks right to a human can still score low, because OKS shrinks fast with distance once the joint-specific tolerance is exceeded. The combination of mediocre inputs and a strict metric explains most of the gap.

The matching step is the other weak point, and it is more interesting because it is fixable. When PoseNet or MoveNet returns no bounding box, the pipeline has to guess which of their single-person detections belongs to which MMPose person. Get that wrong and the average fuses two different people into one skeleton. A single image made this failure visible: one bad single-person prediction, averaged with an otherwise correct MMPose pose, drove that image's APE to 84.22% and its OKS to zero. That case is not representative of the dataset, but it shows the mechanism cleanly. So the honest reading of the positional numbers is that the fused output localises people satisfactorily and, on the matched instances, slightly better than the individual lighter models, yet there is no evidence of a real improvement over the benchmark performance of a strong single detector.

Where does this leave the practical goal? For generating a custom dataset cheaply, coverage and plausible localisation matter more than topping a leaderboard, and on those terms the tool is usable: it labels real footage with no new annotation and recovers a few extra people. But the accuracy ceiling is set by the inputs. Replacing speed-oriented input detectors with accuracy-oriented alternatives would likely improve OKS, although at the cost of runtime and hardware requirements. One detail worth recording for anyone reusing the code: an early version squared the OKS object scale by mistake, which pushed OKS into an implausible range; removing the square restored sane values. Small bug, large effect on the headline number.

5 CONCLUSION

This work built and tested a way to turn three off-the-shelf pose detectors into a single labelling tool, without retraining any of them. The Unified Format maps PoseNet, MoveNet and MMPose into one 27-keypoint skeleton, matches people across models and fuses them by weighted average. On COCO WholeBody it keeps positional error low (APE 3.3%, MSE 1,213.84) and detects marginally more poses than the strongest single model, but it does not reach that model's similarity score (OKS 0.23 vs 0.28). The contribution is therefore a reproducible, honestly evaluated pipeline, not a new accuracy record.

The two limits found here also point the way forward. Instance matching should not rely on bounding boxes that two of the three detectors never produce; a similarity-based vote across keypoints would be more reliable. The fusion itself could weight detectors per instance instead of globally, and flag outliers before averaging so that one wrong pose cannot spoil a correct one. With those changes, and with accuracy-oriented detectors as inputs, the same framework could be further tested as a candidate pipeline for generating datasets suitable for training or fine-tuning a deployable model.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Sova, 2024) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

ALINEZHAD NOGHRE, G., DANESH PAZHO, A., SANCHEZ, J. 2022. *ADG-Pose: Automated Dataset Generation for Real-World Human Pose Estimation*. Cham: Springer International Publishing. ISBN 978-3-031-09282-4. https://doi.org/10.1007/978-3-031-09282-4_22

BLAZEPOSE. 2020. *BlazePose: On-device Real-time Body Pose Tracking*. CVPR Workshop on Computer Vision for AR/VR. https://storage.googleapis.com/mediapipe-assets/Model%20Card%20BlazePose%20GHUM%203D.pdf

COCO CONSORTIUM. 2024. *COCO Dataset: Keypoints Evaluation*. https://cocodataset.org/#keypoints-eval

MMPOSE CONTRIBUTORS. 2020. *OpenMMLab Pose Estimation Toolbox and Benchmark*. https://github.com/open-mmlab/mmpose

MOVENET. 2024. MoveNet: Efficient Human Pose Estimation in the Cloud and on Mobile. *TensorFlow*. https://www.tensorflow.org/hub/tutorials/movenet

REN, S., HE, K., GIRSHICK, R. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:*1506.01497v3. https://doi.org/10.48550/arXiv.1506.01497

SOVA, D. 2024. *Automated neural network learning for higher accuracy human skeleton detection under realistic conditions*. Master's thesis, head: Trenz, O. Brno: Mendel University in Brno.

SUN, K., XIAO, B., LIU, D. 2019. Deep High-Resolution Representation Learning for Human Pose Estimation. *arXiv:* 1902.09212. https://arxiv.org/abs/1902.09212

TOSHEV, A., SZEGEDY, C. 2014. DeepPose: Human Pose Estimation via Deep Neural Networks. In: *IEEE CVPR*. New York: IEEE. https://doi.org/10.1109/CVPR.2014.214

XU, L., LIU, W., XU, J. 2022. *ZoomNAS: Searching for Whole-body Human Pose Estimation in the Wild*. IEEE TPAMI. https://github.com/jin-s13/COCO-WholeBody

ZHANG, Y., CHEN, W. 2022. Decision-level information fusion powered human pose estimation. *Applied Intelligence*. 53, 2161–2172. https://doi.org/10.1007/s10489-022-03623-z

ZHENG, C., WU, W., CHEN, C. 2020. Deep Learning-Based Human Pose Estimation: A Survey. ACM Computing Surveys. *arXiv*:2012.13392 https://doi.org/10.48550/arXiv.2012.13392

Contact information

Damian Sova: e-mail: damian.sova10@gmail.com

Oldrich Trenz: e-mail: oldrich.trenz@mendelu.cz

TREESTORAGE: HIERARCHICAL STORAGE WITH VERSIONING AND INHERITED ACCESS CONTROL IN A UNIVERSITY INFORMATION SYSTEM

Petr Straka[1], Oldrich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_036

ABSTRACT

Large information systems accumulate knowledge that is often transferred informally rather than stored in the system itself. This article describes the design and working prototype of Treestorage, a reusable hierarchical storage component for INET, the economic and personnel information system of Masaryk University. Treestorage is intended to organise agenda-related HTML notes, references to binary documents and parameterised SQL reports in a tree structure, with read and edit rights inherited from parent nodes and version history stored for edited nodes. Before implementing the component, available alternatives at the university were assessed, especially Google Drive, OneDrive and git-based services such as GitHub/GitLab; these options were not selected because they would either keep data and authentication outside INET or fail key requirements such as fine-grained folder-level read access. The implemented data model uses an adjacency list, reads nodes as a flat list and rebuilds the tree in the front end, while inherited rights are resolved through a PostgreSQL materialised view based on a recursive query. The result is a functional prototype that can be tested internally and further integrated after related INET 2.0 components are completed.

Keywords: information systems, hierarchical storage, access control, versioning, materialized view, adjacency list
JEL Code: C88

1 INTRODUCTION

INET, the information system of Masaryk University, has grown since 1998 and was opened to users in 2000 (Měcháček and Kohoutková, 2000). Today it runs hundreds of applications across economics, personnel and study agendas. With that scale comes a problem that has little to do with code. The knowledge of how a particular specialised application is meant to be used, together with the rarely-run SQL reports that answer recurring administrative questions, is held by individual

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

people. When one of them leaves or changes role, the knowledge often goes too. The system offers no shared, access-controlled place to keep notes, document links and reusable reports next to the agenda they belong to.

That is the gap Treestorage was built to close. The thesis set out to design and implement a hierarchical storage component that any INET application can embed, with read and edit rights set on a parent node cascading to everything below it and an earlier state of any node recoverable from its history. One constraint shaped everything else: the component had to live inside INET rather than beside it, so that this supporting knowledge stays under the same authentication and the same data umbrella as the agenda it serves.

1.1 Why not an existing product

The first question is whether a ready-made tool would do. The thesis examined storage options available in the university environment, especially document-management services such as Google Drive and OneDrive and git-based services such as GitHub/GitLab. For the most relevant API-based options, small proof-of-concept clients were prepared to test basic operations such as listing, creating, moving and deleting nodes and setting permissions. Google Drive and OneDrive provide storage and inheriting permissions, but integration would depend on external service accounts, OAuth/Microsoft Entra mechanisms and data stored outside INET. Confluence and Outline were considered only as contextual knowledge-base tools and were not treated as direct replacements for report-aware hierarchical storage.

The git family, represented mainly by GitHub/GitLab, looked promising because versioning is its native strength, yet it fits this task only partially. Git-based services cannot naturally grant read permission on a single folder; access is typically handled at the repository level. They also do not track empty folders, which complicates the preparation of an empty tree structure, and binary documents are not handled as efficiently as textual files. These limitations, together with the need to keep the component inside INET and under its existing authentication and agenda context, were the main reasons for building a custom component rather than adopting an external product directly.

1.2 Storing a tree in a relational database

A hierarchy is a rooted tree, and a relational database can keep one in several ways. The thesis weighed four classical patterns (adjacency list, path enumeration, the closure table and the nested set model) against the tree idioms of a graph database such as Neo4j with its Cypher language and of a document database such as MongoDB (Samoilenko, 2025; Neo4j n.d.; MongoDB n.d.). The MongoDB single-document tree caps at a 16 MB document, which rules it out for an open-ended store.

An adjacency list was chosen: every node simply stores the identifier of its parent. It is the simplest model to write against and the cheapest to keep correct. It also sidesteps the usual objection, that reading a subtree needs recursion in SQL, by not doing the recursion in SQL at all. The whole space is read as one flat list and the tree is reassembled in the front end, a linear-time operation on the client. For the expected size of a Treestorage space this is more than enough, as the next sections show.

2 METHODOLOGY

The work followed an engineering path, not an experimental one: requirements analysis, a survey of existing solutions, database design, and a prototype implementation, closing with an outlook chapter. There is no quantitative benchmark, because the deliverable is a working component inside an existing framework rather than a measured model.

The component sits inside INET 2.0 and is implemented in Java 25, while data are stored in a PostgreSQL relational database. The user interface uses INET's internal web-UI framework, whose components are rendered to HTML through XSL transforms. The dynamic tree is handled on the client side in JavaScript/React and supports node movement through the native HTML drag-and-drop API. Persistence is implemented through JPA entities, which reduces the need to write SQL manually for standard data operations.

Treestorage tables carry the prefix `wdoc_`, the cross-agenda report tables `a_`. A `wdoc_space` row splits storage into independent spaces, one tree per major INET agenda such as Economics, each tied to a domain. The central wdoc_node table holds the space, the parent (the nearest ancestor), the node type, a bilingual name in Czech and English, four generic content columns `(str_val_1, str_val_2, int_val_1, int_val_2)` and audit columns. Every change copies the data columns into a new row of `wdoc_node_log`, and that is how version history is kept. Access rights live in `wdoc_node_acl`, one row per node and person with two booleans, `can_read` and `can_edit`. Reports are defined in `a_query` with their parameters in `a_query_param`. Figure 1 shows the full model.

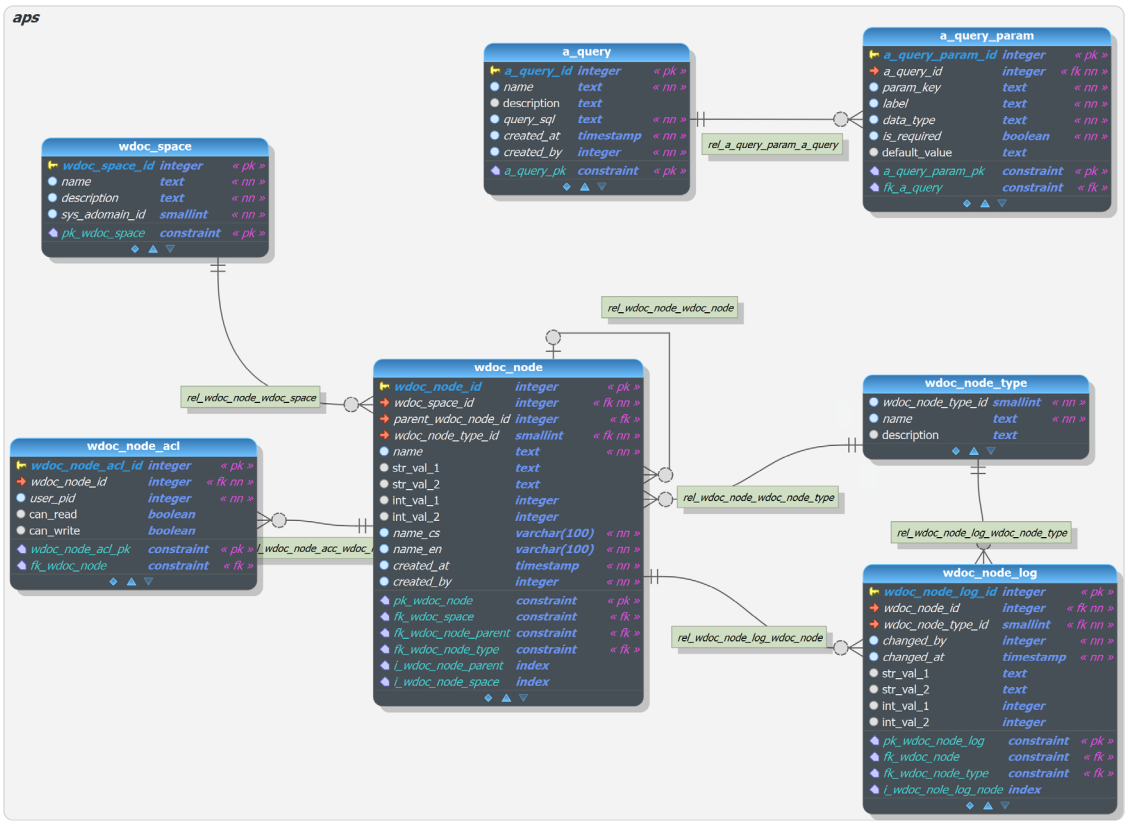


Fig. 1: Entity-relationship diagram of the Treestorage data model, showing the core node table with its versioning log, node types, access-control list and dynamic-query tables (adopted from Straka 2026)

A node is one of four types in this first version: DIR (a folder), HTML (text or rich-text content intended to be edited through the INET WYSIWYG component), DOCUMENT (a reference to binary data expected to be handled through DMSX) and QUERY (a reference to a dynamic SQL report). The four generic content columns allow a single table to carry different node types and make later extension possible without changing the basic structure. The use-case model defines three roles. A reader views permitted nodes, browses history and runs reports. An editor edits and adds nodes and manages rights on relevant parts of the tree. A superuser creates level-zero nodes and maintains the top-level structure.

Inherited rights are the part of the design worth dwelling on, because they are resolved without paying for recursion on every request. Rights cascade from parent to child, but recomputing that for each page load would be wasteful. Instead the effective rights are precomputed in a PostgreSQL materialised view, `wdoc_node_acl_effective`, built from a recursive query: it starts at the root nodes paired with every person who holds at least one explicit right anywhere, then walks down the tree, inheriting the parent's rights unless an explicit ACL row overrides them. A unique index on the person-and-node pair keeps lookups fast, and the view is refreshed only when rights actually change (a database trigger on the ACL table is noted as a likely later addition). Reports get one more safeguard: when a report is saved, an SQL checker parses the body and blocks the save unless every declared variable is also defined in the parameter table, so a report cannot be stored in a state that would fail at run time.

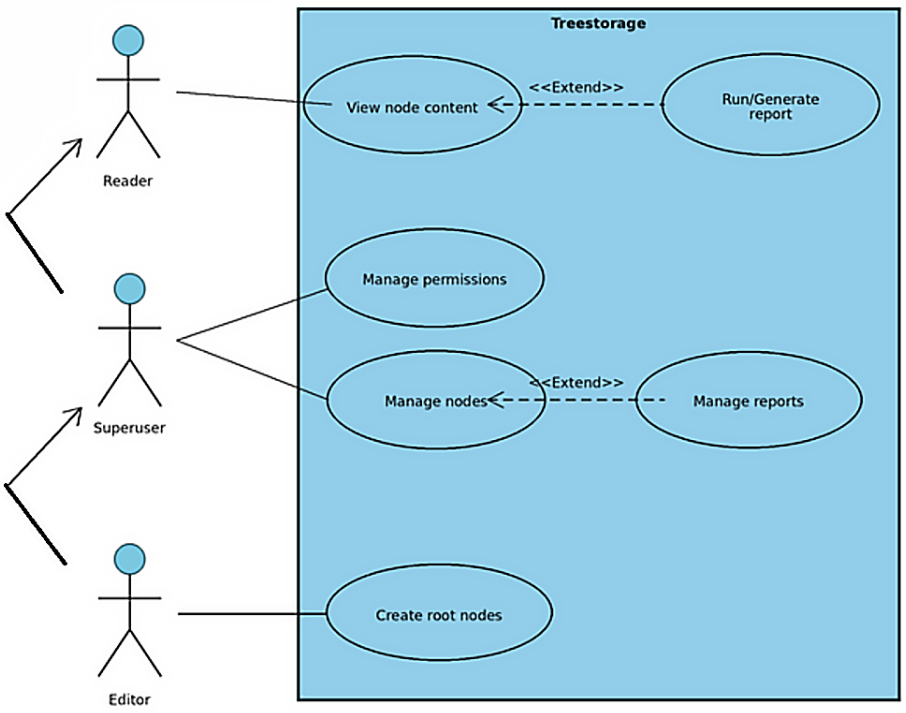


Fig. 2: UML use case diagram of the Treestorage system with the Reader, Superuser and Editor roles and their operations over hierarchical nodes (adopted from Straka 2026)

3 RESULTS

The outcome is a functional prototype, not a set of measurements. Treestorage works as a reusable INET component: it embeds a hierarchical tree into any INET application and, for each node, shows a rich-text note, a document link, or a dynamic SQL report tied to that agenda. Version history is in place (earlier states sit in `wdoc_node_log` and can be reopened), and the inheriting read/edit access control runs through the materialised view rather than per-request recursion. The first version ships four node types, three user roles, three report-parameter input types (text, numeric and a person-id select), and two new report tables (Straka, 2026). Figure 3 shows the dynamic--report component with a generated result table.

A second concrete result is the survey that justified building the component at all. Table 1 should be read as an authorial synthesis of selected figures and constraints reported in chapter 5 of the thesis, not as a direct vendor table. The comparison is primarily functional: whether a service can be integrated into INET, whether it supports the required access-control model, and whether it can hold the intended types of content. Storage capacities and commercial prices are not fully comparable across the alternatives because they use different licensing models and institutional arrangements.

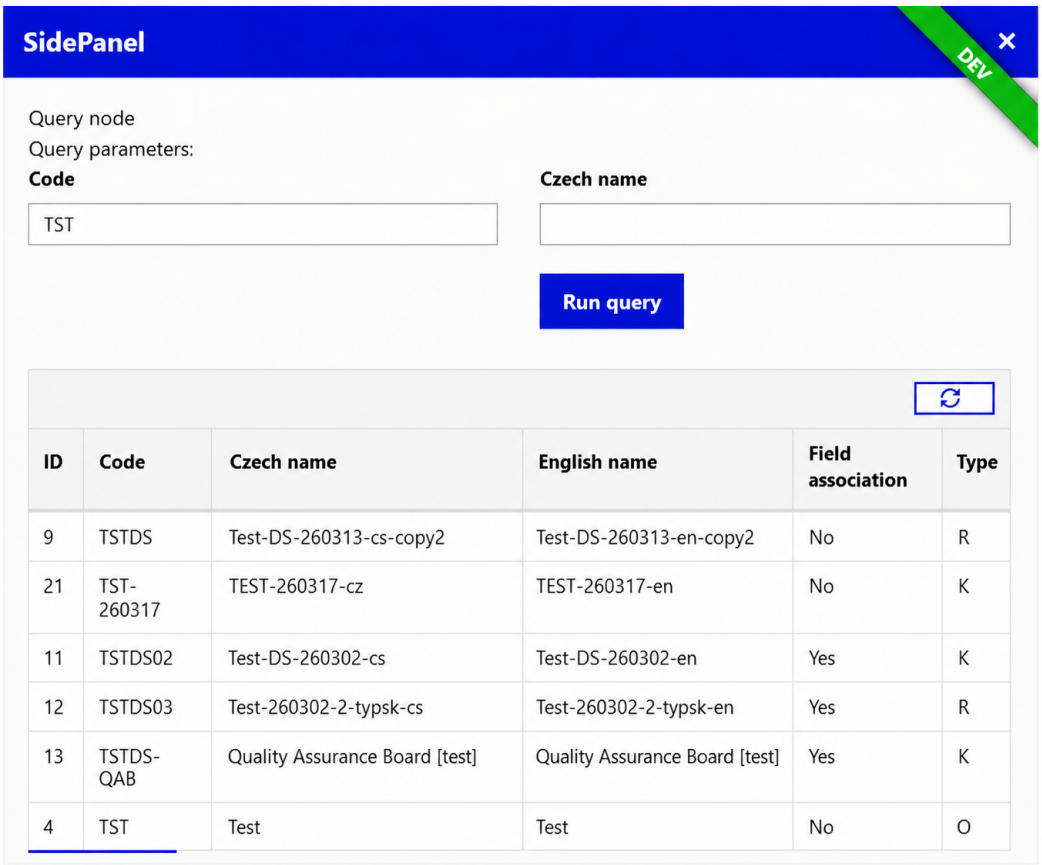


Fig. 3: Web user interface of the dynamic-report component, listing queried nodes with their bilingual names, type and versioning metadata (adopted from Straka, 2026).

Tab. 1 Off-the-shelf storage options surveyed before building Treestorage (authorial synthesis of figures cited in chapter 5; "—" = not stated). Commercial list prices are per user and are not the rates the university actually pays

Criterion	Google Drive	OneDrive (MS 365)	GitHub / git
Education storage	100 TB shared / org (Fundamentals, free)	100 GB staff, 30 GB students	repo limit 10 GB
API / integration	Drive API v3	Microsoft Graph (+ MS Entra)	GitHub REST (git trees)
API rate limit	20 000 req / 100 s	3 000 req / 5 min (+ 100 GB down/h, 50 GB up/h)	metered
File-size limit	—	—	warning > 1 MB, cap 100 MB; LFS 10–250 GiB
List price (per user)	14 USD / 2 TB tier	12.5 USD (M365 Copilot Standard)	4 USD / user
Per-folder read control	yes (inheriting), Google accounts only	yes, also outside MS account	**no (whole repo only)**
Move a node	yes	yes	**only by delete + recreate**
Empty folder	yes	yes	**no**
Binary + versioning	yes	yes	**inefficient (full file per version)**
Indicative cost for all of MU	~3.5 mil. CZK/yr (33 000 students × 5 USD edu)	~5.2 mil. CZK/yr (2024 university licence)	hard to cost

The status is honest about its limits. The component is a functional prototype, testable internally at the university computing centre, and it waits on two INET 2.0 pieces, the WYSIWYG editor and the DMSX binary-upload component, before it can go to real users. It has not yet been measured under load; the assumptions behind that are spelled out in the discussion.

4 DISCUSSION

Why an adjacency list, when the textbook objection (no cheap subtree query in SQL) is so well known? In Treestorage, the objection is weakened by the intended use case. The component reads the nodes of a space as a flat list and rebuilds the tree in the browser, so the server does not repeatedly assemble subtrees through recursive SQL. The expected size of a space is relatively small, and

the design therefore favours simplicity, easy updates and maintainability. More elaborate alternatives such as a closure table or the nested set model would improve some query patterns, but at the cost of more complex write operations and maintenance.

Access control makes the opposite trade. Inheritance is genuinely recursive, so a materialised view backed by a recursive query precomputes it once and refreshes only when rights change, instead of recomputing it on every request. This is the right call when reads vastly outnumber rights changes, which is the normal pattern for stored knowledge: people read notes and run reports far more often than they regrant permissions. The cost is staleness between refreshes, handled for now by refreshing from inside the application. The moment rights start changing often, a database trigger on the ACL table becomes the cleaner answer, and the design already flags that path.

The survey result is worth dwelling on because it is tempting to assume that a hosted service is always the pragmatic choice. For an internal, access-controlled and report-aware store, this is not straightforward. Google Drive and OneDrive solve document storage but do not naturally solve the INET-specific problem of keeping notes, document references and reusable reports directly inside the agenda context. Git solves versioning well, but not fine-grained folder-level read rights or efficient binary-document handling. Treestorage's value is therefore not only storage capacity, but the integration of knowledge, documents and reports into the same controlled INET environment.

5 CONCLUSION

Treestorage delivers a working prototype of a component that a large information system lacked: a hierarchical place where supporting knowledge, document references and reusable reports can be kept under the same authentication and rights model as the agenda they explain. The design uses an adjacency-list tree, browser-side reconstruction, node-level versioning and inherited rights materialised in PostgreSQL. Existing university tools were useful reference points but not direct substitutes because they either externalise the data or do not pro-vide the required integration and permission model. The next step is to complete the dependent INET 2.0 components, verify performance under realistic load and pilot the tool in one agenda before wider use.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Straka, 2026) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

ATLASSIAN. n.d. Confluence — Atlassian. *Atlassian*. https://www.atlassian.com/software/confluence

GITHUB. n.d. GitHub Acceptable Use Policies. *GitHub*. https://docs.github.com/en/site-policy/acceptable-use-policies/github-acceptable-use-policies

GOOGLE. 2026. Google Drive API release notes. *Google for Developers*. https://developers.google.com/workspace/drive/release-notes

MĚCHÁČEK, J., KOHOUTKOVÁ, J. 2000. Intranetový server Informačního systému MU [in Czech]. *Zpravodaj ÚVT MU*. XI(1), 4–7. https://webserver.ics.muni.cz/zpravodaj/articles/194.html

MICROSOFT. 2025. Avoid getting throttled or blocked in SharePoint Online. *Microsoft Learn*. https://learn.microsoft.com/en-us/sharepoint/dev/general-development/how-to-avoid-getting-throttled-or-blocked-in-sharepoint-online

MONGODB. n.d. Model Tree Structures. *MongoDB*. https://www.mongodb.com/docs/manual/applications/data-models-tree-structures/

MOZILLA. n.d. JavaScript. *MDN Web Docs*. https://developer.mozilla.org/en-US/docs/Web/JavaScript

NEO4J. n.d. *Neo4j documentation*. https://neo4j.com/docs/

ORACLE. n.d. What is Java technology and why do I need it? *java.com*. https://www.java.com/en/download/help/whatis_java.html

OUTLINE. n.d. *Outline*. https://www.getoutline.com

POSTGRESQL. n.d. PostgreSQL Contributor Profiles. *The PostgreSQL Global Development Group*. https://www.postgresql.org/community/contributors/

RICHARDS, R. 2025. 2025 Google Workspace for Education Pricing and Licensing Updates. *Trafera*. https://www.trafera.com/blog/2025-google-workspace-for-education-pricing-and-licensing-updates/

SAMOILENKO, S. 2025. Part 1: Nested sets. *Medium*. https://medium.com/@serhiisamoilenko/part-1-nested-sets-239e70f59beb

SEYBOLT, B. 2025. Major Google for Education Licensing and Price Changes. *Cumulus Global*. https://www.cumulusglobal.com/major-google-for-education-licensing-and-price-changes/

STRAKA, P. 2026. *Treestorage — hierarchical storage with versioning and access control* [in Czech]. Master's thesis, head: Trenz, O. Brno: Mendel University in Brno.

