



IPSI TIR
JOURNAL

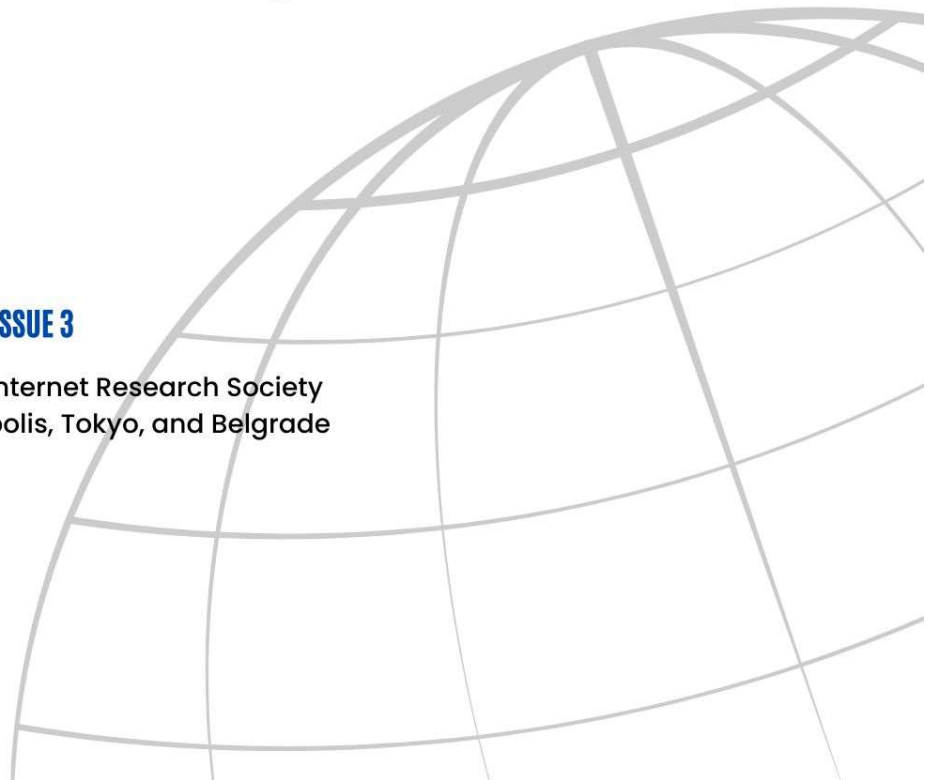
IPSI Transactions on Internet Research

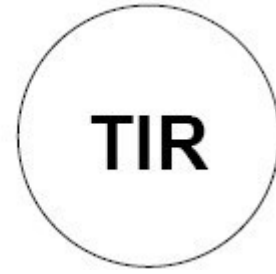