Contact information

Petr Straka: e-mail: xstraka6@node.mendelu.cz

Oldrich Trenz: e-mail: oldrich.trenz@mendelu.cz

WEB APPLICATION FOR GENERATING TECHNICAL DOCUMENTS USING LARGE LANGUAGE MODELS

Martin Šulc[1], Ivo Pisařovic[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_037

ABSTRACT

This paper investigates the viability of leveraging large language models (LLMs) to generate formal corporate documents, moving beyond rigid, traditional template-based approaches. The study aims to verify whether combining reference samples, natural language instructions, and parallel multi-model processing can yield a practical enterprise tool. The proposed microservices architecture utilizes a Node.js/LangChain backend to abstract calls across four LLM providers (OpenAI, Anthropic, Gemini, and Ollama). This is supported by Python microservices for input parsing and multi-format document generation (Word, PDF, Excel), a Vue.js frontend wizard, and PostgreSQL/Redis persistence. System validation involved end-to-end scenarios across five document types, benchmarking seven models, and user testing with ten participants. Results demonstrate the solution's efficacy. Furthermore, comparative analysis reveals that local open-source models achieve generation quality comparable to mid-range cloud services for Czech-language tasks, albeit with trade-offs in increased latency and initial hardware investments.

Keywords: large language models, document generation, web application, microservices
JEL Code: C88, L86

1 INTRODUCTION

Companies work with a large number of formal documents every day – technical documentation, protocols, emergency plans or notes. These documents need to be created, updated and maintained regularly, typically by a worker taking an older document of a similar type, manually editing it, adding current data from various sources and compiling the result into the required form. The whole process is slow and prone to errors. At the same time, there is empirical data that the deployment of generative language models in writing processes significantly increases worker productivity and output quality (Noy and Zhang, 2023), while extensive field experiments in companies have



This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

recorded double-digit productivity increases even for routine text tasks (Brynjolfsson *et al.*, 2025). These observations have framed a broader debate about the opportunities and risks of generative AI in professional practice (Dwivedi *et al.*, 2023).

Available tools for creating corporate documents only partially solve this problem. Template generators such as Docupilot (Docupilot, 2026) operate on the principle of predefined fields into which the system fills in data – an approach useful for repetitive documents with a fixed structure, but failing when it is necessary to adapt the content to a new context or generate text that was not in the template. Document extraction and analysis tools such as Azure AI Document Intelligence (Azure AI, 2025) only solve the opposite direction of the flow, i.e. converting existing documents into structured data. Platforms with integrated artificial intelligence (Writer, 2026) handle content generation, but they mainly focus on marketing and business output, not technical documentation.

Large language models bring new quality, as they can analyze the structure of input documents and generate new content based on requirements formulated in natural language (Ke *et al.*, 2025; Min *et al.*, 2023). However, the LLM market is experiencing rapid development and the number of commercially and locally available models is in the hundreds (Chang *et al.*, 2024), which for practical deployment means the need to introduce an abstraction layer above a specific provider so that model replacement does not require rewriting of application logic. Prompt-based learning has emerged as the central paradigm, where the behavior of the model is controlled by carefully designed system instructions and templates for user inputs without the need for fine-tuning (Liu *et al.*, 2023). At the same time, empirical evaluations across dozens of NLP tasks show that the quality of LLM outputs varies significantly by task and language (Kocoń *et al.*, 2023), which justifies conducting our own comparison on a specific scenario of generating formal documents in Czech. It also remains an open question to what extent current local open-source models available through runtime environments such as Ollama (Ollama, 2026) are able to compete with cloud services in this task.

The aim of the work (Šulc, 2026) on which this contribution is based was to design and implement a web application that fills the gap identified in the market overview. The user uploads older sample documents, formulates requirements in natural language, and the application generates a new formal document in the selected format (Word, PDF or Excel) using LLM. Key assumptions were support for several providers including local deployment, the possibility of parallel comparison of multiple models on the same assignment, a primary focus on the Czech language, and a user interface usable even for participants without detailed knowledge of artificial intelligence. The intention was not to create a production-ready product, but to explore currently available technologies, demonstrate their practical use for a given purpose, and verify the functionality of the resulting solution on real documents of a specific company.

This paper summarizes the design and selected results of the comparison of language models and user testing. The Methodology and Data chapter describes the chosen development approach, the microservices-based architecture, the technologies used, and the testing methodology. The Results chapter presents a comparison of seven models from three cloud providers and four on-premises variants, and selected conclusions from user testing. The Discussion and Conclusions then summarize the achievement of the objectives, identified weaknesses, and recommendations for further development, including considerations of the economic and environmental dimensions of LLM deployment.

2 METHODOLOGY AND DATA

The application was developed using an incremental approach (Hoda *et al.*, 2017), where functionality is delivered gradually in smaller units. The choice is based on the nature of the project, as many design decisions (for example, the optimal structure of the system instruction or the scope of the data model) depend on the knowledge gained only during experiments with real documents. A fully sequential waterfall approach was not suitable here, as some requirements were refined continuously based on the behavior of specific language models.

The system architecture is implemented as a set of microservices run via Docker Compose. This arrangement follows the principles described in Newman (2015) and allows for leveraging the strengths of different technologies where it makes the most sense, ensuring that the failure of one service does not affect the rest of the system, and facilitating independent scaling and deployment of individual components (Jamshidi *et al.*, 2018).

The core of the application is a back-end on the Node.js platform with the Express application framework and the LangChain.js library. This provides a unified abstraction over language model providers (OpenAI, Anthropic, Google Gemini and locally operated Ollama), so that adding another provider only requires implementing the corresponding adapter without interfering with the rest of the application logic (Topsakal and Akinci, 2023). The behavior of the model is controlled by a central system instruction passed to each call along with the user input, which corresponds to the prompt-based learning paradigm (Liu *et al.*, 2023): a single carefully designed instruction maintains output consistency (language, formatting, rules for subsequent modifications) across all models and eliminates duplication in the code. Communication with the client is carried out via REST API and WebSocket connection through the Socket.IO library, which allows sending continuous notifications about the processing status without burdening the server with periodic polling.

Specialized document operations are separated into two Python microservices. MarkItDown Parser uses the open-source MarkItDown library from Microsoft (MarkItDown, 2024) to convert input documents (PDF, DOCX, XLSX, PPTX, RTF) into a unified Markdown format, which is the optimal input for language models, as it preserves the structure (headings, lists, tables) in readable text form without noise. Document Generator then creates output files from structured data in Word (python-docx library), Excel (openpyxl) and PDF (reportlab) formats. The choice of Python for these operations is based on significantly more advanced libraries for working with document formats, which do not have a full-fledged equivalent in the Node.js ecosystem.

The front-end is implemented in Vue.js with the Quasar application framework, which provides an extensive set of components ready for creating a multi-step wizard without the need to integrate third-party libraries (Kaluža and Vukelić, 2018). Persistence combines a PostgreSQL relational database for all persistent data (sessions, files, instructions, results, conversations) with an in-memory Redis storage, which serves as a fast cache, a priority queue for asynchronous instruction processing, and a communication channel for WebSocket notifications. All services are containerized and executable with a single command (Merkel, 2014), eliminating problems with inconsistent environments between development and production installations. User interaction takes the form of a four-step wizard: uploading files and assigning roles, entering instructions, configuring the model (with the possibility of parallel selection of multiple models for comparison), and reviewing the results with export. The architecture of the entire system is depicted in Figure 1.

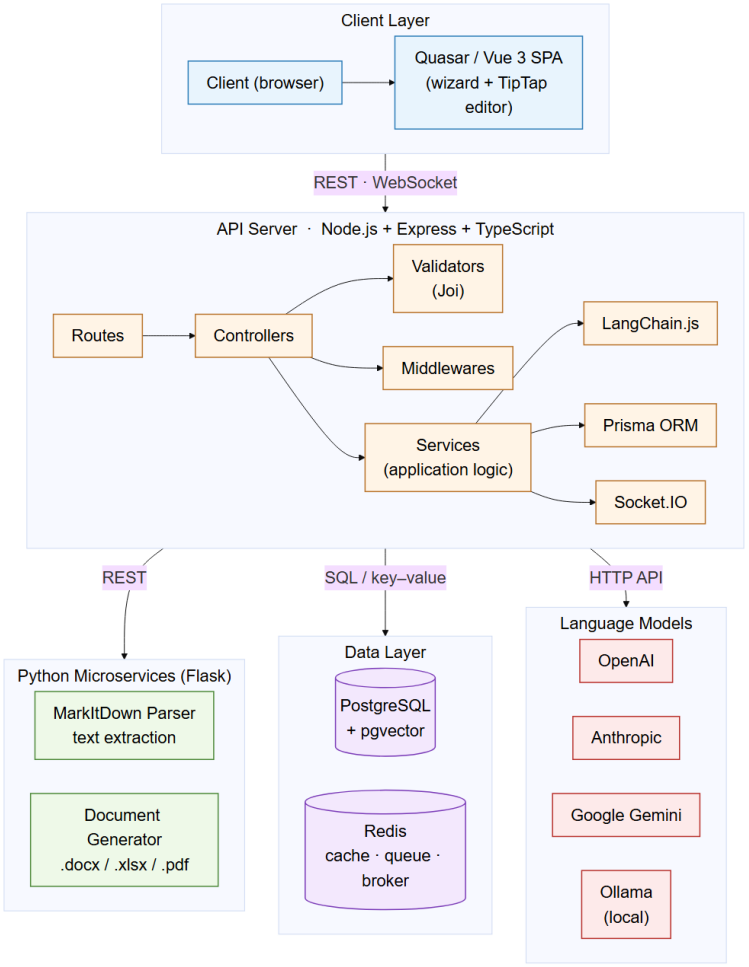


Fig. 1: System architecture with indication of main services (back-end, Python microservices, data layer, client) and their communication links (Šulc, 2026)

Testing was carried out in three layers. Unit and integration tests in the Jest application framework cover validators, provider configuration, REST API endpoints and input file parsing. End-to-end verification of generation quality was performed on five types of documents (technical report, annual report, multi-file project documentation, completed form and additional variants) using prepared scenarios with defined tasks and checklists. Comparison of language models included seven representatives – three cloud-based (OpenAI GPT-5.4, Anthropic Claude Sonnet 4.6, Google Gemini 3 Flash) and four local ones available via Ollama (Qwen 3:32b, Gemma 3:27b, Llama 3.1:8b and Mistral 7b). The design of evaluation criteria (quality of Czech, adherence to the required structure, processing speed, operational cost) is based on the approach of multidimensional LLM assessment across tasks described by Kocoń *et al.* (2023). The measurement was performed on a workstation with an NVIDIA RTX 5090 graphics card with 32 GB VRAM. User testing was based on Nielsen's (2024) recommendation for a combination of task scenarios and a think-aloud protocol and was conducted on a sample of ten participants (seven with an IT background, two economists, one translator), who were tasked with creating a template for future use from the completed form.

3 RESULTS

The result of the work is a functional application that meets the main objectives of the assignment. The application allows uploading documents in key formats (PDF, DOCX, XLSX, CSV, RTF, Markdown), converting them to Markdown via a Python microservice, compiling instructions with the ability to target a specific file or section, parallel calling of several language models from different providers and subsequent generation of structured outputs in Word, Excel and PDF formats. The user interface guides the user through the entire process linearly, and the Socket.IO library sends continuous notifications about the processing status. The specific form of the finished application is illustrated in Figure 2.

3.1 Comparison of language models

Key findings from the comparative testing of the models are summarized in Table 1. Several conclusions emerge from the measured values. Cloud models show significantly lower response times (typically 5–25 seconds) than local models (1–3 minutes), which corresponds to the computing capacity and optimization of the infrastructure of individual providers. The quality of the generated Czech and the ability to adhere to the specified structure are at a high level among the cloud models and reach the highest values in the Claude Sonnet 4.6 model. The observed variability in performance across models corresponds to the general finding from larger comparative studies, according to which no current model represents a clear winner in all tasks and metrics (Kocoń *et al.*, 2023; Min *et al.*, 2023).

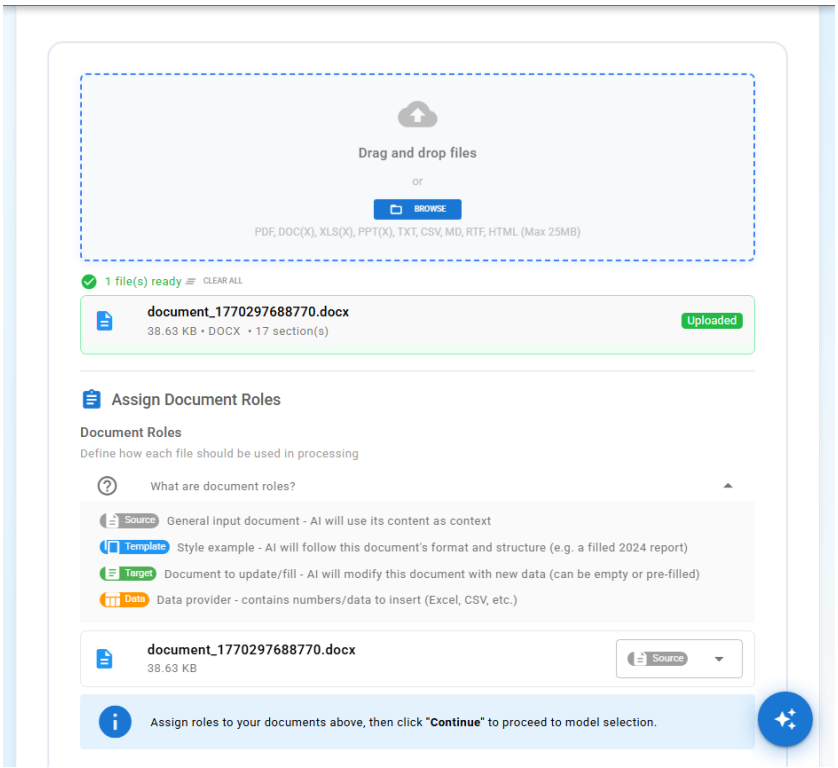


Fig. 2: Sample wizard user interface – first step with drag-and-drop file upload and role assignment (Šulc, 2026)

Tab. 1 Comparison of language models according to key features for generating formal documents in Czech

Model	Type	Context	Price / Specifications	Czech	Structure	Speed
GPT-5.4	cloud	1M	$2.50 / $10 per 1 M stream	8/10	8/10	10–20 seconds
Claude Sonnet 4.6	cloud	1M	3 / 15 USD for 1 M flow	9/10	8/10	15–25 seconds
Gemini 3 Flash	cloud	1M	$0.50 / $4 per 1 M flow	8/10	7/10	5–15 seconds
Qwen 3:32b	local	32,768	32 B / 24 GB VRAM	8/10	7/10	2–3 minutes
Gemma 3:27b	local	128K	27 B / 24 GB VRAM	7/10	6/10	2–3 minutes
Llama 3.1:8b	local	128K	8 B / 8 GB VRAM	6/10	4/10	1–1.5 minutes
Mistral 7b	local	8,192	7 B / 8 GB VRAM	3/10	2/10	≈ 2 minutes

Among the local models, Qwen 3:32b and Gemma 3:27b show significantly the highest quality of Czech and structured output, but they require a graphics card with at least 24 GB VRAM (typically NVIDIA RTX 3090 or 4090). Smaller models (Llama 3.1:8b and Mistral 7b) running on a regular consumer graphics card with 8 GB VRAM are on the verge of usability for generating longer formal texts in Czech – Mistral 7b in particular proved to be insufficient in tests. These findings have a direct impact on the decision to deploy in an organization with strict data protection requirements: a high-quality local alternative to cloud models is available, but at the cost of higher initial investments in hardware and several times longer response times.

3.2 User testing

Of the ten participants in the user test, all completed the task without external assistance. The application was generally rated as clear and no participant was visibly lost in the guide. Most participants used the parallel processing feature and compared the outputs of one local model (Qwen 3:32b or Gemma 3:27b) with the cloud-based GPT-5.4 simultaneously. This functionality provided users with immediate feedback on the differences between the models and made it easier to choose the appropriate variant for a specific type of task.

At the same time, two specific areas with a worse user experience were identified. The first is the file role selection in the initial step, where the labels template, source and target were not intuitively understandable for the task of creating a template from an existing document – the term template does not make intuitive sense here, since a template is only supposed to be an output, not an input. The second is the screen for selecting the output type before starting processing, where the labels of the individual options and their mutual relationship were not easily distinguishable, which led to uncertainty among some participants as to whether the settings were correct. This observation was consistent across professions and there was no significant difference between participants with and without an IT background.

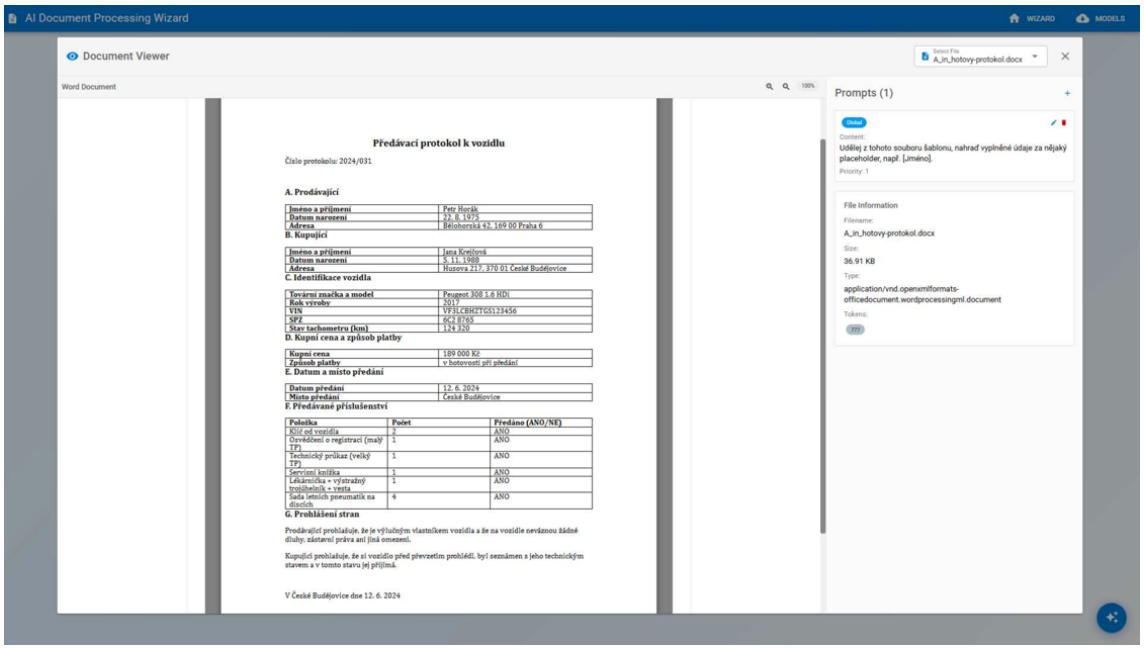


Fig. 3: Document Viewer – preview of an uploaded document along with user's prompt (Šulc, 2026)

A useful side benefit of testing was the discovery of a bug that was not replicated more often during actual development: some generated documents contained fragments of language model meta-comments (for example, sentences like "if you want, I can add..."), even though the system instruction explicitly prohibits such introductory and accompanying phrases. The cause is the undefined behavior of specific models and the need to handle this additionally at the level of subsequent response processing.

The Document Viewer component (see Figure 3), which allows viewing the input document alongside the instructions, was explicitly identified by participants as one of the most beneficial features. The difference in response speed between the cloud and on-premises models was perceived by participants, but in most cases was not considered a barrier. There were minor stylistic differences between the outputs of individual models, but the structure was maintained consistently across all models in most cases, which can be considered an achievement considering the comparison of a top-of-the-line cloud model with on-premises variants running on commonly available hardware. The ability to complete the task depended more on the nature of the specific input document than on the user's experience with information technology, suggesting that the wizard successfully lowers the barrier to entry for users without detailed knowledge of artificial intelligence.

4 DISCUSSION AND CONCLUSIONS

The verification demonstrated that the combination of an abstraction layer over language model providers, a microservice architecture, and a wizard interface represents a functional approach to automating the generation of formal documents that goes beyond the traditional template approach. The application is a fully containerized multi-component solution ready to run with a single command, significantly reducing the need for manual configuration. It fills the gap identified in the

market analysis: a complete open-source solution that would combine input document analysis, new content generation via LLM, support for multiple providers including local deployment, and a focus on technical documentation in Czech was not available at the time of development.

The main value of the resulting application, as revealed by user testing, is the combination of a clear guide, support for multiple AI providers including local variants, and the function of parallel comparison of models on the same assignment. The Document Viewer for simultaneous display of the input document and entered instructions was appreciated by the users as one of the most beneficial functions. The choice of the Vue.js application framework with Quasar proved to be suitable for the rapid creation of a comprehensive guide, which corresponds to the conclusions of comparative studies of front-end technologies (Bielak *et al*., 2022).

The identified weaknesses also indicate the direction of further development. The highest priority should be to rework problematic labels in the initial and final steps of the wizard, or to add contextual help directly in the interface, and to integrate subsequent output processing, which would automatically filter out meta-comments of the language model. The second significant topic is the stability of subsequent modifications within a single session, as the current version works reliably only for the initial instruction, while subsequent modifications of the same session are sensitive to the specific behavior of the model and do not achieve the same stability. The proposed solution includes dividing extensive modifications into smaller steps and more thorough work with the context of previous versions, or the use of structured output and function calls.

From the point of view of economic competitiveness, the work provides empirical evidence for deciding on the strategy for deploying LLM in a company. Selected local open-source models (especially Qwen 3:32b and Gemma 3:27b) achieve a quality comparable to mid-range cloud services for tasks of generating formal documents in Czech, albeit at the cost of several times longer response times and the necessary initial investment in hardware. This trade-off needs to be interpreted in the broader context of computing costs. Schwartz *et al.* (2020) introduced a distinction between Red AI (oriented to maximum performance at the cost of increasing computing demands) and Green AI (taking into account energy and operational efficiency) for deciding on the effectiveness of AI systems. A more detailed analysis of the energy and carbon footprint of operating large models (Patterson *et al*., 2022) shows that choosing a smaller local model can be an economically and environmentally rational alternative for organizations whose tasks do not require top-quality output or that work with sensitive data for which sending content to third-party servers is unacceptable. Conversely, for tasks with high document volumes and tolerance for cloud processing, cloud providers remain the choice with the best balance of speed, quality and operating costs. At the same time, growing empirical evidence on the productivity benefits of generative AI in the corporate environment (Noy and Zhang, 2023; Brynjolfsson *et al*., 2025) suggests that investing in document creation automation, whether via cloud or local deployment, represents a rational decision with a demonstrable return.

In conclusion, it can be stated that the main objectives of the work – design, implementation and verification of a web application for generating formal documents with support for large language models – have been fulfilled. The application was tested on five document types, end-to-end scenarios, seven language models and user testing with ten participants covering both technical and non-technical professions. The identified areas of further development (editing the interface labels, post-processing of the output, stabilization of subsequent edits and deeper integration of reasoning models) represent a direct continuation of the design described here.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Šulc, 2026) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the authors. The authors are responsible for the correctness of the final text.

REFERENCES

AZURE AI. 2025. Azure AI Document Intelligence documentation. *Microsoft*. [Accessed: 2026-02-14] https://learn.microsoft.com/en-us/azure/ai-services/document-intelligence/

BIELAK, D., BOREK, K., PLECHAWSKA-WÓJCIK, M. 2022. Web application performance analysis using Angular, React and Vue.js frameworks. *Journal of the Computer Sciences Institute*. 23, 100–106. https://doi.org/10.35784/jcsi.2827

BRYNJOLFSSON, E., LI, D., RAYMOND, L. R. 2025. Generative AI at Work. *Quarterly Journal of Economics*. 140(2), 889–942. https://doi.org/10.1093/qje/qjae044

CHANG, Y., WANG, X., WANG, J., WU, Y., ZHU, K., CHEN, H., YANG, L. *et al*. 2024. A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology*. 15(3), 1–45. https://doi.org/10.1145/3641289

DOCUPILOT. 2026. *Docupilot*. [Accessed: 2026-02-19]. https://www.docupilot.com

DWIVEDI, Y. K., KSHETRI, N., HUGHES, L., SLADE, E. L., JEYARAJ, A., KAR, A. K. *et al*. 2023. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. I*nternational Journal of Information Management*. 71, art.102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

HODA, R., SALLEH, N., GRUNDY, J., TEE, H. M. 2017. Systematic literature reviews in agile software development: A tertiary study. *Information and Software Technology*. 85, 60–70. https://doi.org/10.1016/j.infsof.2017.01.007

JAMSHIDI, P., PAHL, C., MENDONÇA, N. C., LEWIS, J. W., TILKOV, S. 2018. Microservices: The Journey So Far and Challenges Ahead. *IEEE Software*. 35(3), 24–35. https://doi.org/10.1109/MS.2018.2141039.

KALUŽA, M., VUKELIĆ, B. 2018. Comparison of front-end frameworks for web applications development. *Zbornik Veleučilišta u Rijeci*. 6(1), 261–282. https://doi.org/10.31784/zvr.6.1.19

KE, W., ZHENG, Y., LI, Y., XU, H., NIE, D., WANG, P., HE, Y. 2025. Large Language Models in Document Intelligence: A Comprehensive Survey, Recent Advances, Challenges, and Future Trends. *ACM Transactions on Information Systems*. 44, 1–64. https://doi.org/10.1145/3768156

KOCOŃ, J., CICHECKI, I., KASZYCA, O., KOCHANEK, M., SZYDŁO, D., BARAN, J. *et al*. 2023. ChatGPT: Jack of all trades, master of none. *Information Fusion*. 99, art.101861. https://doi.org/10.1016/j.inffus.2023.101861

LIU, P., YUAN, W., FU, J., JIANG, Z., HAYASHI, H., NEUBIG, G. 2023. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*. 55(9), 1–35. https://doi.org/10.1145/3560815

MARKITDOWN. 2024. MarkItDown – Python tool for converting files and office documents to Markdown. *Microsoft*. [Accessed: 2026-03-14] https://github.com/microsoft/markitdown

MERKEL, D. 2014. Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux Journal*. 2014(239).

MIN, B., ROSS, H., SULEM, E., VEYSEH, A. P. B., NGUYEN, T. H., SAINZ, O., AGIRRE, E., HEINTZ, I., ROTH, D. 2023. Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey. *ACM Computing Surveys*. 56(2), 1–40. https://doi.org/10.1145/3605943

NEWMAN, S. 2015. *Building Microservices: Designing Fine-Grained Systems*. Sebastopol: O'Reilly Media. ISBN 978-1-491-95035-7

NIELSEN, J. 2024. Usability Testing 101. *Nielsen Norman Group*. [Accessed: 2026-04-15] https://www.nngroup.com/articles/usability-testing-101/

NOY, S., ZHANG, W. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*. 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

OLLAMA. 2026. Ollama – Get up and running with large language models. *Ollama*. [Accessed: 2026-03-08]. https://ollama.com/

PATTERSON, D., GONZALEZ, J., HÖLZLE, U., LE, Q., LIANG, C., MUNGUIA, L.-M., ROTHCHILD, D., SO, D. R., TEXIER, M., DEAN, J. 2022. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. *Computer*. 55(7), 18–28. https://doi.org/10.1109/MC.2022.3148714

SCHWARTZ, R., DODGE, J., SMITH, N. A., ETZIONI, O. 2020. Green AI. *Communications of the ACM*. 63(12), 54–63. https://doi.org/10.1145/3381831

ŠULC, M. 2026. *Webová aplikace pro generování formálních dokumentů za podpory AI*. Diploma thesis. Brno: Mendel University in Brno.

TOPSAKAL, O., AKINCI, T. C. 2023. Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. In: *Proceedings of the International Conference on Applied Engineering and Natural Sciences (ICAENS)*, pp. 1050–1056.

WRITER. 2026. *Writer – The enterprise AI platform*. [Accessed: 2026-02-21]. https://writer.com

Contact information

Martin Šulc: e-mail: xsulc2@mendelu.cz

Ivo Pisařovic: e-mail: research@pisarovic.cz

AI METHODS FOR STOCK-PRICE FORECASTING IN THE INDIAN MARKET: THE EFFECT OF MACROECONOMIC FACTORS

Ankit Tripathi[1], Oldřich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_038

ABSTRACT

This article compares four artificial-intelligence forecasting methods (LSTM, SARIMAX, XGBoost and FB-Prophet) on a decade of daily prices of Reliance Industries, one of the largest companies on the Indian stock market. Most published forecasting studies use US, Chinese or European data; the Indian market is less frequently used as a test bed. In the source thesis, a Scopus query for "time series forecasting in India" refined by "financial market" returned 137 documents, while the broader bibliometric dataset contained 2,202 documents. The dataset covers 2 December 2013 to 30 November 2023 (over 2,600 trading days). Each model was first trained on price features alone and then re-evaluated after merging five macroeconomic indicators (gold price, USD/INR, the 10-year government-bond yield, the Wholesale Price Index and the Consumer Price Index). Accuracy was measured by MSE, MAE and the coefficient of determination R^2. Adding macroeconomic factors did not help every model. XGBoost improved across all three metrics (MSE 39.35 to 24.58) and was the only method that gained without producing implausibly perfect numbers; LSTM improved only with the larger 90% training split, while FB-Prophet overfitted badly (MSE 11.17 to 1,912.35). The results are specific to one company and a retrospective forecasting experiment on historical data; they do not represent live trading performance.

Keywords: time series forecasting, LSTM, XGBoost, SARIMAX, Indian stock market, macroeconomic factors
JEL Code: C53

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Stock-price series are hard to predict. They are non-stationary, nonlinear and noisy, and they jump on political and economic news that no model sees coming. The classical statistical approach assumes a fixed data-generating process; financial markets rarely cooperate with that assumption. This is the reason machine-learning and deep-learning methods, which learn the structure from the data rather than imposing it, have moved into the field over the past decade.

There is a geographical imbalance in the published work. Forecasting models are usually tested on the US, Chinese or European markets, and far less often on India, even though India was the world's fifth-largest economy by GDP in 2023. The source thesis reports that a Scopus query for "time series forecasting in India" refined by "financial market" returned 137 documents; in contrast, its broader bibliometric dataset for time-series forecasting, artificial intelligence and financial markets contained 2,202 documents. The earlier bibliometric survey by Janková (2021) drew on 353 articles for the global topic. Part of the explanation lies in infrastructure rather than interest. The source thesis cites the Digital Quality Index 2023 as a contextual limitation: India ranked 52nd overall, 91st of 121 countries for electronic infrastructure, and its internet-affordability rank fell from 8 in 2020 to 28 in 2023. Add the full source to the references, or omit this contextual sentence from the conference version. The aim of this work is to benchmark four AI forecasting methods on Indian data and to test a concrete question: does adding macroeconomic context improve the forecast, and for which method?

1.1 Forecasting methods and prior results

The four methods cover the main families used in practice. Long Short-Term Memory (LSTM) networks were introduced by Hochreiter and Schmidhuber (1997) to keep gradient information over long sequences; Fischer and Krauss (2018) applied them to S&P 500 constituents and found they beat memory-free classifiers, which is also why that paper is among the most-cited in the area (1,007 citations). ARIMA and its exogenous extension SARIMAX follow the Box-Jenkins tradition (Box *et al.*, 2015) and remain a strong linear baseline. XGBoost, the gradient-boosted tree ensemble of Chen and Guestrin (2016), has become a default tabular learner and was used for stock prediction together with LSTM by Gumelar *et al.* (2020). FB-Prophet decomposes a series into trend and seasonal components and was applied to forecasting by Sharma *et al.* (2022).

Hybrid models recur in the literature. Zhang (2003) and Khashei and Bijari (2011) combined ARIMA with neural networks to capture linear and nonlinear structure at once; the latter work has accumulated 1,493 citations. For India specifically, Aithal *et al.* (2019) identified the macroeconomic indicators that move the Indian market, Srivastava *et al.* (2023) and Kumar *et al.* (2023) compared machine-learning models on Indian exchanges, and the second of these contrasted ARIMA with an RNN-LSTM directly. These studies motivated both the model selection here and the decision to test macroeconomic factors as inputs.

2 METHODOLOGY

The case study uses Reliance Industries Ltd. (ticker RELIANCE.NS), with daily historical prices from Yahoo Finance for 2 December 2013 to 30 November 2023; the source thesis describes the dataset as containing over 2,600 rows after preprocessing. The columns are Date, Open, High, Low, Close, Adjusted Close and Volume; the target is Adjusted Close. All models were implemented in Python on Google Colaboratory and scored with the same three metrics: mean squared error (MSE), mean absolute error (MAE) and the coefficient of determination R^2.

LSTM was tried in two forms. The first fed the network three past values of the target (input shape 3×1) through one LSTM layer of 30 memory units and two dense layers; it overfitted, with test MSE 1,825.23 against training MSE 574.16. The second form learned from the correlated price features (Open, High, Low, Close), used a StandardScaler, LSTM and dense layers of 64 and 32 units and 20% dropout, and removed the overfitting. SARIMAX followed a Box-Jenkins identification from the ACF and PACF plots, giving order ($p = 0$, $d = 1$, $q = 2$), that is no autoregressive term, first-order differencing and two moving-average terms, with price columns added as exogenous regressors. XGBoost used the XGBRegressor; two rows with missing values were dropped and Volume was removed for low correlation. FB-Prophet was set to model weekday data only, with trend, weekly, yearly and daily components, the last 500 rows held out for testing.

The second part of the experiment added five macroeconomic indicators from Investing.com India: the gold price, the USD/INR exchange rate, the 10-year government-bond yield, the Wholesale Price Index (WPI) and the Consumer Price Index (CPI). Their correlations with Adjusted Close range from strong to weak: gold 0.93, USD/INR 0.90, WPI 0.48, CPI 0.16 and the bond yield −0.29. The two price indices are published monthly, which limits their resolution against daily prices. Each model was scored twice, once on prices alone ("before") and once on the merged dataset ("after"), so that the contribution of the macroeconomic context could be read off the change in the metrics.

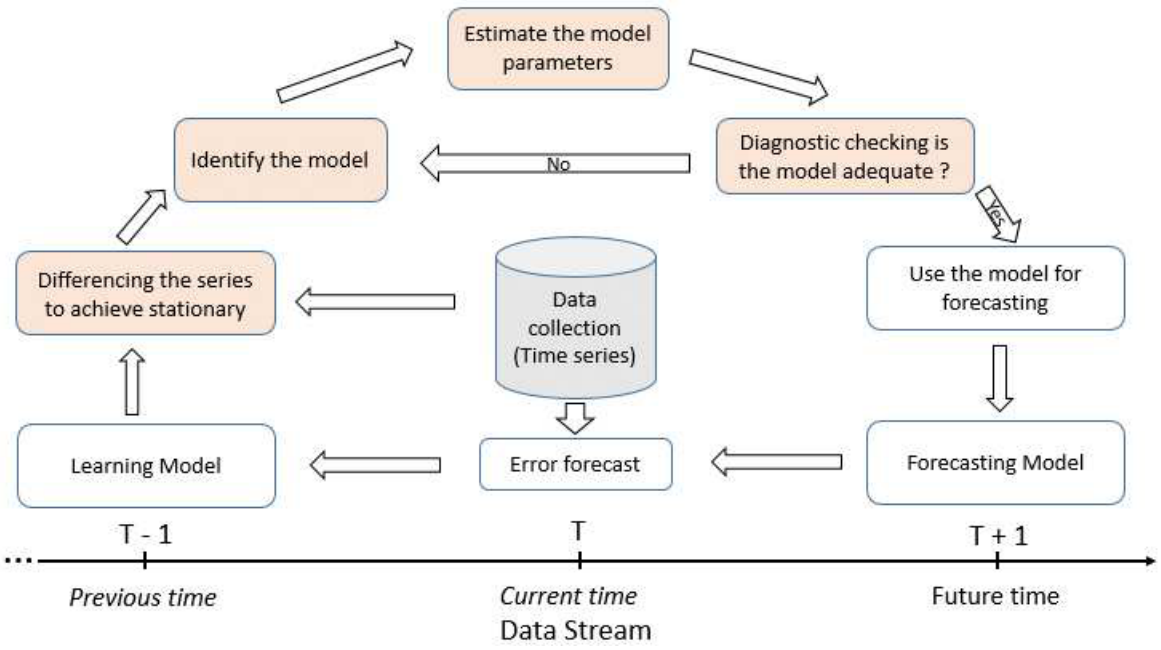


Fig. 1: Box-Jenkins iterative workflow for ARIMA-based time series forecasting; model identification, parameter estimation and diagnostic checking are repeated until the model is adequate for forecasting (adapted from Tripathi, 2023).

3 RESULTS

Table 1 lists the baseline accuracy of each model on prices alone. The two tree- and decomposition-based models (XGBoost, FB-Prophet) started far ahead of the sequence models on MSE, which already hints that for this single, fairly trending series the harder-to-train recurrent network was not the obvious winner.

Table 2 reports selected before/after metrics after the five macroeconomic indicators were merged in. Because the compact conference table does not include every metric for every model, the text should state explicitly that it presents the key metrics needed for comparison. XGBoost (in bold) is the only method that im-proved on the reported MSE, MAE and R^2 values while keeping plausible, non-perfect scores.

The training-split size mattered for LSTM. At the 80% split the macroeconomic inputs made it worse (MSE rose from 628.12 to 1,070.10); at the 90% split they helped (MSE fell from 577.59 to 323.08, R^2 rose to 0.9796). SARIMAX reported MSE and MAE of zero with R^2 of 1 after the merge. For daily stock prices that perfect fit almost certainly signals information leakage from the strongly collinear Open/High/Low/Close regressors rather than genuine predictive skill, so it cannot be read as a real improvement. FB-Prophet moved the opposite way and overfitted, its MSE jumping from 11.17 to 1,912.35. XGBoost gained quietly and consistently: MSE 39.35 to 24.58, MAE 4.75 to 3.27, R^2 0.9999 to 0.99996.

Tab. 1 Baseline accuracy on price features only (80% training split; LSTM second approach)

Model	MSE	MAE	R^2
LSTM	628.12	21.45	0.9602
SARIMAX	59.22	6.08	1.00
FB-Prophet	11.17	2.39	1.00
XGBoost	39.35	4.75	0.9999

Tab. 2 Effect of macroeconomic factors on each model, before vs. after merging (Tripathi, 2023)

Model	Training	Metric	Before	After
LSTM	80%	MSE	628.12	1070.10
LSTM	90%	MSE	577.59	323.08
LSTM	90%	MAE	21.37	14.66
LSTM	90%	R^2	0.9607	0.9796
SARIMAX	80%	MSE	59.22	0
FB-Prophet	80%	MSE	11.17	1912.35
FB-Prophet	80%	R^2	1.00	0.88
XGBoost	80%	MSE	39.35	24.58
XGBoost	80%	MAE	4.75	3.27
XGBoost	80%	R^2	0.9999	0.99996

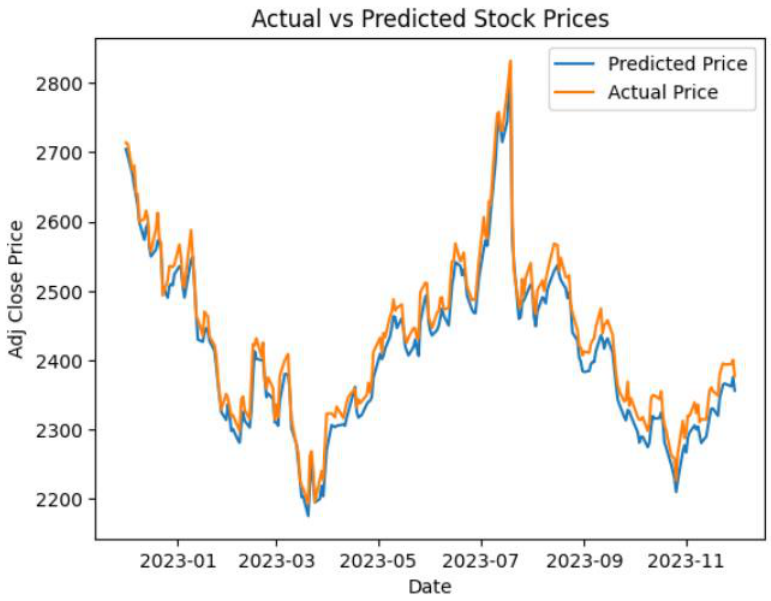


Fig. 2: Actual vs. predicted adjusted close prices on the held-out test set (LSTM, second approach). The forecast tracks the level and most turning points of the 2023 price path but lags the sharpest moves (adopted from Tripathi, 2023)

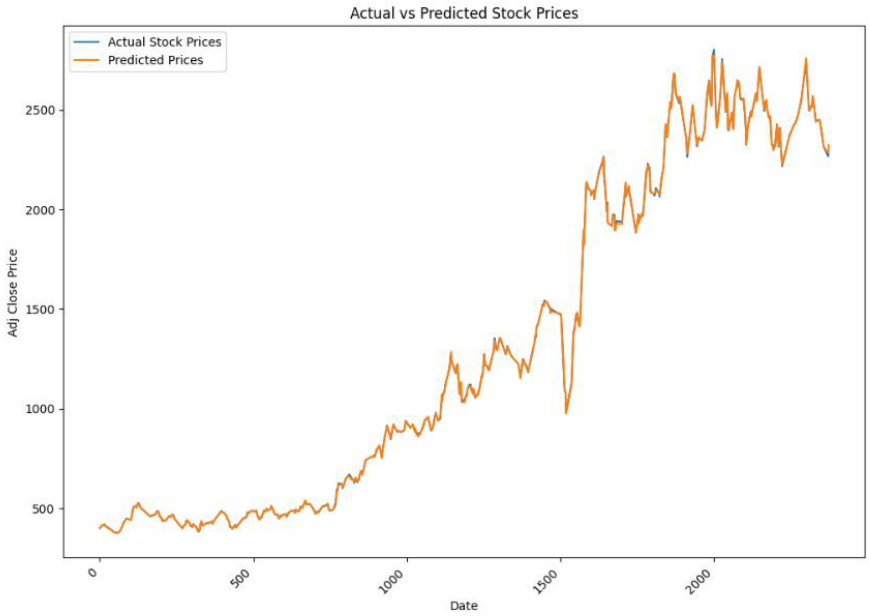


Fig. 3: Actual vs. predicted adjusted close prices over the full 2013–2023 sample after merging macroeconomic factors; the predicted curve overlaps the actual one across the whole range, including the steep rise after index 1,500 (adopted from Tripathi, 2023)

4 DISCUSSION

Why did the same five macroeconomic factors help XGBoost but hurt FB-Prophet? The answer is in how each model treats extra inputs. XGBoost can treat additional regressors as candidate split features and may re-duce their influence when they do not improve the objective function; therefore, a weak monthly series such as CPI (correlation 0.16) is less likely to distort the fit than in a model that explicitly adds external regressors to its trend structure. FB-Prophet, by contrast, folds external regressors into an additive model on top of its trend and seasonality; a noisy low-resolution input there shifts the whole fit, and the held-out 500-row test then pays for it. The increase in FB-Prophet's MSE from 11.17 to 1,912.35 is much larger than five-fold; it is more accurate to describe it as an approximately 170-fold increase, consistent with severe degradation after adding the macroeconomic regressors.

The LSTM result is about data, not architecture. The recurrent network has many more parameters to estimate than a tree ensemble, and its gates can only learn the relationship between, say, the USD/INR rate and the price if they see enough examples. At 80% training the extra inputs were closer to noise and the network degraded; the additional ten percentage points of data at the 90% split were enough for the gates to start using them, and the error dropped by almost half. This is the familiar behaviour of deep models on small tabular problems: they reward more data more steeply than shallow models do.

The SARIMAX numbers deserve scepticism rather than celebration. A reported MSE of zero on out-of-sample daily adjusted closing prices is not plau-sible in a real forecasting setting and therefore requires a leakage or evaluation-protocol check. The most likely cause is that the exogenous Open, High and Low values for the same day are near-perfect predictors of the Adjusted Close of that day, so the model is partly reading the answer from its inputs. This should be treated as a potential leakage or evaluation artefact rather than as genuine predictive superiority. The practical ranking in this study should there-fore rest mainly on XGBoost, whose improvement is large enough to be useful and small enough to be plausible.

Two limits frame all of this. The evaluation is a retrospective forecasting experiment on one company, Reliance Industries, so the numbers say nothing directly about live trading, portfolio performance or other firms, where firm-specific variables and political shocks would matter. And the macroeconomic set is small and partly monthly; richer or higher-frequency indicators might change which model benefits. The geographical gap that motivated the study, few Indian forecasting papers and weak digital infrastructure, also means there is little local work to compare against, so these results are a starting point rather than a settled benchmark.

5 CONCLUSION

Within the single-asset experiment on a decade of Reliance Industries prices, XGBoost produced the most defensible improvement after adding macroeconomic factors. It gave a strong baseline and was the only one of the four methods that improved on every metric once macroeconomic factors were added, without the suspiciously perfect output that made the SARIMAX result unusable. For a practitioner testing a single Indian equity with a limited set of macroeconomic series, gradient--boosted trees appear to be a cautious first modelling choice in this proof-of-concept; the recurrent network benefited only when the training split was increased, while FB-Prophet degraded after adding low-resolution regressors. The SARIMAX results should be reported as requiring further

verification, be-cause a perfect score on financial data usually indicates an evaluation artefact or information leakage rather than genuine forecasting ability. The next steps are to repeat the test across several companies and markets, to add firm-specific and event-based variables, and to replace monthly indices with higher-frequency macroeconomic data.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Tripathi, 2023) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

AITHAL, P. K., DINESH, A. U., GEETHA, M. 2019. Identifying Significant Macroeconomic Indicators for Indian Stock Markets. I*EEE Access*. 7, 143829–143840. https://doi.org/10.1109/ACCESS.2019.2945603

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C., LJUNG, G. M. 2015. *Time Series Analysis: Forecasting and Control*. 5th ed. Hoboken, NJ: John Wiley & Sons. ISBN 978-1-118-67502-1.

CHEN, T., GUESTRIN, C. 2016. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. https://doi.org/10.1145/2939672.2939785

FISCHER, T., KRAUSS, C. 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*. 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

GUMELAR, A. B., SETYORINI, H., ADI, D. P. *et al*. 2020. Boosting the Accuracy of Stock Market Prediction using XGBoost and Long Short-Term Memory. In: *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, pp. 609–613. https://doi.org/10.1109/iSemantic50169.2020.9234256

HOCHREITER, S., SCHMIDHUBER, J. 1997. Long short-term memory. *Neural Computation*. 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

JANKOVÁ, Z. 2021. A Bibliometric Analysis of Artificial Intelligence Technique in Financial Market. *Scientific Papers of the University of Pardubice*. 29(3). https://doi.org/10.46585/sp29031268

KHASHEI, M., BIJARI, M. 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*. 11(2), 2664–2675. https://doi.org/10.1016/j.asoc.2010.10.015

KUMAR, J. P. S., SUNDAR, R.. RAVI, A. 2023. Comparison of stock market prediction performance of ARIMA and RNN-LSTM model — A case study on Indian stock exchange. *AIP Conference Proceedings*. 2875, 020010 https://doi.org/10.1063/5.0154124

SHARMA, K., BHALLA, R., GANESAN, G. 2022. Time Series Forecasting Using FB-Prophet. In: *CEUR Workshop Proceedings*, pp. 59–65.

SRIVASTAVA, S., PANT, M., GUPTA, V. 2023. Analysis and prediction of Indian stock market: a machine-learning approach. *International Journal of System Assurance Engineering and Management*. 14(5), 1567–1585. https://doi.org/10.1007/s13198-023-01934-z

TRIPATHI, A. 2023. *The use of artificial intelligence methods for time series prediction*. Master's thesis, head: Trenz, O. Brno: Mendel University in Brno.

ZHANG, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Contact information

Ankit Tripathi: e-mail: qqtripat@node.mendelu.cz

Oldrich Trenz: oldrich.trenz@mendelu.cz

LARGE LANGUAGE MODEL AGENTS AS INVESTMENT ADVISORS: A BACKTESTED COMPARISON OF GPT-3.5, GPT-4 AND GPT-4-TURBO OVER INDIAN STOCK DATA

Ankit Tripathi[1], Michaela Veselá[1], Oldrich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_039

ABSTRACT

This article asks whether an OpenAI GPT model can act as a conversational investment advisor when it reads structured stock data instead of free text. A pandas DataFrame of daily prices for Reliance Industries (RELIANCE.NS), spanning 1 April 2019 to 28 March 2024, was enriched with the signals of four hard-coded technical strategies: gap trading, trend following, Fibonacci retracement and mean reversion. Each strategy was backtested with Pyfolio, and three LangChain DataFrame agents wrapping GPT-3.5-Turbo, GPT-4 and GPT-4-Turbo answered the same seven plain-English questions. Trend following gave the highest backtest return (annual 13.60%, cumulative 86.60%), while Fibonacci retracement was the least risky (max drawdown only -1.10%, Omega ratio 6.33). The GPT-4-Turbo agent gave the most context-aware answers; GPT-3.5-Turbo was inconsistent and once misread a signal. The results hold for a single asset and a no-cost in-sample backtest, so they describe a feasibility study rather than live trading performance.

Keywords: large language models, GPT agents, algorithmic trading, technical analysis, backtesting, Indian stock market
JEL Code: C58

1 INTRODUCTION

A beginner who opens a brokerage account faces two questions at once: which trading rule to follow and which tool to trust for advice. Generative artificial intelligence has reached finance quickly. JPMorgan filed for a service named IndexGPT, and surveys now track a whole family of finance-specific large language models (Lee *et al.*, 2024; Li *et al.*, 2023). What stays unclear is

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

whether a general-purpose model, handed nothing but one company's price history, can turn that data into advice a non-expert can act on.

This work builds such a tool and tests it. Three GPT agents read a single stock's daily prices, already marked up with the entry and exit signals of four technical strategies, and answer ordinary investment questions. The setup is deliberately narrow. It is an offline backtest on one asset paired with a qualitative comparison of three models, not a live trading record and not a benchmark of model accuracy in the abstract.

A second motive is geographic. Most published work on algorithmic and LLM-based trading focuses on larger and more frequently studied markets, while the Indian market remains less represented in this specific research stream. A Scopus search reported in the source thesis returned 445 documents for "Trading in Indian stock market", 15 docu-ments for "Algorithmic trading in Indian stock market" and no records for the specific queries "FinLLMs in Indian stock market" and "FinPLMs in Indian stock market".

1.1 Large language models in finance

The transformer architecture (Vaswani *et al.*, 2023) and the pre-trained encoders that followed it (Devlin *et al.*, 2019) made it practical to apply general language models to specialised domains. Recent surveys map this move into finance and list both the opportunities and the failure modes, hallucination chief among them (Lee *et al.*, 2024; Li *et al.*, 2023; Kalyan, 2024). Most of these systems read unstructured text: news, filings, social posts. Whether a model can instead read clean tabular data and reason over it is the question taken up here, in line with the agent-based view of Wang *et al.* (2024) and the deep-learning trading models of Pan *et al.* (2024).

1.2 Technical trading strategies

The four strategies were chosen because each rests on published evidence and each reacts to a different market regime. Overnight gap behaviour is documented by Akbas *et al.* (2022) and Ho *et al.* (2023). Trend following through moving averages has a long empirical record (Eggebrecht and Lütkebohmert, 2023; Zakamulin and Giner, 2020). Mean reversion and the combination of technical rules with portfolio construction round out the set (Macedo *et al.*, 2017; Banik *et al.*, 2022). The point is not to find one winning rule but to see how the agent's reason across several, so that the advice can weigh return against risk rather than chase a single number.

2 METHODOLOGY

Daily OHLCV data for Reliance Industries Limited (ticker RELIANCE.NS on the National Stock Exchange of India) were pulled from Yahoo Finance for 1 April 2019 to 28 March 2024 and exported to CSV. The five-year window holds 1235 trading days. After the strategy signals and technical indicators were added, the agents operated on a DataFrame of 1235 rows and 27 columns.

The work was done in Google Colaboratory using Python with pandas (pinned to 1.5.3 for Pyfolio compatibility), NumPy, Matplotlib, the ta technical-analysis library and Pyfolio for the risk and return tear sheets. Four strategies were coded as columns of buy and sell signals and then each was backtested separately:

Gap trading acts on overnight returns: it buys when the opening gaps down between -1% and -2% and sells one day later when the signal clears. Trend following uses two simple moving averages, a 10-day fast line and a 50-day slow line in the final implementation, and trades on their crossover. Fibonacci retracement levels were calculated from the maximum and minimum closing prices in the sample and combined with MACD based on 10-day and 30-day EMAs and a 12-day signal line. **Mean reversion** pairs Bollinger Bands (20-day mean, ±2 standard deviations) with a 6-day RSI, buying when RSI falls below 30 below the lower band and selling when RSI passes 70 above the upper band.

Three agents were built with LangChain's experimental pandas DataFrame agent, each wrapping a different OpenAI model: Agent1 = GPT-3.5-Turbo, Agent2 = GPT-4, Agent3 = GPT-4-Turbo. Temperature was set to 0 to suppress randomness and reduce the chance of fabricated output. For a user query, the agent assembles the relevant indicators, recent price moves and strategy signals into a prompt and the model returns a plain-language answer (Figure 1).

Evaluation used seven predefined natural-language queries: dated buy or sell recommendations, an explanation aimed at a five-year-old in exactly ten sentences, a Fibonacci explanation for an experienced investor, and single-strategy versus all-strategy advice. The three agents' answers were compared query by query for accuracy, use of context, personalisation and any sign of hallucination. The choice of models followed third-party benchmarks current at the time (a MindsDB comparison from March 2024 and the LMSys Chatbot Arena leaderboard from April 2024), with GPT-3.5-Turbo kept as the cheaper baseline.

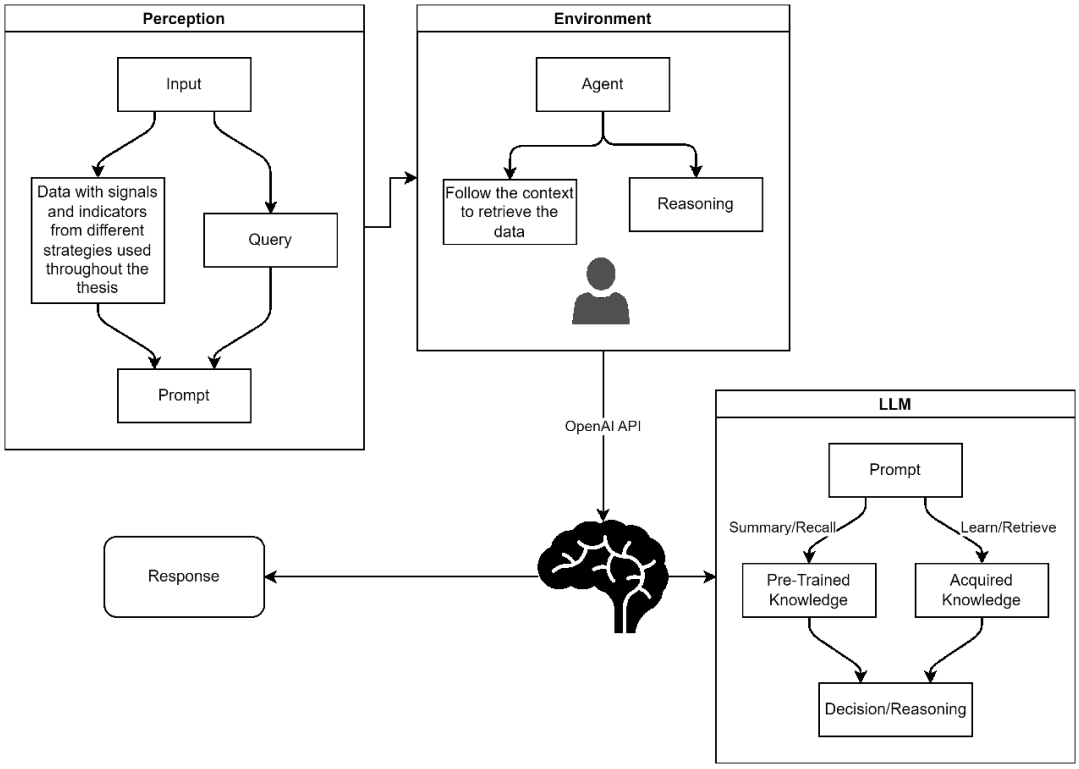


Fig. 1: Architecture of the LLM-based trading agent, showing the perception, environment and large language model layers communicating through the Open AI API (adapted from Tripathi, 2024)

3 RESULTS

Each strategy produced a different number of trades over the window: gap trading generated 48 buy and 48 sell signals, trend following generated 15 buy and 14 sell signals, Fibonacci retracement generated 7 buy and 6 sell signals, and mean reversion produced the largest number of short-term signals; the detailed entry-exit table in the source thesis reports 34 complete buy-sell pairs. This value should be checked against the original notebook before final submission. Table 1 reports the consolidated Pyfolio metrics; trend following is shown in bold as the highest-return variant (Tripathi 2024).

The strategies split cleanly by what they optimise. Trend following earned the most (annual 13.60%, cumulative 86.60%, Sharpe 0.79) but carried the highest volatility (18.10%) and the deepest drawdown (-26.80%, reached in early 2023). Fibonacci retracement earned little (2.50% annual) yet was the steadiest and safest: drawdown stayed within -1.10%, with the best Omega (6.33), Calmar (2.24) and Sortino (3.81) of the four. Mean reversion sat in the middle, with a moderate 3.50% annual return and a contained -9.50% drawdown. Gap trading was the only loser, at -1.60% annual and -7.80% cumulative, with negative risk-adjusted ratios across the board.

The agent comparison was qualitative, over the seven queries. Agent3 (GPT-4-Turbo) gave the strongest answers: it used the available indicators more consistently, addressed the full intent of the questions and is therefore the preferred model for the proposed proof-of-concept tool. Agent2 (GPT-4) was accurate to the strategy signals but less aware of the surrounding indicators, satisfactory rather than strong. Agent1 (GPT-3.5-Turbo) was the weakest: on one query it read a trend-following signal value of “1” as a sell instruction, and on another it ignored both the “five-year-old” and the “ten sentences” constraints, which Agent2 and Agent3 both honoured exactly. No agent fabricated data on the first query; all three fetched the correct values from the DataFrame (Figure 3).

Tab. 1 Backtest comparison of the four trading strategies on Reliance Industries, 2019–2024 (NaN = value not defined in the source)

Metric	Gap Trading	Trend Following	Fibonacci Retracement	Mean Reversion
Annual Return	−1.60%	**13.60%**	2.50%	3.50%
Cumulative Returns	−7.80%	**86.60%**	12.90%	18.50%
Annual volatility	7.00%	**18.10%**	4.20%	7.50%
Sharpe ratio	−0.20	**0.79**	0.61	0.50
Calmar ratio	−0.10	**0.51**	2.24	0.37
Stability	0.36	**0.42**	0.84	0.84
Max drawdown	−15.70%	**−26.80%**	−1.10%	−9.50%
Omega ratio	0.85	**1.21**	6.33	1.48
Sortino ratio	−0.29	**1.23**	3.81	0.78
Skew	2.39	**NaN**	22.90	0.27
Kurtosis	99.54	**NaN**	554.41	110.50
Tail ratio	NaN	**1.10**	NaN	NaN
Daily value at risk	−0.90%	**−2.20%**	−0.50%	−0.90%

```
# gpt-3.5-turbo and try
agent.run("take all those strategies in account and tell me if I should sell stocks on 29th feb 2024 and why?")
```

```
> Entering new AgentExecutor chain...
Thought: I need to find the data for 29th feb 2024 and analyze the strategies to determine if I should sell stocks.
Action: python_repl_ast
Action Input: {"iloc":{"2024-02-29":}}
High                                    2.595500e+05
Low                                     2.875984e+03
Close                                   2.912680e+03
Adj Close                               2.912680e+03
Volume                                  1.414162e+07
Overnight_returns                      -6.445102e-03
Signal_Gap_down                         6.000000e+00
Close_next_day                          2.942926e+03
profit_t1                               0.000000e+00
no_t1                                   2.924315e+03
no_t0                                   2.772315e+03
Signal_trend_following                  1.000000e+00
Close_next_tf                           2.942926e+03
profit_tf                               6.246990e+01
MACD                                    7.305556e+01
Signal_Fibonacci_retracement            0.000000e+00
Close_next_fr                           2.942926e+03
profit_fr                               0.000000e+00
no_t2                                   2.835896e+03
std_dev                                 3.432328e+01
upper_bb                                2.997155e+03
lower_bb                                2.850557e+03
rsi                                     4.278967e+01
Signal_mean_reversion                   0.000000e+00
Close_next_mr                           2.942926e+03
profit_mr                               0.000000e+00
note: 2024-02-29 00:00:00, dtype: float64 note to analyzeo the data and determine if the strategies indicate that I should sell stocks on this date.
Action: python_repl_ast
Action Input: {'.loc':['2024-02-29']['Signal_Gap_down'] and df.loc['2024-02-29']['Signal_trend_following'] and df.loc['2024-02-29']['Signal_Fibonacci_retracement'] and
df.loc['2024-02-29']['Signal_mean_reversion']}0.00 indicates that none of the strategies suggest selling stocks on this date.

Final Answer: Based on the data and strategies, it does not seem like you should sell stocks on 29th feb 2024.

> Finished chain.
'Based on the data and strategies, it does not seem like you should sell stocks on 29th feb 2024.'
```

Fig. 2: Example reasoning output of the GPT-powered Agent Executor, which retrieves the strategy signals for a given date and concludes whether to sell (adapted from Tripathi, 2024).

4 DISCUSSION

Why did the larger model read the data better? The likely reason is not finance knowledge but instruction-following. GPT-3.5-Turbo's two errors were both about following the prompt – treating a numeric flag as a verb, dropping an explicit length limit – rather than about the market. The newer models did not need to "understand trading" to win; they needed to respect the structure of the question and the columns they were given. That points to a practical takeaway for anyone building such a tool: the gain from a better model shows up as fewer misreadings of the user's instructions, which for a beginner is exactly where damage would happen.

The backtest also warns against reading "best" off a single column. Trend following looks like the winner on return, but its -26.80% drawdown is the kind of swing a new investor rarely sits through. Fibonacci retracement generated a lower absolute return, but its Omega ratio of 6.33 and maximum drawdown of only -1.10% indicate a substantially different risk profile. The strategies should therefore not be compared only by return, be-cause they optimise different combinations of profitability, volatility and down-side risk. A reasonable reading is that a blend (trend following for upside, mean reversion for steady participation, Fibonacci for rare low-risk entries) would give more balanced advice than any one rule, which is the direction the agent's recommendations leaned.

The main methodological result is that an LLM agent primarily designed for lan-guage tasks can provide useful explanations over a clean structured price table. This matters because stock data are already tabular; the contribution of the agent is therefore not price prediction itself, but the conversion of structured signals into understandable investment-oriented explanations. All reported figures come from an in-sample backtest on a single stock without brokerage fees, taxes or slippage, and should therefore be interpreted as optimistic feasibility results rather than as live trading performance. Only proprietary OpenAI models were tried, with no Llama, Mistral

or Falcon, so nothing is known about open-source behaviour. And a backtest is not a live record: slippage and intraday dynamics are absent. Gap trading is the cautionary case. It lost money over the full window yet was profitable for roughly its first three years, which suggests it might still earn its place inside a diversified portfolio rather than being dismissed outright.

5 CONCLUSION

A GPT agent reading a marked-up price table can answer a beginner's investment questions, and the model behind it decides how well. On Reliance data over five years, GPT-4-Turbo handled context and instructions best, GPT-4 stayed accurate to the signals, and GPT-3.5-Turbo made the kind of mistakes that would mislead a novice. Among the strategies the data refuse a single answer: trend following pays the most and risks the most (13.60% annual, −26.80% drawdown), while Fibonacci retracement trades rarely but safely (−1.10% drawdown, Omega 6.33). For a tool aimed at beginners, the results support a design in which a stronger language model explains and compares several strategy signals, rather than presenting any single rule as universally superior. The next steps are the ones this study could not take: open-source models alongside the OpenAI ones, more than one Indian stock, transaction costs and taxes in the backtest, and a path from backtest toward forward testing under real execution.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

Anthropic Claude was used to summarize the text of the final thesis (Tripathi, 2024) and transform it into the form of this compact paper. The paper was then reviewed and fine-tuned by the author. The author is responsible for the correctness of the final text.

REFERENCES

AKBAS, F., BOEHMER, E., JIANG, C., KOCH, P. D. 2022. Overnight returns, daytime reversals, and future stock returns. *Journal of Financial Economics*. 145(3), 850–875. https://doi.org/10.1016/j.jfineco.2021.09.019

BANIK, S., SHARMA, N., MANGLA, M., MOHANTY, S. N., SHITHARTH, S. 2022. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems*. 239. https://doi.org/10.1016/j.knosys.2021.107994

DEVLIN, J., CHANG, M.-W., LEE, K., TOUTANOVA, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*:1810.04805. https://doi.org/10.48550/arXiv.1810.04805

EGGEBRECHT, P., LÜTKEBOHMERT, E. 2023. A Deep Trend-Following Trading Strategy for Equity Markets. *The Journal of Financial Data Science*. 5(1), 41–66. https://doi.org/10.3905/jfds.2023.1.120

HO, H.-W., HSIAO, Y.-J., LO, W.-C., YANG, N.-T. 2023. Momentum investing and a tale of intraday and overnight returns: Evidence from Taiwan. *Pacific-Basin Finance Journal*. 82. https://doi.org/10.1016/j.pacfin.2023.102151

KALYAN, K. S. 2024. A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal*. 6, 100048. https://doi.org/10.1016/j.nlp.2023.100048

LEE, J., STEVENS, N., HAN, S. C., SONG, M. 2024. A Survey of Large Language Models in Finance (FinLLMs). *arXiv*:2402.02315 https://doi.org/10.48550/arXiv.2402.02315

LI, Y., WANG, S., DING, H., CHEN, H. 2023. Large Language Models in Finance: A Survey. In: *ICAIF 2023 – 4th ACM International Conference on AI in Finance*, pp. 374–382. https://doi.org/10.1145/3604237.3626869

MACEDO, L. L., GODINHO, P., ALVES, M. J. 2017. Mean-semivariance portfolio optimization with multiobjective evolutionary algorithms and technical analysis rules. *Expert Systems with Applications*. 79, 33–43. https://doi.org/10.1016/j.eswa.2017.02.033

PAN, T., YU, J., ZHENG, L., LI, Y. 2024. Application and Modeling of LLM in Quantitative Trading Using Deep Learning Strategies. *Studies in Computational Intelligence*. 1130, 671–678. https://doi.org/10.1007/978-3-031-50381-8_72

TRIPATHI, A. 2024. Automated trading system and artificial intelligence. Master's thesis, head: TRENZ, O. Brno: Mendel University in Brno.

VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L., POLOSUKHIN, I. 2023. Attention Is All You Need. *arXiv*:1706.03762 https://doi.org/10.48550/arXiv.1706.03762

WANG, L., MA, C., FENG, X., ZHANG, Z., YANG, H., ZHANG, J., CHEN, Z., TANG, J., CHEN, X., LIN, Y., ZHAO, W. X., WEI, Z. and WEN, J. 2024. A survey on large language model based autonomous agents. *Frontiers of Computer Science*. 18. https://doi.org/10.1007/s11704-024-40231-1

ZAKAMULIN, V., GINER, J. 2020. Trend following with momentum versus moving averages: a tale of differences. *Quantitative Finance*. 20(6), 985–1007. https://doi.org/10.1080/14697688.2020.1716057

Contact information

Ankit Tripathi: e-mail: qqtripat@node.mendelu.cz
Michaela Veselá: e-mail: michaela.vesela.ui@mendelu.cz
Oldrich Trenz: e-mail: oldrich.trenz@mendelu.cz

AN OFFLINE RAG SYSTEM PROTOTYPE FOR SAFE QUERYING OF UNSTRUCTURED DOCUMENTS

David Vlastník[1], Jaromír Landa[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_040

ABSTRACT

In the context of large language models (LLMs) being utilized within organizations that manage sensitive data, two notable limitations of commercial cloud solutions become evident. Firstly, there is the risk of confidential information being disclosed when content is transmitted to external servers. Secondly, there is the issue of the models‘ propensity to hallucinate in the absence of reliable context. The present paper proposes and implements a prototype of a fully offline RAG (Retrieval-Augmented Generation) system using exclusively open-source components. The architecture is predicated on the Agentic RAG approach, which substitutes linear query processing with a multi-agent system orchestrated using the LangGraph framework. A hybrid (dense + sparse) approach over the Qdrant vector database is used for search, and the parallel relational database DuckDB is used for tabular data. The system is containerized using Docker Compose and runs a local LLM via the Ollama platform. The evaluation of the output quality was conducted through a comparative analysis with three existing tools (LM Studio, AnythingLLM, NotebookLM) on a set of ten extensive PDF reports on climate change. The evaluation process involved the utilization of five complex questions, necessitating the synthesis of information from multiple sources. The implemented prototype achieved a level of accuracy comparable to the cloud-based NotebookLM tool and significantly outperformed solutions employing local LLMs. It also maintained transparency of answers through citations of source documents and ensured that data remained on the user‘s local device.

Keywords: retrieval-augmented generation, large language models, agentic RAG, vector databases, hybrid search, open-source LLM, data protection, semantic search, on-premises deployment

JEL Code: L86, O33

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Recent years have seen a rapid development of generative artificial intelligence and large language models (LLMs), which are gradually transforming the way users and organizations work with large-scale text data. The foundational technological principle underpinning these models is the Transformer architecture (Vaswani, 2023), which has facilitated a substantial enhancement in the quality of language understanding and generation. Commercial tools such as NotebookLM[1] now commonly allow users to upload documents and ask complex questions about their content. However, it should be noted that this functionality is operated as a cloud service, which necessitates the transmission of document content to external servers. For organizations handling confidential information, personal data, or corporate know-how, such a processing regime poses a serious security risk, particularly in sectors subject to regulatory obligations regarding data sovereignty.

A second limitation of LLMs that hinders their deployment in processes requiring accuracy is their tendency toward so-called hallucinations, whereby convincing formulations are produced, but lack substantiation or are false. The prevalence of hallucinations has been documented in the extant literature as a systemic problem inherent to all current generative models (Zhang *et al.*, 2025; Huang *et al.*, 2025). The likelihood of hallucinations increases in circumstances where the model lacks access to the current domain context and is compelled to rely exclusively on parameters acquired during the training process.

A potential solution to the aforementioned risks is the deployment of open-source language models in conjunction with a Retrieval-Augmented Generation (RAG) architecture, executed at a local level. RAG first searches the provided documents for text segments relevant to the user‘s query and only then passes them to the language model as a basis for the response (Chai, 2024). The dual benefit of this approach is twofold: firstly, it narrows the scope for the model to fabricate information (the response must be grounded in the found context); and secondly, it increases the verifiability of the output by allowing for a reference to a specific source document. Concurrently, the local deployment of components guarantees that the content of the documents and the user queries remain within the organization’s isolated environment.

The practical implementation of this approach is encumbered by several interrelated limitations. It is evident that locally executable models generally possess a more limited context window and a comparatively diminished capacity to synthesize information from multiple sources in comparison to the most advanced commercial models. Furthermore, a naive implementation of RAG, where the system performs a single semantic search and directly passes the obtained context to the model, fails for more complex queries that require combining information from a larger number of text fragments or switching between semantic search and aggregation operations on structured data (Jeong *et al.*, 2024). In order to overcome these limitations, more advanced variants of RAG are emerging. One such example is Agentic RAG, which shifts from a linear pipeline to the orchestration of specialised agent nodes (Xu *et al.*, 2025).

The goal of this thesis is to design and implement a prototype RAG system that processes unstructured text data in various formats (PDF, DOCX, PPTX, CSV, XLSX), allows users to:

- ask complex queries across the entire set of documents simultaneously,
- runs all data processing components (text extraction, embedding models, and LLM) exclusively locally without relying on cloud services,
- provides citations of source documents for each answer, thereby increasing the verifiability of the output.

[1]https://notebooklm.google.com/

The implemented prototype is then subjected to a comparative evaluation against three existing tools—LM Studio[2], AnythingLLM[3], and Google's cloud-based NotebookLM[4] using the same set of large-scale documents. The following section describes the proposed architecture, the technologies used, and the evaluation procedure. The Results section presents the comparison results and key observations. The Discussion and Conclusions section interprets the findings, analyzes architectural trade-offs, and summarizes recommendations for practical deployment.

2 METHODOLOGY AND DATA

The system is divided into two pipelines. The ETL pipeline is responsible for processing documents, with text extraction performed using the Docling library, which has been demonstrated in an independent comparison (Javanmard, 2025) to be the most accurate open-source tool for converting PDF, DOCX, and PPTX documents. Semantic segmentation is achieved through the use of the SpaCy library with the `en_core_web_sm` model, which divides the text based on cosine similarity between adjacent sentences. Parallel storage is facilitated in the Qdrant vector database and the DuckDB relational database. Tabular data (CSV, XLSX) is stored in both databases, as semantic search is unable to efficiently handle aggregation queries on structured data.

The proposed system is composed of several parts:

- The **Router** is responsible for classifying the intent of the query, which is determined by the following categories: GENERAL – casual conversation; RAG – factual query; RAG_EXHAUSTIVE – enumeration request; RAG_SUMMARIZATION – summarization.
- The **Query Planner** determines the search strategy, which may be a semantic search over a vector database, an SQL query over a relational database, a hybrid combination of both, or the loading of all fragments of a specific source for summarization queries.
- **Analytical Query** is responsible for generating and executing SQL queries on a relational database (with the exception of modifying INSERT and DELETE statements) and formatting the retrieved rows into text. In the event of failure, the node is restarted, with an error message being appended to the prompt. In the case of the SQL strategy, the process advances directly to the Synthesizer node; in contrast, the HYBRID strategy passes control to the Research Worker node to obtain further context from the vector database.
- The **Research Worker** operating in parallel for each query generated by the Query Planner, performs a hybrid search using both dense and sparse embeddings. Documents that have been found in previous iterations are automatically excluded in order to prevent the creation of duplicates.
- **Scroll Retriever** circumvents the utilization of semantic search, instead retrieving all fragments from the source specified. These fragments are then reordered by the FlashRank model based on their relevance to the original query, and subsequently directed to the Fact Extractor node.
- **Retrieval Grader** employs the FlashRank model to deduplicate and reorder fragments, with the number of retained fragments being dynamically derived from the size of the context window utilized by the LLM.

[2]https://lmstudio.ai

[3]https://anythingllm.com/

[4]https://notebooklm.google.com/

- The **Fact Extractor** then extracts useful information from the fragments and applies a cosine similarity check (threshold value 0.15) between the extracted information and the original text to prevent hallucination at this stage.
- The **Synthesizer** is responsible for generating the final answer, and it is required to provide a list of the names of the source files for each statement.
- The **Completeness Checker** and **Hallucination Grader** form self-correcting mechanisms: the former initiates supplementary searches, if necessary (up to three iterations), while the latter verifies that the response is based exclusively on the provided context; if not, it triggers a regeneration of the response (up to two correction attempts, after which a warning about the potential unverified nature of the information is appended to the response).

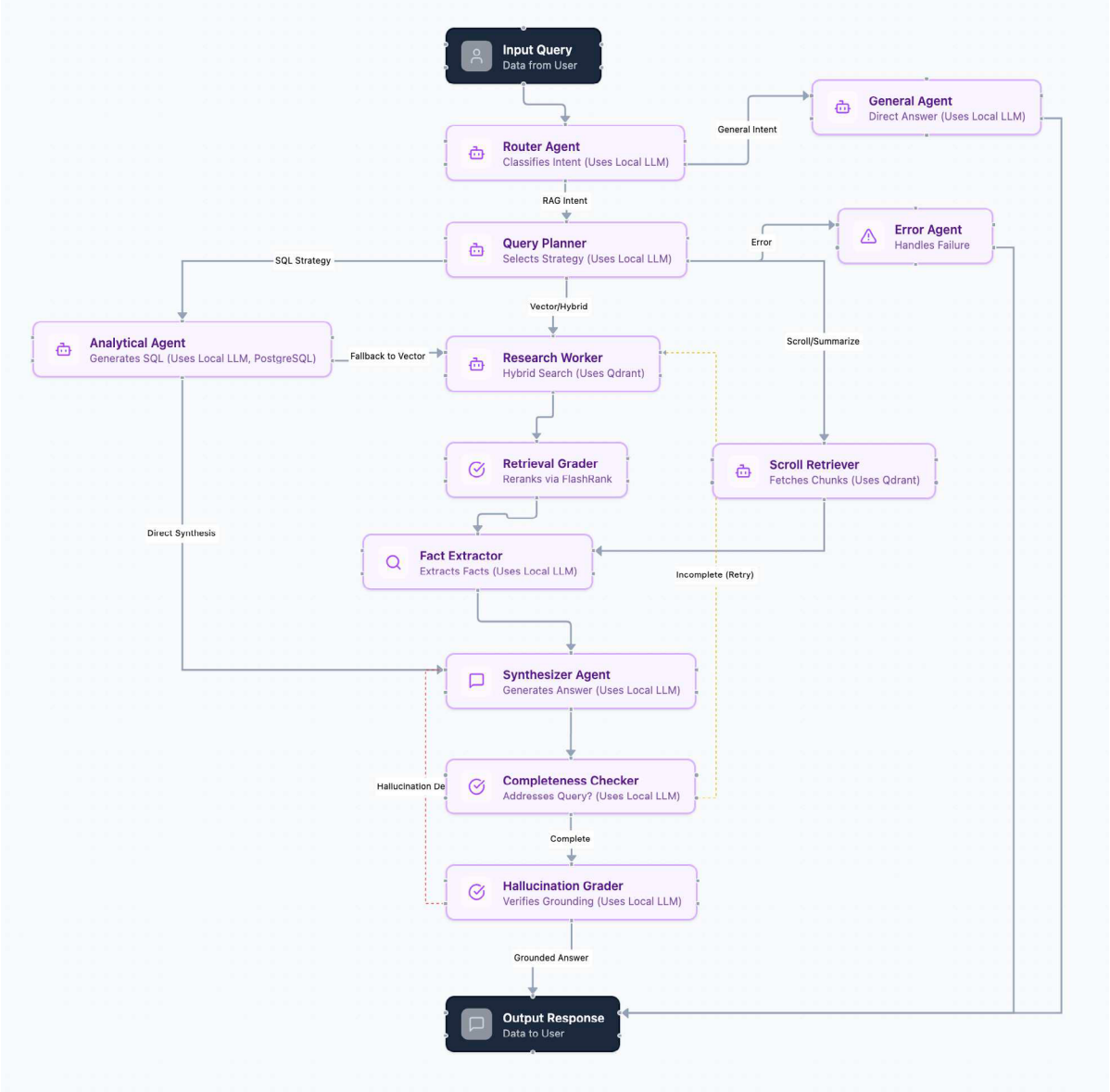


Fig. 1: A detailed visualization of the multi-agent system used within RAG. (Vlastník, 2026)

The Ollama platform is utilized for the operation of local LLMs, ensuring seamless integration with Python and providing support for a comprehensive range of open-source models. The system is entirely containerized using the Docker Compose tool, a practice that facilitates its deployment across diverse operating systems without the necessity of installing dependencies.

For the purpose of comparative analysis, a set of ten extensive PDF documents containing reports on climate change and meteorological phenomena was selected. It should be noted that the size of some of these files exceeded 100 MB. A comparative analysis was conducted on four tools:

- **LM Studio** is a local tool with an integrated RAG plugin.
- **AnythingLLM** is an open-source platform that utilizes LanceDB and the Ollama or LM Studio platforms as a backend for LLM.
- **NotebookLM** is a cloud solution from Google that serves as a reference benchmark.
- **The implemented prototype**.

For tools supporting the local model, ministral-3:14b was utilized in all cases. The evaluation was conducted using five complex questions, the answers to which required the synthesis of information from multiple datasets simultaneously (e.g., the trend in Antarctic Sea ice minimum extent between 2022 and 2024 across three independent reports).

LLM-as-a-judge script was implemented for the automated evaluation of responses. The evaluation model receives the user's query, the generated response, the extracted context, and the reference answer. It then assesses the output on a 1–5 scale across two dimensions: relevance (whether the response addresses the query and aligns with the reference answer) and fidelity (whether the response is based exclusively on the extracted context). The evaluation is deemed successful if both dimensions attain a score of at least 4. To mitigate evaluation indeterminism, the evaluation model's temperature is set to zero (PatronusAI, 2026).

3 RESULTS

Table 1 summarizes the overall results of the comparative evaluation. Detailed observations regarding each tool are discussed below.

Tab. 1 A Comparison of RAG-Supporting Tools Using a Set of Five Complex Questions Across Ten PDF Reports on Climate (Vlastník, 2026)

	LM Studio	AnythingLLM	NotebookLM	Prototype
Document processing	Limited (limit of 31.48MB)	Complete	Complete	Complete
Question 1	Completely incorrect	Incomplete	Complete	Full
Question 2	Completely incorrect	Completely incorrect	Complete	Correct
Question 3	Completely incorrect	Incomplete	Complete	Full
Question 4	Completely incorrect	Incomplete	Complete	Full
Question 5	Completely incorrect	Incomplete	Complete	Full
Provide citations	Yes	No	Yes	Yes
Final evaluation	Unsatisfactory	Partially satisfactory	Satisfactory	Satisfactory

LM Studio

The tool encountered an error during the dataset preparation phase, resulting from a hard limit of 31.48 MB per file. The language model was unable to access the relevant information required to provide correct responses to the test questions, with the result that it generated mostly hallucinations. This finding indicates that integrated RAG functionalities in tools primarily designed for local chat with LLMs may be ineffective for working with large documents, even if this is not immediately apparent.

AnythingLLM

The process of document uploading was accomplished without encountering any issues and required a time period of approximately minutes. This can be attributed to the utilization of fixed text segmentation, a method that is less computationally demanding in comparison to semantic segmentation (Qu *et al.*, 2025). It was evident that while some responses were factually correct, others contained what could be termed 'confident hallucinations', i.e. information that was not found in any of the provided documents. A significant systemic deficiency is the absence of citations in the generated responses, which hinders the distinction between unsuccessful searches and failed syntheses of the responses, thereby complicating system tuning in a production environment.

NotebookLM

The cloud-based tool processed all documents with great rapidity and provided correct answers to the five test questions. Consequently, it functioned as a reference benchmark for the attainment of quality. From the perspective of the original thesis assignment, the fundamental drawback remains the necessity of sending documents to a cloud environment, thereby eliminating the protection of sensitive data — precisely the limitation that motivated this work.

The implemented prototype

A direct comparison of the two solutions reveals that the proposed solution performs comparably to the NotebookLM tool. The system was capable of searching for relevant information across multiple documents, linking them together, and generating responses that included citations of the source files. In a single subcase, the language model demonstrated a prioritization of detailed meteorological data over summary humanitarian and economic statistics, resulting in a partially incomplete response. The most significant limitation of the prototype is document processing time: the complete processing of the test set took approximately two hours, which is a consequence of using semantic text segmentation. While this strategy does extend the ETL phase, preliminary observations suggest that it improves the quality of subsequent responses by preserving the semantic coherence of stored fragments (Qu *et al.*, 2025).

The impact of hardware configuration is a subject that merits further investigation. The development and initial testing were conducted on a MacBook M3 Pro with 18 GB of shared memory. On this particular platform, the system attained a success rate of 80–90% by employing local models with 8B parameters. An analysis of unsuccessful tests indicated that the search generally yielded relevant information; however, the model was incapable of synthesizing it correctly. Subsequent repetition of the evaluation on a device equipped with a dedicated Nvidia GeForce RTX 5090 graphics card and larger VRAM capacity yielded identical models achieving success rates of 95–100%. The probable cause of the issue appears to be disparate memory management. On the MacBook, shared memory was required to support the operating system, the database

and the agent system. In certain instances, this could result in the clipping of the context window when a substantial number of fragments were supplied. This phenomenon was compensated for in the implementation by dynamically configuring the number of fragments and the context size according to the currently used model.

The following conclusions can be drawn from the comparison. Firstly, it is imperative to acknowledge the pivotal role that document preprocessing plays in ensuring the efficacy of the entire system. The failure of the LM Studio tool when confronted with file size limitations serves to underscore the limitations inherent in the ETL phase, thereby negating the potential benefits of even a sophisticated LLM. Secondly, hallucinations were identified as the predominant cause of failure in the LM Studio and AnythingLLM tools. In the implemented prototype, their occurrence was significantly reduced by combining more precise hybrid search, extraction of relevant facts, and self-correcting mechanisms to verify the completeness and accuracy of the response. Thirdly, cross-document synthesis necessitates a high-quality search. In the event of retrieval failure, a more powerful language model is unable to compensate for the information deficit. Fourthly, it is imperative to incorporate citations in order to ensure the verifiability and calibration of the system. In conclusion, a local solution that is built on open components has demonstrated competitiveness against the cloud-based NotebookLM, despite utilizing a considerably reduced language model and offering a mere command-line interface.

4 DISCUSSION AND CONCLUSIONS

The transition from the Naive RAG to the Agentic RAG approach (Xu *et al.*, 2025) was identified as a pivotal architectural decision. In the context of more complex queries, the linear sequence of context retrieval and response generation reaches a limit imposed by the model‘s context window and the one-time nature of the search. The replacement of the pipeline with a multi-agent system featuring a planning node (Query Planner), parallel search (Research Worker), filtering (Retrieval Grader), and iterative supplementary search (Completeness Checker) resulted, according to observations made during the development process, in the most significant enhancement to the system‘s overall success rate in a single step. These findings are consistent with the broader trend of RAG architectures shifting from a static to an agent-based paradigm, as described in the literature (Jeong *et al.*, 2024).

A secondary, yet substantial, decision was the integration of semantic and lexical search components. The utilization of metadata extracted by a language model as a pre-filter, a process which was subjected to initial testing, paradoxically resulted in outcomes that were inferior to those achieved by a pure hybrid search combining dense and sparse embeddings, a combination that is natively supported by the Qdrant database. The probable cause of this issue is the lack of reliability of smaller local models in the process of keyword extraction, in addition to errors in the specialized agent that had to link these words between the query and the metadata in the database.

The processing of structured data necessitated the integration of a parallel relational database. Semantic search has not been designed for aggregate queries on tables of the type “select the ten highest values”. Following an unsuccessful attempt to deploy PostgreSQL, whose syntax and schemas smaller local models could not consistently generate, the built-in analytical database DuckDB was selected. This proved effective even with less powerful models when interacting with the SQL agent.

The evaluation yielded several management-relevant implications for organizations considering the deployment of RAG on sensitive documents. Firstly, it is evident that local deployment with open LLMs is currently a viable alternative to cloud solutions, at least for organizations that can tolerate longer document processing times and the absence of a graphical user interface. Secondly, the accuracy of responses is contingent on computational power, particularly graphics memory capacity. The disparity in memory capacity between 18 GB of shared memory on a laptop and the dedicated VRAM of a high-performance graphics card resulted in a notable enhancement in success rates, from 80–90% to 95–100%, for models of identical design. Consequently, investment in hardware constitutes an indispensable component within the cost calculation for such a solution. Thirdly, the choice between RAG and model fine-tuning remains open. RAG is suitable where the document corpus grows dynamically and a cited response is required. Conversely, fine-tuning is advantageous when consistent style and deep domain expertise are needed (Ailoitte, 2026).

The selected architectural design is also subject to inherent limitations. Reasoning models, which engage in an internal chain of reasoning prior to outputting a result, have been demonstrated to be ineffective within Agent RAG. The high latency of these models, which is exacerbated by the number of agent nodes, results in a disproportionate overhead for simple decision-making tasks (such as intent classification in Router). In the context of multi-agent systems, it is therefore more appropriate to employ standard models and, if necessary, reserve the reasoning model solely for the final synthesis of the response. The limitations of the system are evident in a number of ways. Firstly, the test set is limited in size, comprising a mere five questions across a single domain, namely climate reports. Secondly, the absence of formal statistical evaluation is notable. Thirdly, the restriction to a command-line interface hinders usability for non-technical users.

The present paper has demonstrated that a RAG system prototype built exclusively on open-source components and operated entirely offline can achieve a quality comparable to that of a leading cloud-based competitor when performing factual queries on a large collection of unstructured documents. The key building blocks that were identified in this study are: firstly, the Agentic RAG architecture; secondly, hybrid search over a vector database; thirdly, a parallel relational database for tabular data; and finally, a multi-stage self-correcting mechanism (which includes a cosine similarity check during fact extraction, response completeness verification, and hallucination check). The primary constraints pertain to the protracted document processing time, attributable to semantic segmentation, the reliance on local hardware performance, and a user interface constrained to the command line. Consequently, the subsequent enhancement of the system is centered on the optimization of the ETL pipeline and the development of a desktop application that would facilitate the utilization of the solution by non-technical users. From the perspective of broader implications for organizational competitiveness, the prototype demonstrates that protecting corporate information assets while utilizing generative AI does not necessarily entail a significant compromise in output quality — provided the organization is willing to invest in appropriate local hardware and accept the current functional limitations of open models.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

Several AI tools were used in the preparation of this article. They assisted with summarizing source material, translating passages between languages, and identifying typographical and stylistic issues in the text. All output produced by these tools was reviewed and verified. Summaries were checked against the original sources, translations were assessed for accuracy, and suggested

corrections were evaluated before being accepted. The tools served as an aid to the writing process; the final content reflects the authors‘ judgment and remains the authors‘ responsibility.

REFERENCES

AILOITTE. 2026. RAG vs fine-tuning: which AI approach should businesses choose? *Ailoitte*. https://www.ailoitte.com/insights/rag-vs-fine-tuning.

CHAI. 2024. RAG 101: Ep 1 — retrieval-augmented generation (RAG), why? *Medium*. [Accessed: 2026-03-07] https://medium.com/@thechaiexperiment/rag-101-ep-1-retrieval-augmented-generation-rag-basics-aaf7cafdd0ce

HUANG, L., YU, W., MA, W., ZHONG, W., FENG, Z., WANG, H., CHEN, Q., PENG, W., FENG, X., QIN, B., LIU, T. 2025. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*. 43(2), 1–55. https://doi.org/10.1145/3703155

JAVANMARD, A. 2025. PDF data extraction benchmark 2025: comparing Docling, Unstructured, and LlamaParse for document processing pipelines. *Procyons*. [Accessed: 2026-01-24] https://procycons.com/en/blogs/pdf-data-extraction-benchmark/

JEONG, S. *et al*. 2024. Adaptive-RAG: learning to adapt retrieval-augmented large language models through question complexity. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* Volume 1: Long Papers. Mexico City: Association for Computational Linguistics, pp. 7036–7050.

PATRONUSAI. 2026. RAG evaluation metrics: best practices for evaluating RAG systems. *Patronus AI*. [Accessed: 2026-03-10] https://www.patronus.ai/llm-testing/rag-evaluation-metrics

QU, R., TU, R., BAO, F. S. 2025. Is semantic chunking worth the computational cost? In: *Findings of the Association for Computational Linguistics: NAACL 2025*. Albuquerque: Association for Computational Linguistics, pp. 2155–2177. https://doi.org/10.18653/v1/2025.findings-naacl.114

VASWANI, A. 2023. Attention is all you need. *arXiv*:1706.03762v7 https://arxiv.org/html/1706.03762v7

VLASTNÍK, D. 2026. *Prototyp RAG (Retrieval-augmented generation) pro analýzu velkého množství nestrukturovaných textů*. Master‘s thesis. Brno: Mendel University in Brno. Faculty of Business and Economics, Department of Informatics.

XU, X. *et al*. 2025. Agentic RAG with human-in-the-retrieval. In: *2025 IEEE 22nd International Conference on Software Architecture Companion (ICSA-C)*. Odense, Denmark. IEEE. https://doi.org/10.1109/ICSA-C65153.2025.00074.

ZHANG, Y. *et al*. 2025. Siren‘s song in the AI ocean: a survey on hallucination in large language models. *Computational Linguistics*. 51(4), 1373–1418. https://doi.org/10.1162/COLI.a.16

Contact information

David Vlastník: e-mail: xvlastni@node.mendelu.cz

Jaromír Landa: e-mail: jaromir.landa@mendelu.cz

THE ROLE OF HAPTIC PRODUCT EXPERIENCE IN CONSUMER EVALUATION: AN EYE-TRACKING AND PUPILLOMETRY STUDY OF REUSABLE BOTTLES

Jan Vrána[1], Stanislav Mokrý[1], Michal Pšurný[1]

[1]Department of Marketing and Trade, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_041

ABSTRACT

This study examines how physical and digital product presentation shape consumer evaluation in the context of reusable water bottles. The research focuses on whether the possibility of direct tactile interaction with a product influences perceived attractiveness, purchase intention, certainty of evaluation, and pupil response. An experimental design combining eye-tracking, pupillometry, and a post-exposure questionnaire was used to compare electronic product presentation with physical presentation involving product handling. The findings suggest that tactile interaction did not systematically change perceived product attractiveness. However, for selected products, the form of presentation affected purchase intention and certainty of evaluation, indicating that haptic information may become relevant when material or functional product characteristics are salient. The pupillometry results showed differences between physical and electronic presentation, although pupil dilation was not consistently associated with product attractiveness or purchase intention. Overall, the study contributes to sensory marketing research by showing that physical product experience may influence selected aspects of consumer decision-making, while its effect depends on the specific product and evaluated attribute. The results also highlight both the potential and limitations of pupillometry as a complementary method for studying consumer responses to haptic product presentation.

Keywords: haptics, product presentation, eye-tracking, pupillometry, consumer decision-making, sensory marketing, reusable bottles
JEL Code: M300, M310

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Online shopping has become an increasingly common way of purchasing a wide range of goods and services. This trend is reflected in the growing number of online shoppers and the increasing revenues generated by e-commerce. In the online environment, products are typically presented through several formats, most commonly photographs accompanied by product descriptions, zoom functions, product videos (Boardman and McCormick, 2019), or virtual mirrors (Cho and Schwarz, 2012). These presentation formats are designed to introduce the product to consumers, communicate its features, and motivate purchase.

Their main purpose is to provide consumers with relevant product information that supports the purchase decision-making process. However, most online presentation formats rely primarily on visual stimuli, or on a combination of visual and auditory stimuli. As a result, consumers are unable to touch, try, or physically experience the product due to the barrier between the virtual and physical shopping environments. Product presentation can therefore range from purely digital or multimedia-based formats to direct physical contact with the product, where information is acquired not only through vision and hearing but also through other senses, particularly touch.

This study is situated within the field of sensory marketing and focuses on the effect of product presentation form on consumer decision-making. Specifically, it compares two forms of product presentation: electronic and physical. Reusable water bottles were selected as the product category because their material, weight, shape, and handling characteristics may be relevant to consumer evaluation. The study uses eye-tracking with pupillometry to capture changes in pupil size, which may reflect consumers' physiological responses to product stimuli. These data are complemented by a questionnaire assessing consumers' perceptions of the different presentation forms.

2 LITERATURE REVIEW

2.1 Haptics and its effect on the consumer decision-making process

Although online shopping and various forms of digital product presentation have become common, e-commerce may still be perceived as less suitable for evaluating products with specific physical characteristics, such as material, texture, workmanship, and overall quality (Suh *et al.*, 2018). These attributes are often difficult to assess online, although they may play an important role in the consumer decision-making process. Their limited accessibility can therefore complicate product evaluation and increase the perceived risk associated with purchase decisions (Verhagen *et al.*, 2016). The absence of physical contact is particularly relevant because sensory perception influences emotions, memories, preferences, product choice, and product use (Suh *et al.*, 2018). Touch is especially important in retail settings, where direct product contact is a natural part of product evaluation and consumer behaviour (Holbrook, 1983).

Previous research suggests that the possibility of touching a product can influence consumers' attitudes, purchase decisions, trust in the product, and perceived product attractiveness. Direct physical contact enables consumers to evaluate the physical characteristics of a product more accurately. Peck and Childers (2003) argue that consumers often prefer to touch products before purchase, particularly when material characteristics are relevant to product evaluation. Vision

and touch are among the dominant senses through which consumers explore product characteristics (Balaji *et al.*, 2011). Similarly, Schifferstein and Cleiren (2005) suggest that when both visual and tactile information are available, consumers may form a stronger and more confident evaluation of the product.

The importance of touch in consumer behaviour has also been demonstrated in several empirical studies. Balaji *et al.* (2011) examined the effect of tactile and visual stimuli on product attitude, product evaluation, and purchase intention. Their findings indicate that tactile product information can significantly influence product evaluation. In their study, tactile stimuli had a stronger effect on attitude and purchase intention than visual stimuli. Participants also evaluated products more positively and with greater certainty when they were able to touch them; in contrast, the absence of touch could negatively affect product perception. Similar conclusions were reported by Subhash *et al.* (2019), who examined the role of touch in overall product evaluation. Their study, conducted in the United States and replicated in India, found that touch had a positive effect mainly for products with characteristics that needed to be explored haptically, such as blankets. For products for which touch was less relevant, such as flash drives, this effect was not confirmed. However, Vieira (2012) did not find evidence that touch necessarily changes purchase probability.

2.2 Tactile research using neuroscientific and psychophysiological methods

Haptics has also been studied using neuroscientific and psychophysiological methods. Compared with self-report methods, these approaches can provide more immediate and less verbally mediated data on consumer responses. They may also reduce some limitations associated with retrospective questioning, because self-reported answers are typically collected after exposure to the stimulus and may be influenced by memory, interpretation, or response bias. However, only a limited number of studies have examined haptics in relation to marketing and consumer decision-making using such methods (Vrána and Mokrý, 2020).

Park *et al.* (2015, 2018), for example, investigated the relationship between the physical characteristics of a product and consumer preference and satisfaction using EEG. Their findings showed an association between preference and gamma oscillations. Francis *et al.* (1999) also demonstrated that different types of touch, including pleasant, neutral, and painful touch, are associated with activity in different brain areas. Despite their potential, neuroscientific methods are not yet widely used in this research area, partly because they are often costly, technically demanding, and time-consuming (Vrána and Mokrý, 2020). For this reason, less demanding methods, such as eye-tracking and pupillometry, may offer a useful alternative for studying consumer behaviour.

Pupillometry is based on the measurement of changes in pupil size. According to Loewenfeld (1993), pupil size is not influenced only by the amount of light in the environment, but may also respond to non-visual stimuli, emotions, and cognitive processes. Pupil responses have been linked to mental effort, surprise, attention, decision certainty, belief formation, and information gain (Zénon, 2019). Pupil dilation has also been associated with food preference, taste, and music perception (Sirois and Brisson, 2014). These findings suggest that pupillometry may be useful for capturing physiological responses that accompany product evaluation.

Bertheaux *et al.* (2020) used pupillometry to measure emotional responses during tactile interaction with different material surfaces. Their results showed changes in pupil size during the first two seconds of touch, and the pattern of pupil response was consistent with findings reported by Bradley *et al.* (2008). Pupil dilation was observed in response to both positively and negatively

evaluated stimuli when compared with neutral stimuli. Similar use of pupil dilation as an indicator of consumer response can be found in Laeng *et al.* (2016), who focused on wine labels, and in Ramsøy *et al.* (2017), who examined clothing and found a relationship between pupil dilation and product preference.

2.3 Hypotheses development

The main objective of this study is to determine whether the form of product presentation, i.e. electronic or physical, influences consumer decision-making in the context of reusable water bottles. To achieve this objective, the study combines eye-tracking and pupillometry with questionnaire-based evaluation. The following hypotheses were formulated in relation to the research objective.

H1: There is no difference between electronic and physical product presentation in perceived bottle attractiveness.

This hypothesis is based on Peck and Wiggins (2006), who suggest that the possibility of touching a product may increase its perceived attractiveness.

H2: There is no difference between electronic and physical product presentation in purchase intention.

This hypothesis is based on McCabe *et al.* (2003), Peck and Wiggins (2006), and Balaji *et al.* (2011), whose studies indicate that consumers may be more likely to purchase a product when they are able to touch it before making a decision.

H3: There is no difference between electronic and physical product presentation in certainty of product evaluation.

This assumption follows Grohmann *et al.* (2007) and Balaji *et al.* (2011), who suggest that consumers may be more certain about their evaluation when they can physically interact with the product.

H4: There is no difference in pupil response between electronic and physical product presentation.

This hypothesis builds on the assumption that pupil dilation may reflect mental effort or arousal (Kahneman, 2013). Since physical product presentation involves product handling and tactile exploration, it may elicit a different pupil response than electronic presentation.

H5a: There is no difference in pupil response across levels of perceived attractiveness during electronic product presentation.

H5b: There is no difference in pupil response across levels of perceived attractiveness during physical product presentation.

H6a: There is no difference in pupil response across levels of purchase intention during electronic product presentation.

H6b: There is no difference in pupil response across levels of purchase intention during physical product presentation.

These hypotheses are based on previous studies suggesting that pupil dilation may occur in response to emotionally or motivationally relevant stimuli, including both positively and negatively evaluated stimuli, compared with neutral stimuli (Bertheaux *et al.*, 2020; Bradley *et al.*, 2008; Laeng *et al.*, 2016; Ramsøy *et al.*, 2017). In the present study, pupil response is therefore examined in relation to perceived attractiveness and purchase intention under both electronic and physical presentation conditions.

3 METHODOLOGY AND DATA

3.1 Preparatory phase of research

Reusable water bottles were selected as stimuli for this study. This product category was chosen because reusable bottles differ in several physical characteristics, such as material, weight, shape, and surface properties. These characteristics may be relevant for product evaluation and may increase the importance of touch during physical product presentation (Klatzky and Lederman, 1992, 1993; Chen *et al.*, 2009; Vieira, 2012).

In total, 25 reusable bottles were selected. Their prices ranged from 39 to 999 CZK, their volume ranged from 0.5 to 1 litre, and they represented six different types of material. The products were divided into two comparable groups for the purpose of A/B testing. Bottle no. 25 was used only during the preparatory phase and was not included in the main analysis.

Graphical materials were prepared for each stimulus. These materials consisted of real product photographs obtained from online sources and were supplemented with specific product information, including manufacturer, product name, volume, weight, height, material, price, and additional characteristics.

The eye-tracking experiment was created in SR Research Experiment Builder, version 2.2.61, and consisted of two versions, A and B. Both versions included 24 reusable bottles. In version A, the first 12 bottles from the first product group were accompanied by a hand symbol, indicating that participants would see the bottle on the screen and could also touch and manipulate the physical product. The remaining 12 bottles were accompanied by an eye symbol, indicating that the product would be presented only visually on the monitor. Version B used the opposite assignment of presentation forms to product groups. This design ensured that each product was presented in both conditions across participants, while each participant evaluated a given product in only one presentation condition. Therefore, comparisons between electronic and physical presentation for individual products were treated as between-group comparisons.

The experiment was conducted in a room with covered windows in order to reduce changes in pupil size caused by variations in ambient light. Participants also wore noise-cancelling headphones, which reduced surrounding noise by 27 dB.



Fig. 1: An overview of reusable bottles used in the research

3.2 Research realization phase

The research was conducted at the Faculty of Business and Economics, Mendel University in Brno, between 3 March and 12 March 2020. The sample consisted of 22 participants, including 10 men and 12 women, mostly students aged 18–26. All participants took part in two experimental sessions, which were conducted exactly seven days apart. Eye-tracking data were recorded using a stationary SR Research EyeLink 1000 Plus eye-tracker.

Before each product stimulus was displayed, either a hand symbol or an eye symbol appeared on the monitor for 6,000 ms. If the eye symbol was displayed, participants were informed that the product would be presented only on the monitor. If the hand symbol was displayed, participants saw the product on the monitor and were also given the corresponding physical bottle, which they could touch and manipulate.

All product stimuli, symbols, and fixation crosses were presented on a white background (RGB: 255, 255, 255). Each product stimulus was displayed for 25,000 ms. During the physical presentation condition, a shading desk was placed under the chin rest to prevent participants from



Fig. 2: The experimental procedure: participant handling a test bottle

looking directly at the physical bottle in their hands instead of the image displayed on the monitor. The sequence of stimuli was randomized so that electronic and physical presentation conditions alternated throughout the experiment.

After all stimuli had been presented, participants were informed that the eye-tracking part of the experiment had ended. They then completed an electronic questionnaire displayed on the monitor. In the questionnaire, participants evaluated each bottle in terms of perceived attractiveness, purchase intention, and certainty of evaluation using a seven-point Likert scale.

3.3 Data processing and statistical analysis

The collected data were analysed using IBM SPSS Statistics 25, Microsoft Excel, and SR Research EyeLink Data Viewer. Eye-tracking data were exported for the right eye only, as this was considered sufficient for the purposes of the study. Several variables were used for data analysis and hypothesis testing.

Questionnaire responses were measured on a seven-point Likert scale and were therefore treated as ordinal data. The pupillometric part of the analysis was based on a dilation/constriction coefficient, which captured changes in pupil size during the first two seconds of stimulus exposure. This time window was selected because previous research suggests that the most pronounced pupil response to tactile stimulation may occur shortly after stimulus onset (Bertheaux *et al.*, 2020).

Pupil data were recorded at a sampling frequency of 1,000 Hz. For easier processing, the data were averaged in EyeLink Data Viewer before further analysis. The pupil-size change coefficient was then calculated using the following formula:

$$[d(t) - d(25)] / d(25) \quad \textbf{(1)}$$

In this formula, d(25) represents the mean pupil size during the 25-second stimulus presentation, while d(t) represents pupil size at time t during the first two seconds of exposure. The resulting pupil-size change coefficients were averaged across the first two seconds and used as the main pupillometric variable in hypothesis testing.

The hypotheses concerning the effect of presentation form on perceived bottle attractiveness, purchase intention, and certainty of evaluation were tested using the Wilcoxon rank-sum test. This test was selected because each participant evaluated a given product in only one presentation condition; therefore, the comparison between electronic and physical presentation was based on independent groups. Analysis of variance (ANOVA) or the non-parametric Kruskal–Wallis test was used to determine the effect of perceived attractiveness and purchase intention on pupil-size change in the two presentation conditions. When the assumption of normality was violated, the non-parametric test was used.

4 RESULTS

4.1 Hypothesis 1

As shown in Fig. 3, perceived bottle attractiveness varied for some products depending on the form of presentation. The largest differences were observed for bottles no. 4, 3, and 1, for which physical presentation led to higher attractiveness ratings. In contrast, bottles no. 16, 23, 24, and particularly bottle no. 17 received higher attractiveness ratings in the electronic presentation condition.

However, the Wilcoxon test did not confirm a statistically significant difference in perceived attractiveness between the two presentation forms. Bottles no. 4 and 17 were closest to statistical significance. For bottle no. 4, physical presentation had a positive effect on perceived attractiveness, whereas for bottle no. 17 the opposite pattern was observed.

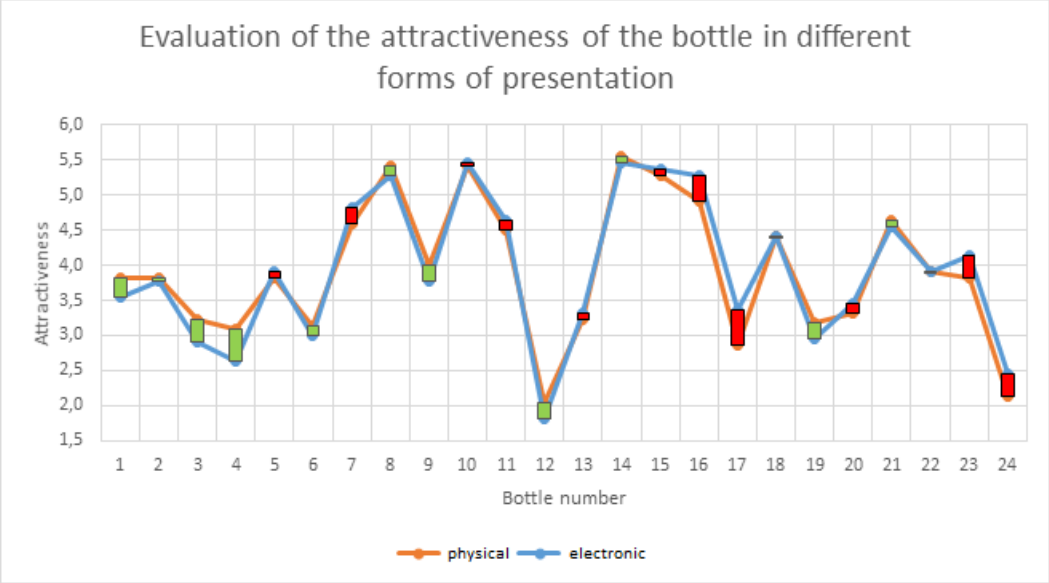


Fig. 3: Bottle attractiveness ratings by presentation form (1 – very unattractive; 7 – very attractive)

4.2 Hypothesis 2

As shown in Fig. 4, purchase intention tended to be more positively affected by physical presentation than by electronic presentation for several products. Positive differences were observed particularly for bottles no. 12 and 13, which were also among the bottles with the lowest overall purchase intention ratings. In contrast, bottle no. 7 showed a substantial decrease in purchase intention in the physical presentation condition.

The Wilcoxon test confirmed a statistically significant effect of presentation form on purchase intention for two bottles. Although differences were also observed for several other bottles, these differences were not statistically significant. Physical presentation had a negative effect on purchase intention for bottle no. 7, whereas a positive effect was observed for bottle no. 12.

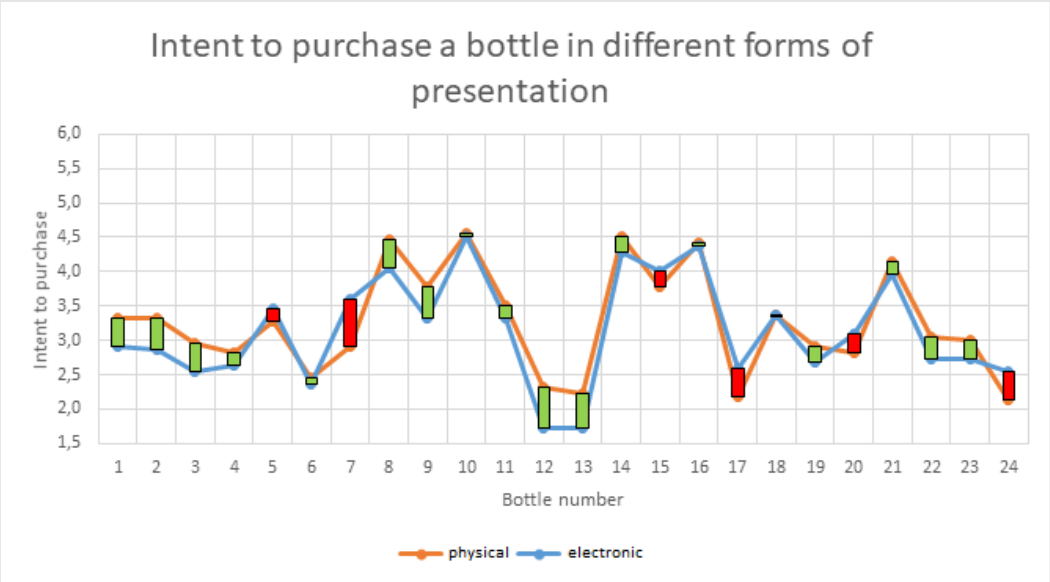


Fig. 4: Purchase intention ratings by presentation form (1 – I am certainly not interested to purchase; 7 –I am certainly interested to purchase)

4.3 Hypothesis 3

The final self-reported variable examined was participants' certainty of evaluation for individual bottles in relation to presentation form. As shown in Fig. 5, differences in certainty ratings were generally small across all products. Mean certainty values ranged between 5.2 and 6.2, indicating relatively high certainty in both presentation conditions.

The Wilcoxon test confirmed a statistically significant effect of presentation form only for bottle no. 20. For the remaining bottles, the differences between electronic and physical presentation were not statistically significant.

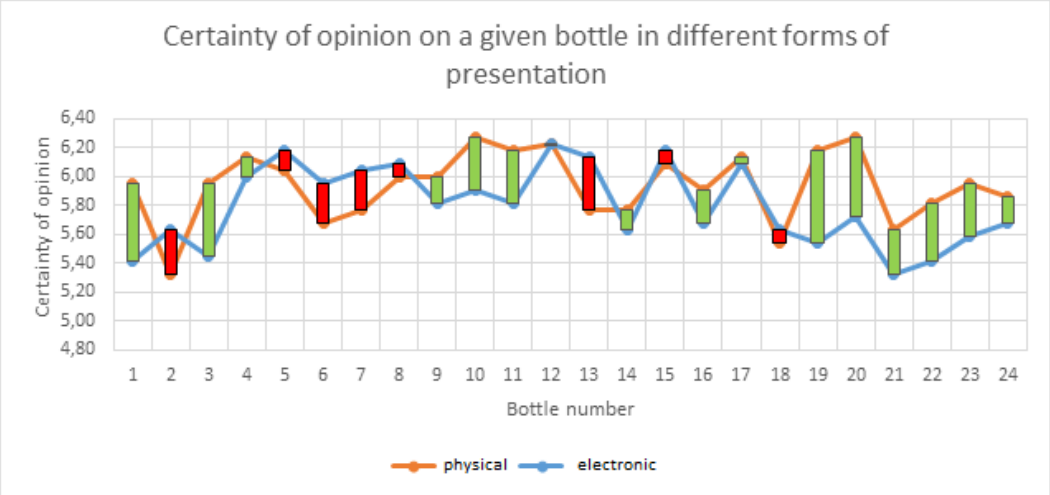


Fig. 5: : Certainty of evaluation by presentation form (1 – I am not certain at all; 7 – I am completely certain) 7 –I am certainly interested to purchase)

4.4 Hypothesis 4

The results showed that presentation form affected pupil-size change in 17 bottles, representing 71% of the analysed products. In the remaining seven bottles, specifically bottles no. 2, 7, 8, 9, 10, 11, and 22, the mean pupil size was higher in the physical presentation condition than in the electronic presentation condition, but these differences were not statistically significant.

Overall, the results indicate a general tendency towards stronger pupil response during physical product presentation. This pattern may be related to the additional effort involved in handling and manipulating the bottles during the physical presentation condition.

4.5 Hypothesis 5a, 5b

Pupil size was used as the dependent variable, while perceived attractiveness was used as the independent variable. The hypotheses were tested using all bottles simultaneously. For the electronic presentation condition, the assumption of normality was met, and analysis of variance was therefore used. For the physical presentation condition, the assumption of normality was not met, and the non-parametric Kruskal–Wallis test was applied.

The results did not confirm a statistically significant effect of perceived attractiveness on pupil-size change in either presentation condition. In both cases, the p-values were above the selected significance level of $\alpha = 0.05$. Therefore, the null hypotheses were not rejected. The physical presentation condition was also tested using the parametric variant of the test for verification, and the results remained unchanged.

4.6 Hypothesis 6a, 6b

The effect of purchase intention on pupil-size change was also not statistically significant. Although the p-values were lower than those obtained for perceived attractiveness, they remained above the selected significance level of $\alpha = 0.05$. Therefore, the null hypotheses were not rejected for either presentation condition.

Figures 6 and 7 illustrate the course of pupil-size change during the first two seconds of stimulus exposure. The bottles were divided into three groups according to purchase intention. The first group included the eight bottles with the highest purchase intention ratings, the second group included eight bottles with medium purchase intention ratings, and the third group included the eight bottles with the lowest purchase intention ratings.

Only small differences between these groups were observed, particularly around 1,000 ms after stimulus onset. However, the statistical tests did not confirm these differences as significant.

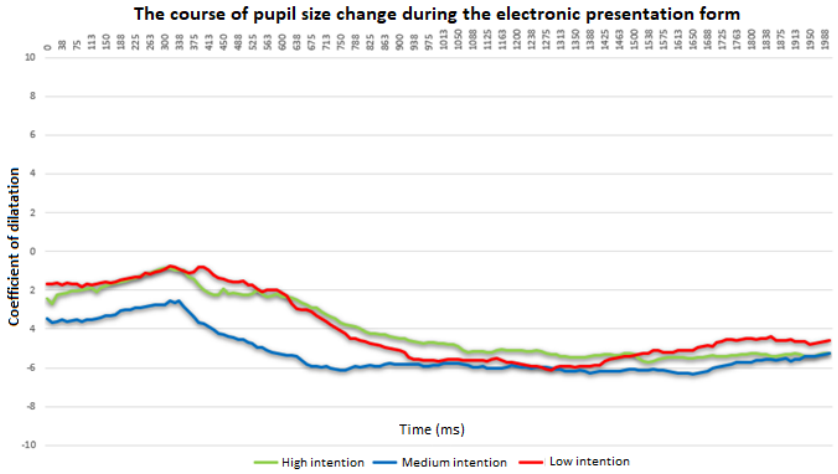


Fig. 6: Pupil-size change during the first two seconds of exposure according to purchase intention in the electronic presentation condition

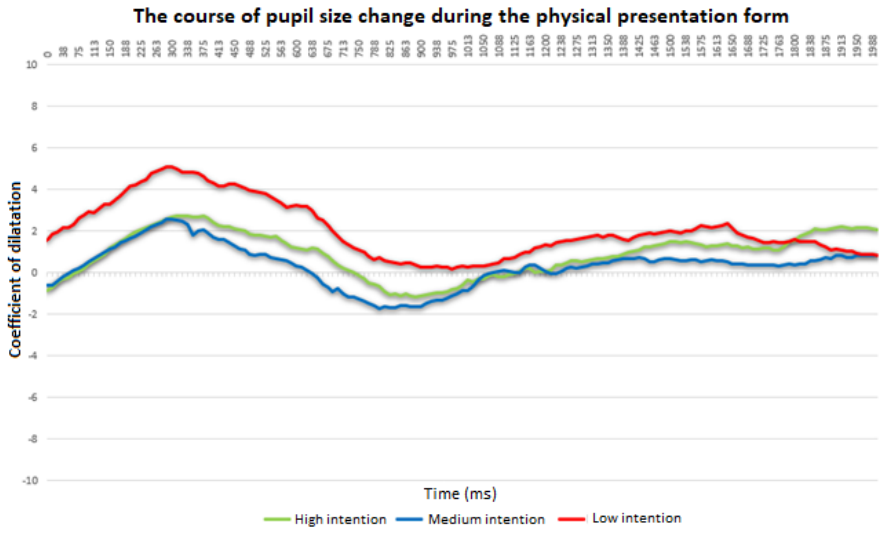


Fig. 7: Pupil-size change during the first two seconds of exposure according to purchase intention in the physical presentation condition

5 DISCUSSION AND CONCLUSION

The results of this study did not confirm a significant effect of presentation form on perceived product attractiveness. The difference between the present findings and those reported by Subhash *et al.* (2019) may be explained by the fact that this study focused specifically on perceived attractiveness, whereas Subhash *et al.* (2019) examined overall product evaluation. Although these constructs are closely related, they are not identical. Another possible explanation is the product category used in the research. In the study by Subhash *et al.* (2019), presentation form affected mainly products with haptically relevant characteristics, such as blankets and sweaters, while this effect was not observed for products for which touch was less important, such as flash drives and DVDs. It is therefore possible that, in the case of reusable water bottles, haptic characteristics such as material, weight, or surface properties are related more to perceived usability and quality than to perceived attractiveness.

The results also indicate that presentation form may affect purchase intention to some extent. Similar conclusions have been reported in previous research. Balaji *et al.* (2011) found a positive effect of physical product presentation on purchase intention for tissue papers, whereas Vieira (2012) did not find a relationship between touch and purchase probability in the case of hand soaps. These findings are also consistent with McCabe (2003), who suggests that physical product examination is more important for material products, where characteristics such as material, weight, temperature, and surface texture are relevant, than for geometric products, where shape and size are the main attributes. In the present study, a statistically significant effect of presentation form was found only for two bottles. This may suggest that the haptic characteristics of most bottles were not sufficiently distinctive to substantially influence purchase intention.

Another hypothesis concerned the effect of touch on certainty of product evaluation. Previous studies by Grohmann *et al.* (2007) and Balaji *et al.* (2011) suggest that tactile input may increase consumers' certainty when evaluating products. In the present study, however, this effect was confirmed only for bottle no. 20. For bottles no. 1, 3, and 19, a similar tendency was observed, but the differences were not statistically significant. One possible explanation is that, during the electronic presentation condition, written descriptions provided detailed information about product characteristics that would otherwise be accessible mainly through touch. As a result, participants may have been able to form a relatively clear evaluation of the products even without physical contact. The availability of detailed product information may therefore have increased certainty in the electronic presentation condition and reduced the additional value of touch.

The relationship between presentation form and pupil-size change was also examined. The results showed that physical presentation was associated with pupil dilation in most of the analysed bottles. Since pupil size can be influenced by many factors that are difficult to fully control, these results should be interpreted with caution (Sirois and Brisson, 2014). The stronger pupil response observed during physical presentation may reflect the additional sensory, attentional, or motor demands associated with handling the product. However, it cannot be interpreted as a direct indicator of preference or emotional response without considering other possible explanations.

Previous studies by Bertheaux *et al.* (2020), Bradley *et al.* (2008), Laeng *et al.* (2016), and Ramsøy *et al.* (2017) suggest that pupil dilation may occur in response to both positively and negatively evaluated stimuli when compared with neutral stimuli. In this study, pupil-size change was examined in relation to perceived attractiveness and purchase intention under both presentation conditions. However, the results did not confirm a statistically significant relationship between pupil-size change and these evaluation variables. Although small differences between

the best-rated, medium-rated, and lowest-rated groups were visible in the time course of pupil-size change, particularly during the first second after stimulus onset, these differences were not statistically significant. One possible explanation is the absence of sufficiently extreme evaluations. The mean ratings of most bottles ranged from 1.9 to 5.4 on a seven-point scale, which means that highly positive evaluations were largely absent. Since pupil dilation is expected to be more pronounced for highly positive or highly negative stimuli, the limited range of evaluations may have reduced the likelihood of detecting significant differences.

As online shopping continues to grow, the barrier between virtual and physical product experience remains an important topic for consumer research. This study contributes to sensory marketing by examining how electronic and physical product presentation influence consumer evaluation and pupil response. The findings suggest that touch does not necessarily increase perceived attractiveness, but it may influence purchase intention and certainty of evaluation for selected products. The results also show that pupillometry can capture differences between presentation forms, although its interpretation requires caution.

Future research should further examine product categories with more distinctive haptic characteristics. In the case of reusable water bottles, future studies could include metal bottles and bottles with more diverse shapes, such as rectangular designs. It would also be useful to separate the two presentation conditions more clearly. In the electronic condition, participants could evaluate only product photographs or videos on a monitor, whereas in the physical condition they could freely handle and visually inspect the real product without simultaneous presentation on the screen. This would make it possible to compare digital and physical product experience more directly and reduce potential overlap between the two presentation forms.

Acknowledgments

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, co-financed by Operational Programme Research, Development and Education.

REFERENCES

BALAJI, M. S., RAGHAVAN, S., JHA, S. 2011. Role of tactile and visual inputs in product evaluation: a multisensory perspective. *Asia Pacific Journal of Marketing and Logistics*. 23(4), 513–530. https://doi.org/10.1108/13555851111165066

BERTHEAUX, C., TOSCANO, R., FORTUNIER, R., ROUX, J.-C., CHARIER, D., BORG, C. 2020. Emotion measurements through the touch of materials surfaces. *Frontiers in Human Neuroscience*. 13, 455. https://doi.org/10.3389/fnhum.2019.00455

BOARDMAN, R., MCCORMICK, H. 2019. The impact of product presentation on decision-making and purchasing. *Qualitative Market Research: An International Journal*. 22(3), 365–380. https://doi.org/10.1108/QMR-09-2017-0124

BRADLEY, M. M., MICCOLI, L., ESCRIG, M. A., LANG, P. J. 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*. 45(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

CHEN, X., SHAO, F., BARNES, C. J., CHILDS, T., HENSON, B. 2009. Exploring relationships between touch perception and surface physical properties. *International Journal of Design*. 3(2), 67–76.

CHO, H., SCHWARZ, N. 2012. I like your product when I like my photo: misattribution using interactive virtual mirrors. *Journal of Interactive Marketing*. 26(4), 235–243. https://doi.org/10.1016/j.intmar.2012.03.003

FRANCIS, S., ROLLS, E. T., BOWTELL, R., MCGLONE, F., O'DOHERTY, J., BROWNING, A., CLARE, S., SMITH, E. 1999. The representation of pleasant touch in the brain and its relationship with taste and olfactory areas. *NeuroReport*. 10(3), 453–459. https://doi.org/10.1097/00001756-199902250-00003

GROHMANN, B., SPANGENBERG, E. R., SPROTT, D. E. 2007. The influence of tactile input on the evaluation of retail product offerings. *Journal of Retailing*. 83(2), 237–245. https://doi.org/10.1016/j.jretai.2006.09.001

HOLBROOK, M. B. 1983. On the importance of using real products in research on merchandising strategy. *Journal of Retailing*. 59(1), 4–23.

KAHNEMAN, D. 2013. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

KLATZKY, R. L. ,LEDERMAN, S. J. 1992. Stages of manual exploration in haptic object identification. *Perception & Psychophysics*. 52(6), 661–670. https://doi.org/10.3758/BF03211702

KLATZKY, R. L., LEDERMAN, S. J. 1993. Toward a computational model of constraint-driven exploration and haptic object identification. *Perception*. 22(5), 597–621. https://doi.org/10.1068/p220597

LAENG, B., SUEGAMI, T., AMINIHAJIBASHI, S. 2016. Wine labels: an eye-tracking and pupillometry study. *International Journal of Wine Business Research*. 28(4), 327–348. https://doi.org/10.1108/IJWBR-03-2016-0009

LOEWENFELD, I. E. 1993. *The Pupil: Anatomy, Physiology, and Clinical Applications*. Detroit, MI: Wayne State University Press.

MCCABE, D. B., NOWLIS, S. M. 2003. The effect of examining actual products or product descriptions on consumer preference. *Journal of Consumer Psychology*. 13(4), 431–439. https://doi.org/10.1207/S15327663JCP1304_10

PARK, W., KI, D., KIM, D.-H., KWON, G. H., KIM, S.-P., KIM, L. 2015. EEG correlates of user satisfaction of haptic sensation. In: *2015 IEEE International Conference on Consumer Electronics (ICCE). IEEE*. pp. 569–570. https://doi.org/10.1109/ICCE.2015.7066529

PARK, W., KIM, D.-H., KIM, S.-P., LEE, J.-H., KIM, L. 2018. Gamma EEG correlates of haptic preferences for a dial interface. *IEEE Access*. 6, 22324–22331. https://doi.org/10.1109/ACCESS.2018.2827023

PECK, J., CHILDERS, T. L. 2003. Individual differences in haptic information processing: the "Need for Touch" scale. *Journal of Consumer Research*. 30(3), 430–442. https://doi.org/10.1086/378619

PECK, J., WIGGINS, J. 2006. It just feels good: customers' affective response to touch and its influence on persuasion. *Journal of Marketing*. 70(4), 56–69. https://doi.org/10.1509/jmkg.70.4.056

RAMSØY, T. Z., JACOBSEN, C., FRIIS-OLIVARIUS, M., BAGDZIUNAITE, D., SKOV, M. 2017. Predictive value of body posture and pupil dilation in assessing consumer preference and choice. *Journal of Neuroscience, Psychology, and Economics*. 10(2–3), 95–110. https://doi.org/10.1037/npe0000073

SCHIFFERSTEIN, H. N. J., CLEIREN, M. P. H. D. 2005. Capturing product experiences: a split-modality approach. *Acta Psychologica*. 118(3), 293–318. https://doi.org/10.1016/j.actpsy.2004.10.009

SIROIS, S., BRISSON, J. 2014. Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*. 5(6), 679–692. https://doi.org/10.1002/wcs.1323

JHA, S., BALAJI, M. S., STAFFORD, M. B. R., SPEARS, N. 2020. Haptic information processing and need for touch in product evaluation. *Journal of Consumer Marketing*. 37(1), 55–64. https://doi.org/10.1108/JCM-02-2018-2554

SUH, Y., KIM, S., KIM, W. 2018. Visual tactile attributes in online product presentations for improving purchase intention. *International Journal of Contents*. 14(2), 7–15.https://doi.org/10.5392/IJOC.2018.14.2.007

VERHAGEN, T., VONKEMAN, C., VAN DOLEN, W. 2016. Making online products more tangible: the effect of product presentation formats on product evaluations. *Cyberpsychology, Behavior, and Social Networking*. 19(7), 460–464. https://doi.org/10.1089/cyber.2015.0520

VIEIRA, V. A. 2012. An evaluation of the Need for Touch scale and its relationship with need for cognition, need for input, and consumer response. *Journal of International Consumer Marketing*. 24(1–2), 57–78. https://doi.org/10.1080/08961530.2012.650139

VRÁNA, J., MOKRÝ, S. 2020. Haptics and its effect on consumers' intentions using neuroscientific methods: literature review. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*. 68(2), 451–457. https://doi.org/10.11118/actaun202068020451

ZÉNON, A. 2019. Eye pupil signals information gain. *Proceedings of the Royal Society B: Biological Sciences*. 286(1911), 20191593. https://doi.org/10.1098/rspb.2019.1593

Contact information

Jan Vrána: e-mail: vranjes@email.cz

Stanislav Mokrý: e-mail: xmokry@mendelu.cz

Michal Pšurný: e-mail: xpsurny@mendelu.cz

FOREIGN DIRECT INVESTMENT, OIL REVENUES, AND TRADE AS DETERMINANTS OF ECONOMIC GROWTH IN GCC COUNTRIES

Evans Yeboah[1]

[1]Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_042

ABSTRACT

This study investigates the roles of foreign direct investment (FDI), oil revenues, and trade in shaping economic growth in Gulf Cooperation Council (GCC) countries over the period 1995–2024. The analysis employs the pooled mean group autoregressive distributed lag (PMG-ARDL) model to capture both long-run and short-run dynamics, and uses fully modified ordinary least squares (FMOLS) for robustness. The results indicate that FDI and oil rents exert positive, statistically significant effects on long-run per capita income. In contrast, merchandise trade does not exhibit a significant contribution to growth. The short-run estimates confirm the presence of a stable adjustment mechanism toward the long-run equilibrium. These findings highlight the dominant role of resource revenues and external capital inflows in the GCC growth process and point to the need for policies that promote diversification and strengthen non-oil sectors.

Keywords: FDI, merchandise trade, oil rents, economic growth
JEL Code: F43, F21, O11, Q32

1 INTRODUCTION

Economic growth remains a significant indicator of economic performance, reflecting improvements in income, employment, and overall welfare. Traditional growth theories emphasize capital accumulation, labour expansion, and technological progress (Solow, 1956; Swan, 1956; Saleem *et al.*, 2020), while more recent approaches incorporate external drivers such as foreign direct investment, trade, and natural resource endowments. FDI contributes to growth through capital inflows, technology transfer, and integration into global production networks (Borensztein *et al.*, 1998; Yimer, 2022; Duodu and Baidoo, 2020). Trade openness enhances efficiency and allows countries to exploit comparative advantages (Frankel and Romer, 1999; Sirag *et al.*, 2018). However, the growth effects of FDI and trade may differ across economies depending on structural characteristics such as resource dependence, institutional quality, and sectoral composition.

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

In resource-dependent economies, natural resource revenues often remain the dominant source of growth, shaping both macroeconomic stability and long-term development paths. In the Gulf Cooperation Council (GCC) countries, Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates, oil revenues have historically driven economic expansion. Nevertheless, dependence on hydrocarbons exposes these economies to external shocks, especially oil price volatility (Sachs and Warner, 1995; Alfaro *et al.*, 2004). Although oil revenues can support infrastructure development and public investment, excessive reliance on hydrocarbons may weaken economic diversification and increase vulnerability to external market fluctuations. This has intensified policy efforts toward economic diversification, with increasing emphasis on FDI and trade as alternative growth drivers. Several GCC countries have therefore introduced long-term economic transformation programs aimed at strengthening non-oil sectors and attracting foreign investment.

Although previous studies have examined the relationships among FDI, trade openness, oil revenues, and economic growth, the empirical findings remain mixed in resource-rich economies. Some studies report positive contributions of FDI and trade to economic growth, while others find limited or insignificant effects due to continued dependence on oil revenues and structural constraints. In addition, much of the existing literature examines these variables separately rather than within a unified empirical framework. The limited number of studies focusing specifically on GCC countries also suggests the need for updated empirical evidence that jointly evaluates the relative importance of FDI, trade, and oil rents in influencing economic growth.

This study investigates the long-run and short-run effects of FDI, merchandise trade, and oil rents on economic growth in GCC countries over the period 1995–2024. By applying the PMG-ARDL approach, the analysis accounts for heterogeneous short-run dynamics while imposing a common long-run relationship. The study contributes to the literature by jointly assessing the roles of foreign investment, trade integration, and oil revenues within a unified framework, thereby providing updated evidence on the determinants of growth in resource-dependent GCC economies.

2 METHODOLOGY

2.1 Data Source

This study uses annual panel data for selected GCC countries, namely Bahrain, Kuwait, Oman, Qatar, and Saudi Arabia, covering the period 1995–2024. All variables are obtained from the World Bank's World Development Indicators (WDI). The United Arab Emirates is excluded due to data availability constraints. Despite the exclusion of the United Arab Emirates due to data limitations, the selected sample still captures the major structural characteristics of GCC resource-dependent economies. Economic growth is proxied by GDP per capita growth (GDPPCG). The explanatory variables include foreign direct investment (FDI), merchandise trade (MERCT), and oil rents (ORT). FDI is measured as net inflows as a percentage of GDP, merchandise trade is measured as the sum of exports and imports of goods relative to GDP, while oil rents represent the share of oil revenues in GDP. All explanatory variables are expressed as percentages of GDP to improve comparability across countries and over time. The selection of variables is consistent with both theoretical and empirical literature on growth in resource-dependent economies, where external capital inflows, trade integration, and natural resource revenues are considered important determinants of macroeconomic performance.

2.2 Model Specification

To examine the relationship between the variables, the baseline model is specified in Equation 1.

$$GDPPCG_{it} = \beta_1 FDI_{it} + \beta_2 ORT_{it} + \beta_3 MERCT_{it} + \varepsilon_{it} \quad \textbf{(1)}$$

where $GDPPCG_{it}$ represents GDP per capita growth, FDI denotes foreign direct investment inflows, MERCT captures merchandise trade, and ORT represents oil rents. The error term ε accounts for unobserved country-specific effects. A positive relationship is expected between FDI and economic growth due to technology transfer, capital accumulation, and productivity spillovers. Merchandise trade is also expected to positively influence growth through increased market access and efficiency gains. However, the effect of oil rents may be either positive or negative depending on whether oil revenues promote productive investment or increase dependence on the resource sector.

2.3 Estimation Technique

The pooled mean group autoregressive distributed lag estimator proposed by Pesaran *et al.* (1999) is employed to estimate both long-run and short-run relationships. The approach allows short-run coefficients, intercepts, and error variances to differ across countries while constraining long-run coefficients to be homogeneous. This is appropriate given the structural similarities among GCC economies alongside potential short-run heterogeneity. The PMG-ARDL approach is appropriate for this study because GCC economies share similar long-run structural characteristics associated with oil dependence, trade integration, and investment policies, although short-run adjustment dynamics may differ across countries. Another advantage of the PMG-ARDL framework is that it can be applied when variables are integrated of mixed order I(0) and I(1), provided none of the variables are integrated of order I(2). In addition, the dynamic lag structure of the ARDL specification helps reduce potential endogeneity and simultaneity bias by incorporating lagged values of the explanatory variables.

The use of a cointegration-based framework also minimizes the possibility of spurious regression and allows the distinction between short-run dynamics and long-run equilibrium relationships. For robustness, fully modified ordinary least squares (FMOLS) is also applied to validate the long-run estimates. FMOLS is employed as a robustness estimator because it corrects for potential endogeneity and serial correlation associated with cointegrated panel regressions. The estimator produces consistent long-run coefficients by accounting for the possible correlation between the regressors and the error term. The PMG-ARDL baseline model is in Equation 2.

$$y_{it} = a_i + \sum_{j=1}^{p} \beta_{ij} y_{it-j} + \sum_{j=1}^{q} \delta_{ij} K_{it-j} + \varepsilon_{it} \quad \textbf{(2)}$$

Where y_{it} is the dependent variable, a_i is the constant, β_{ij} is the lag coefficient of the dependent variable y_{it-j}, K_{it-j} represents the regressors with δ_{ij} as their corresponding coefficients, and ε_{it} is the error term in the model.

3 RESULTS

3.1 Descriptive Statistics

Table 1 presents descriptive statistics for the selected variables based on 150 observations. GDP per capita growth exhibits considerable volatility, with a small positive mean and a negative median, indicating uneven growth performance across GCC countries and periods. The wide gap between the maximum value (24.93) and the minimum value (−11.86) further reflects substantial fluctuations in economic performance, likely associated with oil price cycles and external economic shocks. Foreign direct investment shows substantial dispersion, reflecting periodic inflows rather than stable investment patterns. The large variation between the maximum and minimum FDI values suggests that foreign investment inflows differ significantly across GCC economies and over time, partly due to differences in investment policies, market size, and sectoral diversification efforts.

Merchandise trade accounts for a large share of GDP, highlighting the high degree of trade exposure in GCC economies. The relatively high average trade ratio confirms the openness of GCC economies to international markets through hydrocarbon exports and import dependence. Oil rents display wide variation, confirming the central role of hydrocarbon revenues and their sensitivity to global price fluctuations. The substantial dispersion in oil rents also demonstrates the heterogeneity of resource dependence among GCC countries, with some economies relying more heavily on oil revenues than others.

Tab. 1 Descriptive Statistics

	GDPPCG	MERCT	ORT	FDI
Mean	0.284	77.42	29.03	2.324
Median	−0.183	75.70	28.26	1.112
Maximum	24.93	129.3	58.36	29.52
Minimum	−11.86	40.60	2.67	-2.760

3.2 Cross-Sectional Dependence and Unit Root Tests

Cross-sectional dependence tests in Table 2 reveal strong evidence of interdependence among GCC economies for GDP per capita growth, merchandise trade, and oil rents, suggesting the presence of common shocks and regional spillovers. This outcome is expected given the strong economic integration of GCC countries through oil markets, regional trade linkages, and similar macroeconomic structures. Accordingly, second-generation panel unit root tests are employed. The CIPS results in Table 3 indicate that the variables exhibit mixed orders of integration but become stationary after first differencing, satisfying the requirements for panel ARDL estimation. Since none of the variables are integrated of order I(2), the PMG-ARDL framework remains appropriate for estimating both short-run and long-run relationships.

Tab. 2 Cross-sectional dependence test

Test	GDPPCG	FDI	MERCT	ORT
Breusch-Pagan LM	24.02***	15.06	67.7***	196***
Pesaran Scaled LM	3.136***	1.131	12.90***	41.61***
Bias-corrected Scaled LM	3.050***	1.045	12.82***	41.53***
Pesaran CD	3.322***	1.666	6.140***	13.65***

Note: ***indicates statistical significance at the 1%

Tab. 3 CIPS unit root results

Variable	Constant	Constant & Trend
GDPPCG	−2.576***	−2.271
ΔGDPPCG	−4.821***	−4.803***
MERCT	−0.962	−1.999
ΔMERCT	−4.269***	−4.083***
FDI	−3.055***	−3.230***
ΔFDI	−5.177***	−5.078***
ORT	−1.911	−2.493
ΔORT	−4.115***	−4.346***

Note: ***indicates statistical significance at the 1%

3.3 PMG-ARDL Estimation Results

The PMG-ARDL long-run estimates in Table 4 indicate that foreign direct investment and oil rents exert positive, statistically significant effects on economic growth in GCC countries. The coefficient of FDI is positive and statistically significant at the 1% level, indicating that increases in foreign capital inflows contribute positively to GDP per capita growth. Specifically, a one-unit increase in FDI is associated with approximately a 1.27-unit increase in economic growth, suggesting that foreign investment plays an important role in promoting capital accumulation, technological transfer, and productivity enhancement in GCC economies. Oil rents also positively influence growth, reflecting their role in financing infrastructure development and public investment in resource-rich economies. The positive coefficient on oil rents confirms the continued importance of hydrocarbon revenues as a major source of economic expansion in GCC countries despite ongoing diversification efforts. Merchandise trade, however, does not exhibit a statistically significant long--run effect on growth. This finding suggests that trade expansion alone may not be sufficient to generate sustained income growth when exports remain heavily concentrated in the energy sector, and non-oil export diversification remains limited. In the short run, the error-correction term is negative and statistically significant. The coefficient of -0.667 implies that approximately 66.7% of short-run disequilibrium is corrected within one period, indicating a relatively rapid adjustment toward long-run equilibrium following economic shocks.

Tab. 4 Estimated coefficients from the PMG model

	Coefficient	Std. Error	t-Statistic	p-value
Long run				
FDI	1.266	0.279	4.529	0.000
ORT	0.296	0.052	5.621	0.000
MERCT	−0.030	0.070	−0.440	0.660
Constant	3.138	9.143	0.343	0.732
Short-run				
ECT (-1)	−0.667	0.210	−3.173	0.002
ΔFDI	−0.841	0.716	−1.174	0.242
ΔORT	−0.317	0.302	−1.049	0.296
ΔMERCT	−0.297	0.431	−0.689	0.491

3.4 Robustness Analysis

FMOLS estimates in Table 5 corroborate the baseline PMG-ARDL findings. Foreign direct investment and oil rents remain positive and statistically significant determinants of economic growth, confirming the robustness of the long-run relationship. The consistency of coefficient signs and statistical significance across both PMG-ARDL and FMOLS estimations strengthens the reliability of the empirical results and reduces concerns regarding model instability. Merchandise trade continues to show no significant effect, reinforcing the view that trade-driven growth in the GCC may depend on structural diversification rather than trade volume alone. The robustness results, therefore, support the argument that economic growth in GCC countries remains more strongly associated with foreign investment inflows and oil-based revenues than with merchandise trade expansion.

Tab. 5 FMOLS estimated results

Variable	Coefficient	Std. Error	t-Statistic	p-value
FDI	0.214	0.081	2.633	0.009
ORT	0.179	0.037	4.802	0.000
MERCT	0.032	0.032	1.002	0.317

4 DISCUSSION

The empirical findings indicate that foreign direct investment and oil rents are the main drivers of economic growth in GCC countries, while merchandise trade does not exert a statistically significant long-run effect. The PMG-ARDL estimates show that FDI has a relatively large positive coefficient, suggesting that foreign capital inflows contribute substantially to GDP per capita growth in GCC economies. The positive contribution of oil rents further reflects the continued reliance of GCC economies on hydrocarbon revenues as a primary source of income and public investment financing. This outcome is consistent with resource-based growth arguments, where natural resource revenues support infrastructure development, fiscal expansion, and public sector investment. The positive and statistically significant oil rent coefficient confirms that hydrocarbon revenues remain important to economic activity despite ongoing diversification policies across the GCC region. However, such dependence also exposes these economies to external shocks and fluctuations in global oil prices (Mahmood *et al.*, 2022). Consequently, periods of declining oil prices may weaken fiscal sustainability and reduce the capacity of governments to maintain growth--enhancing expenditures.

The positive and significant effect of foreign direct investment supports the view that external capital inflows enhance growth through technology transfer, productivity improvements, and integration into international markets. This finding aligns with the predictions of endogenous growth theory, which emphasizes the role of knowledge spillovers and capital accumulation in sustaining long-run growth. In the GCC, FDI may also contribute to ongoing diversification efforts by supporting the development of non-oil sectors (Kahouli and Chaaben, 2022). The magnitude and significance of the FDI coefficient further suggest that policies aimed at attracting high-quality and technology-intensive foreign investment could generate substantial long-run growth benefits for GCC economies.

In contrast, the insignificant effect of merchandise trade suggests that trade openness alone is not sufficient to drive growth in resource-dependent economies. This may be explained by the structure of GCC exports, which remain heavily concentrated in hydrocarbons with limited value addition. As a result, increases in trade volume do not necessarily lead to broader economic development or productivity gains. The insignificant trade coefficient implies that the benefits of trade may depend more on export composition and industrial diversification than on trade intensity itself. This finding indicates that the growth impact of trade depends on the composition and diversification of exports rather than trade openness alone.

The results suggest that economic growth in the GCC is primarily shaped by resource revenues and capital inflows, while the contribution of trade remains conditional on structural transformation. Policies aimed at enhancing diversification, promoting value-added exports, and strengthening the absorptive capacity of FDI are therefore essential for achieving sustainable long-run growth. In particular, GCC governments may strengthen long-run economic resilience by investing oil revenues in technology-driven industries, manufacturing, renewable energy, logistics, and human capital development to reduce excessive dependence on hydrocarbons.

5 CONCLUSIONS

This study examined the determinants of economic growth in selected GCC countries over the period 1995–2024, focusing on foreign direct investment, oil rents, and merchandise trade within a PMG-ARDL framework, with FMOLS estimates providing robustness. The findings show that oil rents and FDI are significant long-run drivers of per capita income growth, confirming the dominant role of resource revenues and external capital inflows in the GCC growth process. The empirical results indicate that FDI exerts a relatively stronger growth effect compared to the other explanatory variables, highlighting the importance of foreign capital inflows in supporting diversification and productivity improvements. In contrast, merchandise trade does not exert a statistically significant influence on long-run growth, indicating that trade expansion alone is insufficient in the absence of export diversification.

These results suggest that sustained economic growth in the GCC depends on effective management of resource revenues and the ability to attract productive foreign investment. While trade remains an important component of economic activity, its growth-enhancing role appears to depend on structural transformation toward more diversified and value-added exports. Policy efforts should therefore focus on reducing dependence on hydrocarbons, strengthening non-oil sectors, and improving the institutional and economic conditions necessary to maximize the benefits of FDI. The study therefore supports ongoing economic diversification strategies across GCC countries aimed at promoting private sector development, technological advancement, and broader economic transformation. Overall, the findings emphasize the importance of balancing resource-based growth with long-term diversification strategies to achieve sustainable economic development in GCC economies.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ALFARO, L., CHANDA, A., KALEMLI-OZCAN, S., SAYEK, S. 2004. FDI and economic growth: the role of local financial markets. *Journal of International Economics*. 64(1), 89-112. https://doi.org/10.1016/S0022-1996(03)00081-3

BORENSZTEIN, E., DE GREGORIO, J., LEE, J. W. 1998. How does foreign direct investment affect economic growth? *Journal of International Economics*. 45(1), 115-135. https://doi.org/10.1016/S0022-1996(97)00033-0

DUODU, E., BAIDOO, S. T. 2020. How does quality of institutions affect the impact of trade openness on economic growth of Ghana? *Cogent Economics & Finance*. 8(1), 1812258. https://doi.org/10.1080/23322039.2020.1812258

FRANKEL, A., ROMER, D. 1999. Does trade cause growth? *American Economic Review*. 89(3), 379-399.

KAHOULI, B., CHAABEN, N. 2022. Investigate the link among energy consumption, environmental pollution, foreign trade, foreign direct Investment, and economic growth: Empirical evidence from GCC countries. *Energy and Buildings*. 266, 112117. https://doi.org/10.1016/j.enbuild.2022.112117

MAHMOOD, H., ASADOV, A., TANVEER, M., FURQAN, M., YU, Z. 2022. Impact of oil price, economic growth and urbanization on CO2 emissions in GCC countries: asymmetry analysis. *Sustainability*. 14, 4562. https://doi.org/10.3390/su14084562

PESARAN, M., SHIN, Y., SMITH, R. 1999. Pooled mean group estimation of dynamic heterogeneous panels. *Journal of American Statistical Association*. 94, 621-631.

ROMER, M. 1990.Endogenous technological change. *Journal of Political Economy*. 98(5), S71-S102.

SACHS, D., WARNER, M. 1995. *Natural resource abundance and economic Growth*. NBER Working Paper No. 5398. https://doi.org/10.3386/w5398

SALEEM, H., SHABBIR, M., BILAL KHAN, M. 2020. The short-run and long-run dynamics among FDI, trade openness and economic growth: using a bootstrap ARDL test for co-integration in selected South Asian countries. *South Asian Journal of Business Studies*. 9(2), 279-295. https://doi.org/10.1108/SAJBS-07-2019-0124

SIRAG, A., SIDAHMED, S., ALI, H. 2018. Financial development, FDI and economic growth: evidence from Sudan. *International Journal of Social Economics*. 45(8), 236-1249. https://doi.org/10.1108/IJSE-10-2017-0476

SOLOW, R. 1956. A Contribution to the theory of economic growth. *The Quarterly Journal of Economics*. 70(1), 65–94. https://doi.org/10.2307/1884513

SWAN, W. 1956. Economic growth and capital accumulation. *Economic Record*. 32, 334-361. https://doi.org/10.1111/j.1475-4932.1956.tb00434.x

YIMER, A. 2022. The effects of FDI on economic growth in Africa. *The Journal of International Trade & Economic Development*. 32(1), 2–36. https://doi.org/10.1080/09638199.2022.2079709

Contact information

Evans Yeboah: e-mail: xyeboah1@mendelu.cz

FOMO EFFECT AND GENERATION Z'S SENSITIVITY TO MARKETING COMMUNICATION TOOLS IN THE CLOTHING INDUSTRY

Marek Záboj[1], Lenka Procházková[1], Karolína Malinová[1]

[1]Department of Marketing and Trade, Faculty of Business and Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_043

ABSTRACT

The paper examines the influence of the FOMO effect (Fear of Missing Out) on the consumer behaviour of Generation Z in the clothing industry. The aim is to identify factors influencing purchasing decisions with an emphasis on the role of FOMO and to formulate recommendations for marketing communication differentiated according to consumer types. Empirical data come from a quantitative questionnaire survey conducted in October 2025 ($n = 295$) among respondents born in 1995–2010. Using hierarchical cluster analysis (Ward's method, Euclidean distance), respondents were divided into four segments according to the degree of FOMO experienced. Six hypotheses were tested using multiple linear regression (H1–H3) and Spearman's correlation coefficient (H4–H6). The results show that FOMO is a statistically significant predictor of sensitivity to time limited offers and sales promotions ($p < 0.001$) and sensitivity to brand communication content ($p < 0.001$). At the same time, a positive association was confirmed between the level of FOMO, the intensity of social media use and the frequency of following influencers. On the other hand, FOMO did not appear to be a predictor of the frequency of online purchases or the consideration of sustainability. The results suggest differentiated recommendations for brand communication, which for individuals with high FOMO emphasizes urgency and visual dynamism, while for segments with low FOMO, credibility, transparency and value content.

Keywords: consumer behaviour, generation Z, FOMO effect, marketing communication, clothing industry, cluster analysis, consumer segmentation, social networks, limited time offer
JEL Code: M31, M37, D12

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

1 INTRODUCTION

Generation Z, defined in this study by the year of birth 1995–2010 (Suryadi *et al.*, 2025), represents a significant purchasing group whose consumption behaviour is shaped by constant contact with digital technologies and social networks. Young consumers respond to a rapid flow of information, visual content and short interactions with brand content, with the clothing industry being one of the segments in which their preferences are most pronounced. Studies focusing on the purchasing behaviour of this cohort in fashion confirm that Generation Z combines rational choice attributes (price, quality) with high sensitivity to visual stimuli on platforms such as Instagram and TikTok (Djafarova and Bowes, 2021; Wojdyla and Chi, 2024).

In the context of the digital environment, the psychological phenomenon known as FOMO – the fear of missing out on an experience, opportunity or social event – is gaining importance. This construct was psychometrically operationalized by Przybylsky *et al.* (2013) using the FoMOS scale, which has become the dominant measurement tool in subsequent research. Studies published in the last decade show that FOMO is associated with impulsive buying tendencies, higher sensitivity to marketing strategies that work with scarcity and exclusivity, and increased influence of influencers on purchasing decisions (Bekman, 2022; Dinh *et al.*, 2023; Oak and Markandeya, 2025).

In the clothing industry, FOMO is gaining particular importance. Rapidly changing trends, limited collections, visual inspiration on social media, and marketing communications emphasizing exclusivity create an environment in which the feeling of missing out can immediately turn into impulsive buying. Bläse *et al.* (2023) describe in this context that the FOMO effect supports unsustainable purchasing behaviour in the fast fashion segment, as it increases the pressure to react quickly and weakens the influence of value considerations. Pham *et al.* (2023) confirm that FOMO increases the intention to purchase trendy fashion products among young consumers on social media, Wu *et al.* (2021) experimentally demonstrate the effect of scarcity-based promotional strategies on impulsive buying in an online environment.

Empirical research on the impact of the FOMO effect in the Czech Generation Z remains limited in the professional literature. The Czech context has so far been described primarily from the perspective of the general purchasing behaviour of Generation Z (Kunc *et al.*, 2024), not focusing on the psychological factors of sensitivity to marketing stimuli. This contribution is trying to fill this gap. The aim of the study is to identify factors influencing the consumption behaviour of Generation Z in the clothing industry with an emphasis on the role of the FOMO effect, and based on the results, to formulate recommendations for brand marketing communication. It is assumed that the degree of FOMO is not homogeneous among consumers and that differences in the intensity of experiencing this phenomenon are reflected in different sensitivity to marketing tools, digital behaviour and the role of social networks in the purchasing decision-making process. To empirically verify these assumptions, six hypotheses (H1–H6) were formulated, which are specified in the following section.

2 METHODOLOGY AND DATA

The empirical part of the study is based on a quantitative questionnaire survey conducted online on the Google Forms platform. The questionnaire was intended for members of Generation Z born between 1995 and 2010 (Suryadi *et al.*, 2025); respondents outside this interval were excluded from the analysis by a filter question. Data collection took place between 17 and 26 October 2025, distributed primarily via social networks with subsequent dissemination by respondents. Due to this method of collection, the research sample cannot be considered representative of the entire population of Generation Z in the Czech Republic, which is a limitation of the study that must be considered when interpreting the results. Before starting the collection, a pilot study was conducted on a sample of five respondents to verify the comprehensibility of the items. The total number of valid observations was n = 295.

The questionnaire was structured into thematic blocks focusing on digital behaviour, frequency and method of clothing purchase, factors influencing choice, attitudes towards sustainability, influence of social environment, marketing stimuli and brand communication tools. The key component was the scale for measuring FOMO, translated from the original ten-item FoMOS scale by Przybylský *et al.* (2013). Most items were measured on a five-point Likert scale (1 = disagree / does not motivate, 5 = agree / motivates). The final part contained identification questions.

The internal consistency of the multi-item scales used was verified using Cronbach's α coefficient (Rabušic *et al.*, 2018), with $\alpha \geq 0.7$ being considered the minimum acceptable limit. A composite average score was then created for each scale, which was used in hypothesis testing.

Hierarchical cluster analysis was used to identify consumer segments, as it does not require a predetermined number of clusters and allows for the selection of a solution based on visual inspection of the dendrogram (Meloun and Militký, 2004). Euclidean distance was chosen as the distance metric, and Ward's method (Ward.D2) was used for the merging process, which minimizes intra-cluster variance. The input variables were ten items of the FoMOS scale. The calculations were performed in the Jamovi program.

Two statistical techniques were used to test the hypotheses. Hypotheses H1–H3, expressing the relationship between the level of FOMO and sensitivity to a specific type of marketing or product stimuli, were tested by multiple linear regression with control variables (age, gender, education) and an accompanying F-test of the overall model (Hendl, 2012). Hypotheses H4–H6, expressing the relationship between the level of FOMO and the frequency of a specific behaviour measured on an ordinal scale, were tested by Spearman's correlation coefficient ρ as a non-parametric measure of monotonic relationship (Rabušic *et al.*, 2018).

Hypotheses stated:

(1) H1: ***Higher levels of FOMO are associated with higher sensitivity to time-limited offers and sales promotions.***

(2) H2: ***Higher levels of FOMO are associated with higher sensitivity to brand communication content.***

(3) H3: ***Higher levels of FOMO are associated with stronger consideration of ethics and sustainability when purchasing clothing.***

(4) H4: ***Higher levels of FOMO are associated with more intensive use of social media in purchasing decisions.***

(5) H5: ***Higher levels of FOMO are associated with higher frequency of following influencers.***

(6) H6: ***Higher levels of FOMO are associated with higher frequency of online shopping.***

3 RESULTS

The research sample (n = 295) consisted mainly of women (75.3%), men represented 24.7%. The most numerous age groups were 24–26 years (42.7%) and 21–23 years (32.9%); 53.9% of respondents reported a university education. In terms of the size of the place of residence, large urban agglomerations with over 100,000 inhabitants prevailed (47.8%). Almost half of the sample came from the South Moravian Region (49.5%), which reflects the nature of the questionnaire distribution.

Descriptive results showed that Generation Z buys clothes mostly several times a year (43.4%) or about once a month (34.2%). E-shops (M = 3.59) and brick-and-mortar stores (M = 3.38) dominate among sales channels; purchasing via shopping buttons on social networks is among the least used channels (M = 1.64). When choosing clothes, the key attributes are the quality of the material (M = 3.96) and the price (M = 3.95), while sustainability and production ethics (M = 2.83), trendiness (M = 2.75) and brand (M = 2.54) play a secondary role. Purchasing decisions are most strongly influenced by the Instagram platform (74.2% of respondents) and brand websites (49.2%).

The internal consistency of the scales used was satisfactory to good: FOMO $\alpha = 0.814$, time-limited offers and sales promotion $\alpha = 0.786$, brand communication $\alpha = 0.744$, sustainability $\alpha = 0.716$. All values met the minimum threshold of $\alpha \geq 0.7$. Within the sustainability scale, the item related to personalized offers was removed to improve consistency, and the item "I would rather buy

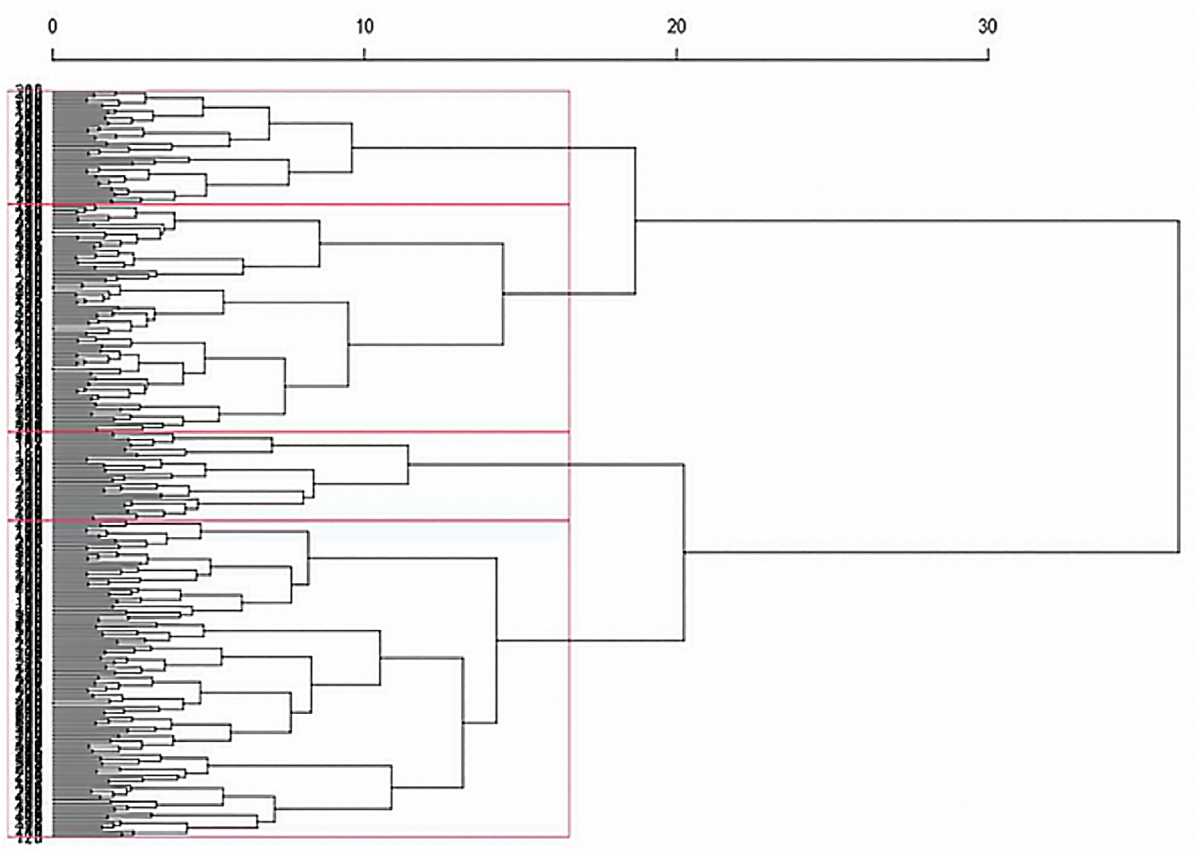


Fig. 1: Dendrogram of hierarchical cluster analysis with four segments marked according to the FOMO measure

trends than sustainable products" was reverse-coded. Hierarchical cluster analysis of the ten items of the FoMOS scale led to the division of respondents into four clusters based on visual inspection of the dendrogram (see Figure 1). Segment 1 (n = 35) represents individuals with a high level of FOMO – emotionally sensitive, socially impressionable and impulsively reacting to stimuli from the digital environment. Segment 2 (n = 125) includes individuals with a medium level of FOMO, characterized as socially engaged but rationally making decisions based on value alignment with the brand. Segment 3 (n = 45) represents individuals with a low to medium level of FOMO, whose decision-making is pragmatic and oriented towards price and quality. Segment 4 (n = 90) includes individuals with a low level of FOMO, who are value-based, plan purchases and place an above-average emphasis on sustainability.

The results of hypothesis testing are summarized in Table 1. Hypothesis H1 was confirmed: multiple regression showed a statistically significant model ($F(8, 286) = 6.83$; $p < 0.001$) explaining 16.0% of the variability in sensitivity to time-limited offers and sales promotion ($R^2 = 0.160$). FOMO turned out to be a strong positive predictor ($p < 0.001$). Of the control variables, only gender showed a statistically significant effect – women showed slightly higher sensitivity than men ($p = 0.030$). Hypothesis H2 was also confirmed: the model was statistically significant ($F(8, 286) = 4.38$; $p < 0.001$) and explained 10.9% of the variability in sensitivity to brand communication content ($R^2 = 0.109$). FOMO was the strongest predictor ($p < 0.001$), gender and the age group 18–20 also showed a significant effect. Hypothesis H3 was not confirmed: FOMO did not appear to be a significant predictor of considering ethics and sustainability when purchasing clothing; only gender showed a significant effect. Hypotheses H4 and H5 were confirmed: Spearman's coefficient showed a positive correlation between the level of FOMO and the intensity of social media use (H4) and the frequency of following influencers (H5). The strongest manifestation of this connection was recorded in the segment with high FOMO. Hypothesis H6, which assumes a higher frequency of online shopping in individuals with higher FOMO, was not confirmed; preferences for online and offline channels were similar across segments.

Tab. 1 Summary of hypothesis testing results

Hypothesis	Method	Key indicators	Result
H1:	Multiple linear regression	$F(8, 286) = 6,83$; $R^2 = 0,160$; FOMO $p < 0,001$	Confirmed
H2:	Multiple linear regression	$F(8, 286) = 4,38$; $R^2 = 0,109$; FOMO $p < 0,001$	Confirmed
H3:	Multiple linear regression	FOMO statistically insignificant; gender significant	Not confirmed
H4:	Spearman's ρ	Positive correlation; $\rho = 0.3009$; $p = 0.00001$	Confirmed
H5:	Spearman's ρ	Positive correlation; $\rho = 0.1969$; $p = 0.00067$	Confirmed
H6:	Spearman's ρ	Correlation statistically insignificant; $\rho = 0.1032$; $p = 0.07669$	Not confirmed

4 DISCUSSION AND CONCLUSIONS

The results of the empirical investigation confirm that the FOMO effect represents a significant psychological factor modulating the reactions of Generation Z to selected marketing communication tools, while in other areas – the choice of sales channel, sustainability orientation – its influence was not observed. These mixed findings are in line with the current literature. The importance of the mechanism of scarcity, exclusivity and time pressure for purchase intention is confirmed by the meta-analysis of Barton *et al.* (2022), specifically for FOMO in Generation Z in fast fashion by Oak and Markandeya (2025) and Pham *et al.* (2023). Wu *et al.* (2021) experimentally show that scarcity-based sales promotion directly increases impulse buying in an online environment, which corresponds to the positive relationship between FOMO and sensitivity to time constraints captured in hypothesis H1.

Confirmation of hypothesis H2 completes the framework in brand communication. Higher FOMO is associated with higher sensitivity to visually attractive, emotionally tuned and exclusivity-evoking messages. This result corresponds with the observation of Djafarová and Bowes (2021), according to which Instagram intensively shapes impulsive purchases in fashion among Generation Z, as well as with the work of Dinh *et al.* (2023), who identify FOMO as a moderator of the influence of influencers on purchase intention. Bekman (2022) states that communication practices, working with the feeling of missing out, generate higher engagement, and strengthen short-term purchase motivations.

The failure to confirm hypothesis H3 deserves more detailed discussion. The literature suggests that FOMO may weaken sustainable purchasing behaviour, as it promotes quick and impulsive reactions that contradict the consideration of the long-term consequences of consumption (Bläse *et al.*, 2023). However, in this research, FOMO did not emerge as a statistically significant predictor of sustainability-oriented behaviour; gender showed a more significant effect. The outcome of the segmentation analysis is interesting: the segment with the highest FOMO declared a relatively strong interest in sustainability and at the same time a preference for trendy products, which corresponds to the phenomenon of a discrepancy between declared values and actual behaviour. The segment with the lowest FOMO, on the other hand, attached the highest importance to sustainability and showed the highest level of purchase planning. In this context, Wojdyla and Chi (2024) state that Generation Z declares sustainability attitudes, but their translation into purchasing behaviour remains inconsistent.

Confirmation of hypotheses H4 and H5 shows that a higher level of FOMO is systematically linked to more intense movement in the digital environment and higher exposure to influencer content. Because of the cross-sectional design of the survey, it is not possible to decide on the direction of causality – it is likely a mutually reinforcing relationship, which is also pointed out by Dinh *et al.* (2023). On the contrary, failure to confirm hypothesis H6 shows that FOMO does not differentiate Generation Z in terms of the choice of shopping channel. This result corresponds to the observation of Kunc *et al.* (2024), according to which the shopping behaviour of Czech Generation Z is primarily omnichannel and searching for products online is naturally combined with shopping in brick-and-mortar stores.

From a marketing practice perspective, the results provide differentiated recommendations. For segments with higher FOMO, short visually dynamic videos, time-limited discounts, promo codes and communication of limited collections appear to be effective. For segments with medium FOMO, linking price incentives with brand authenticity, story campaigns and "behind the scenes" content is effective. Segments with low FOMO respond best to content focused on the quality of materials, transparency of origin, certification and production ethics; brand websites are a key verification channel for them.

The study's limitations lie mainly in the unrepresentativeness of the sample, the predominance of women (75.3%) and the regional concentration of respondents in the South Moravian region. Gender inequality could have influenced mainly the results regarding the sensitivity to communication content and sustainability, where gender turned out to be a significant variable. Furthermore, the cross-sectional design does not allow for a causal interpretation of the observed relationships. Future research suggests longitudinal monitoring of the stability of FOMO in the same consumer over time and intergenerational comparison with the emerging Alpha generation, which can be expected to have even more intense exposure to digital stimuli.

The main contribution of the study lies in demonstrating that the FOMO effect in the Czech Generation Z does not act as a universal consumption driver, but rather as a specific modulator of sensitivity to selected categories of marketing stimuli. The segmentation framework based on the intensity of FOMO experience allows for the formulation of an empirically based differentiation of the communication strategy of clothing industry brands.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

BARTON, B., ZLATEVSKA, N., OPPEWAL, H., 2022. Scarcity tactics in marketing: A meta-analysis of product scarcity effects on consumer purchase intentions. *Journal of Retailing*. 98, 741–758. https://doi.org/10.1016/j.ijresmar.2024.06.003

BEKMAN, M., 2022. The Effect of FoMO (Fear of Missing Out) on Purchasing Behavior in Public Relations Practices. *Journal of Selcuk Communication*. 15(2), 528–557. https://doi.org/10.18094/JOSC.1116808.

BLÄSE, R., FILSER, M., KRAUS, S., PUUMALAINEN, K., MOOG, P., 2023. *Non-sustainable buying behavior: How the fear of missing out drives purchase intentions in the fast fashion industry*. Business Strategy and the Environment. https://doi.org/10.1002/bse.3509.

DINH, T. C. T., WANG, M., LEE, Y., 2023. How does the fear of missing out moderate the effect of social media influencers on their followers' purchase intention? *SAGE Open*. July–September 2023, 1–13. https://doi.org/10.1177/21582440231197259.

DJAFAROVA, E., BOWES, T., 2021. ‚Instagram made me buy it': Generation Z impulse purchases in fashion industry. *Journal of Retailing and Consumer Services*. 59, 102345. https://doi.org/10.1016/j.jretconser.2020.102345

HENDL, J., 2012. *Přehled statistických metod: analýza a metaanalýza dat*. 4th ed. Praha: Portál. ISBN 978-80-262-0200-4.

KUNC, J., NOVOTNÁ, M., REICHEL, V., KŘÍŽAN, F., 2024. *Bricks or clicks? Factors influencing shopping behavior of Generation Z*. Equilibrium. https://doi.org/10.24136/eq.2999.

MELOUN, M., MILITKÝ, J., 2004. *Statistická analýza experimentálních dat*. Praha: Academia. ISBN 80-200-1254-0.

OAK, P. P., MARKANDEYA, A. A., 2025. FOMO and impulse buying: a behavioral study of Gen Z in the fashion market. *International Journal of Engineering and Management Research*. 15(3), 75–88. https://doi.org/10.5281/zenodo.15833671.

PHAM, T. T., NGUYỄN, T. G. H., DAO, T. K. L., 2023. The impact of the fear of missing out on purchasing trendy fashion products among young consumers on social media platforms. International *Journal of Advanced Multidisciplinary Research and Studies*. 3(5), 1240–1249.

PRZYBYLSKI, A. K., MURAYAMA, K., DEHAAN, C. R., GLADWELL, V., 2013. Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior*. 29(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

RABUŠIC, L., SOUKUP, P., MAREŠ, P., 2018. *Statistická analýza sociálně vědních dat*. Brno: Masarykova univerzita. ISBN 978-80-210-9247-1

SURYADI, N., FIRDAUSIAH, R. R. A., FASIEH, M. F. I., HAKIM, A., PARWATI, K. Y., 2025. Online consumption behavior of Gen Z in Indonesia post-COVID-19 pandemic: The role of digital technology. *General Management*. 26(207), 402–412. https://doi.org/10.47750/QAS/26.207.41

WOJDYLA, W., CHI, T., 2024. Decoding the fashion quotient: an empirical study of key factors influencing U.S. Generation Z's purchase intention toward fast fashion. *Sustainability*. 16, 5116. https://doi.org/10.3390/su16125116.

WU, Y., XIN, L., LI, D., YU, J., GUO, J., 2021. How does scarcity promotion lead to impulse purchase in the online market? A field experiment. I*nformation & Management*. 58(1), 103283. https://doi.org/10.1016/j.im.2020.103283.

Contact information

Marek Záboj: e-mail: marek.zaboj@mendelu.cz

Lenka Procházková: e-mail: lenka.prochazkova@mendelu.cz

Karolína Malinová: e-mail: xmalinov@seznam.cz

VISUALIZATION OF ALGORITHMS FOR GENERATING PSEUDORANDOM NUMBERS

Štěpán Zálešák[1], Radim Farana[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

DOI: https://doi.org/10.11118/ecos2026_044

ABSTRACT

This article deals with the design and implementation of an interactive web application intended for the visualization and comparison of two basic algorithms for generating pseudorandom numbers (PRNGs) — the Linear Congruential Generator (LCG) and the Linear Feedback Shift Register (LFSR). The work responds to the problem of the abstractness of teaching these algorithms and offers a tool that allows users to experiment with parameters, monitor the internal states of the generators, and analyze the quality of the resulting sequences using integrated analytical tools such as a histogram of value distribution, period detection, and a scatter plot to reveal the lattice structure.

Keywords: pseudorandom number, generator, algorithm visualization, web application
JEL Code: C63

1 INTRODUCTION

The generation of pseudorandom numbers is a critical part of many areas of computer science, including cryptography, computer simulations, the gaming industry, and machine learning. Unlike true random number generators (TRNGs), which utilize physical processes, pseudorandom generators (PRNGs) are deterministic algorithms. Based on an initial value (seed), they produce sequences that exhibit statistical properties of randomness, with their main advantages being reproducibility and high computational efficiency.

1.1 Linear Congruential Generator (LCG)

LCG is one of the oldest PRNG algorithms (Knuth 1998). It is defined by the recurrence relation:

$$x_i = (ax_{i-1} + c)(\mathrm{mod}\ m) \tag{1}$$

This article is licensed under CC BY-SA 4.0 Attribution-ShareAlike 4.0 International license.
https://creativecommons.org/licenses/by-sa/4.0/deed.en

where m is the modulus, a is the multiplier, and c is the increment, the choice of which affects the period and statistical quality of the generated sequence. Despite its simplicity, LCG suffers from inherent weaknesses, such as a lattice structure (n-tuples of values lying on hyperplanes) and a short period of lower bits for modules in the form of a power of two. From the literature, we know good LCG parameters for achieving a large period value and good statistical properties of the generated sequence of numbers (L'Ecuyer 1999).

1.2 Linear Feedback Shift Register (LFSR)

LFSR is based on bitwise operations. It consists of a register and a feedback function (usually XOR of selected bits — so-called taps). There are two basic variants: Fibonacci and Galois. The maximum period of an LFSR of length n is $2^n - 1$, which is achieved by using a primitive polynomial. LFSR is efficient, especially in hardware applications, but due to its linearity, it is not suitable for cryptography (Singer 2022).

2 APPLICATION DESCRIPTION

The resulting application is implemented using standard web technologies (HTML5, CSS3, JavaScript). The user interface is localized into Czech and English (Zálešák 2026).

2.1 Visualization Implementation

The LCG visualization displays the calculation process directly in the context of the mathematical formula, where variable values change dynamically in each step, respecting the standards for algorithm visualization (Naps *et al.* 2002). For LFSR, emphasis is placed on displaying the bit shift in the register and the feedback calculation, while the application supports both Fibonacci and Galois variants. The user has the option to step through the algorithm in both directions, even several times in a row, and thus correctly understand the behavior of both algorithms, see Fig. 1. After the user reaches the end of the algorithm, a statistical evaluation of the generated series of numbers is displayed, see Fig. 2 and Fig. 3.

2.2 Analytical Tools

To evaluate the quality of generators, the application contains:

- **Histogram:** Graphically represents the uniformity of the distribution of generated values, see Fig. 2. +
- **Period Detection:** Empirically determines the length of the period by generating values until the initial state is encountered, see Fig. 2.
- **Scatter Plot:** Displays pairs of consecutive numbers (xi, xi+1). This tool is key for visually detecting the lattice structure of LCG with an inappropriate choice of multiplier, see Fig. 3.

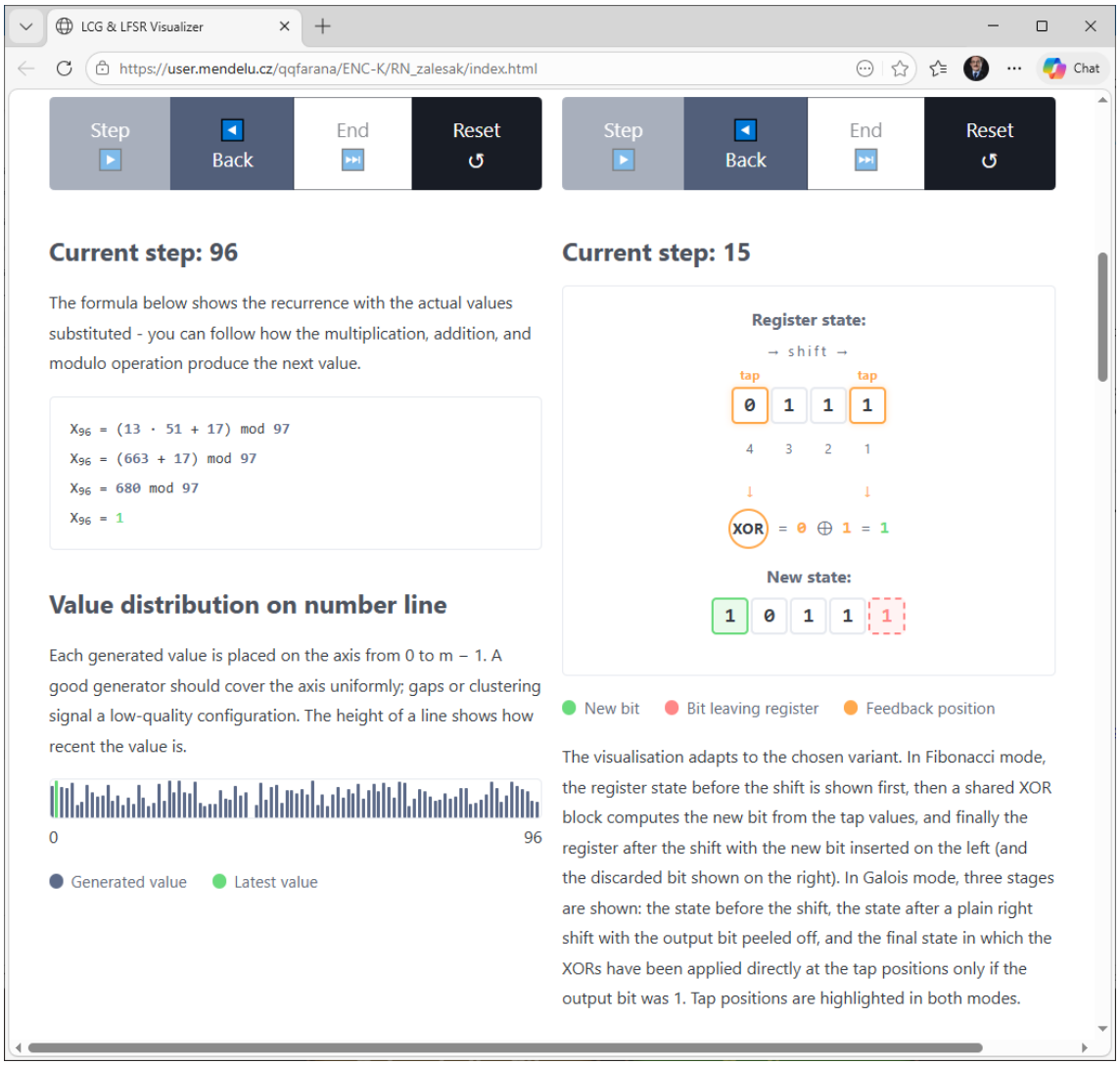


Fig. 1: Random numbers generation

3 TESTING RESULTS

The functionality and didactic benefit of the application were verified by user testing on a sample of computer science students. Testing took the form of a controlled scenario, after which respondents filled out a questionnaire focused on usability and clarity.

Tab. 1 Testing results

Evaluation Criterion	Average Rating (1–5, 5 is best)
Clarity of LCG visualization	4.7
Clarity of LFSR visualization	4.5
Contribution of analytical tools	4.8
Overall intuitiveness of the interface	4.6

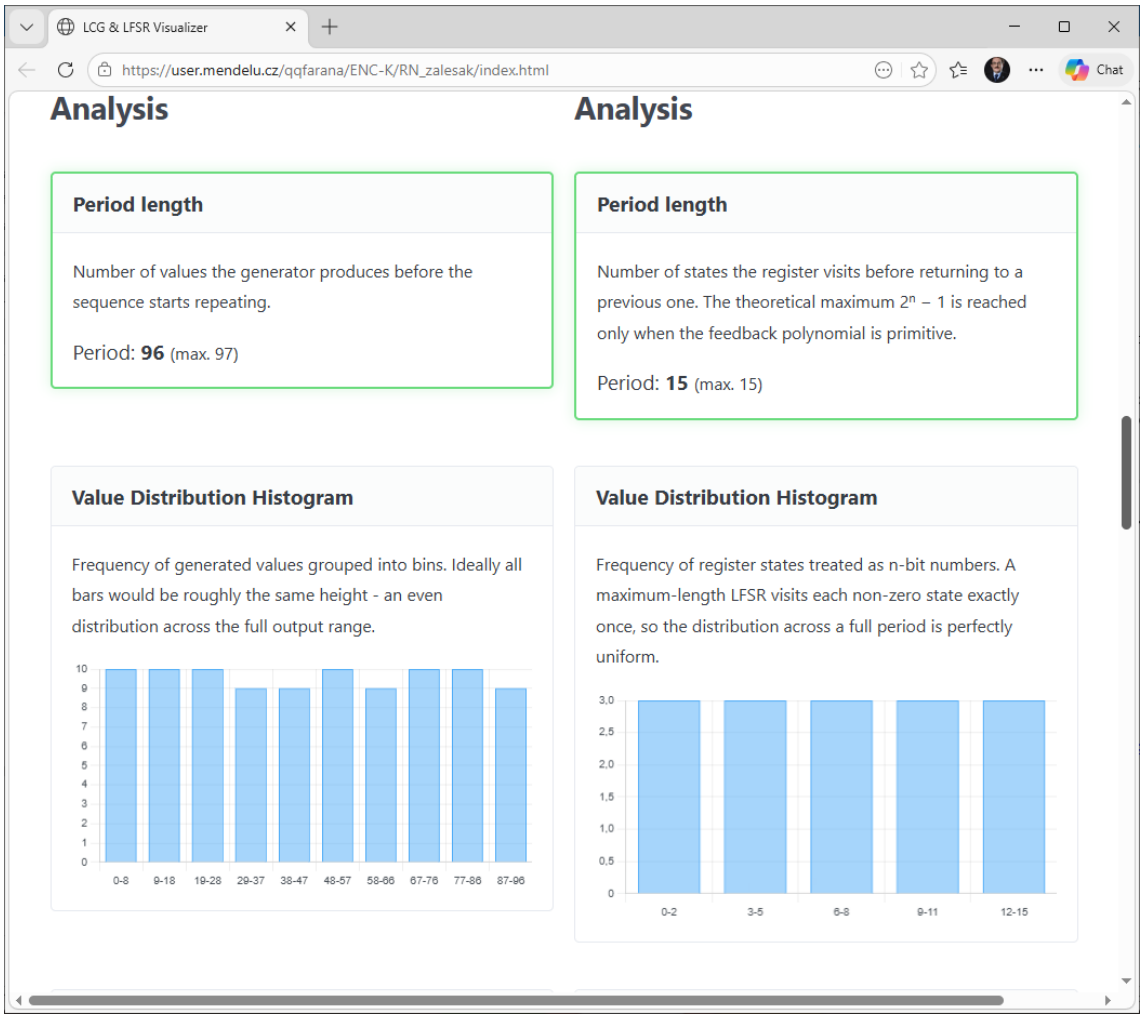


Fig. 2: Obtained period and histogram comparison

Users rated the period detection tool and the scatter plot most highly, as they allowed them to immediately see the influence of „bad“ parameters on the output quality. Identified problems included a higher difficulty in understanding the Galois variant of LFSR for beginners, which was subsequently addressed by adding context-sensitive help.

4 EVALUATION

The proposed application successfully demonstrates the differences in the principles and properties of LCG and LFSR. While LCG excels in integer arithmetic, LFSR offers efficient bitwise operations. Thanks to interactive comparison, users can observe that to achieve comparable quality and period, both algorithms require a specific, mathematically grounded choice of parameters. The application proved to be an effective didactic tool that supports the active involvement of students in the learning process.

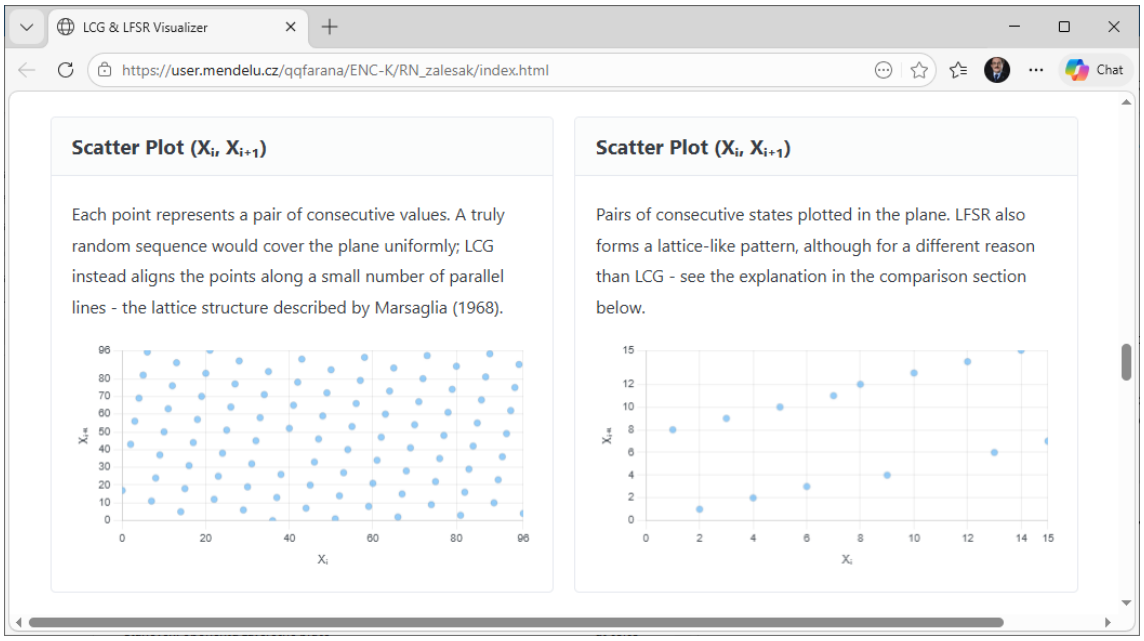


Fig. 3: Scatter plot for both generators

5 CONCLUSIONS

The created web application fulfills the set goals and provides a comprehensive environment for studying PRNG algorithms. The combination of step-by-step visualization and real-time statistical analysis represents an innovative approach to teaching this subject. Future expansions may include the implementation of other modern algorithms, such as Mersenne Twister, or the integration of more advanced statistical batteries of randomness tests. Thanks to its clarity, it can also help people who work with them and are not familiar with the algorithms used to understand the context associated with generating pseudorandom numbers.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-financed from the Operational Programme Research, Development and Education.

REFERENCES

HULL, T. E., DOBELL, A. R. 1962. Random Number Generators. *SIAM Review*. 4(3), 230-254. ISSN 0036-1445. https://doi.org/10.1137/1004061

KNUTH, D. E. 1998. *The Art of Computer Programming*. Volume 2: Seminumerical Algorithms. 3rd ed. Upper Saddle River, NJ: Addison-Wesley, c1998. ISBN 0-201-89684-2.

L'ECUYER, P. 1999. Tables of Linear Congruential Generators of Different Sizes and Good Lattice Structure. *Mathematics of Computation*. 68(225), 249–260. ISSN 1088-6842. https://doi.org/10.1090/S0025-5718-99-00996-5

NAPS, T. L. *et al*. 2002. Exploring the Role of Visualization and Engagement in Computer Science Education. In: *ITiCSE-WGR ,02: Working group reports from ITiCSE on Innovation and technology in computer science education*. New York: ACM, pp. 131-152. ISBN 978-1-4503-7449-1. https://doi.org/10.1145/960568.782998

SINGER, D. G. 2022. Secrets of Linear Feedback Shift Registers. *Mathematics Magazine*. 95(4), 368-375. ISSN 0025-570X. https://doi.org/10.1080/0025570x.2022.2094166

ZÁLEŠÁK, Š. 2026. *Web presentation and comparison of LCG and LFSR algorithms for generating pseudorandom numbers* [in Czech]. Bachelor's thesis, head: Farana, R. Brno: Mendel University in Brno.

Contact information

Štěpán Zálešák: e-mail: stepan.zalesak@outlook.com

Radim Farana: e-mail: radim.farana@mendelu.cz

28th ANNUAL INTERNATIONAL CONFERENCE

ECONOMIC COMPETITIVENESS AND SUSTAINABILITY 2026

PROCEEDINGS

Editors: Petr David, Hana Vránová

Publisher: Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
Edition: First, 2026

ISBN 978-80-7701-102-0 (online ; pdf)
https://doi.org/10.11118/ecos2026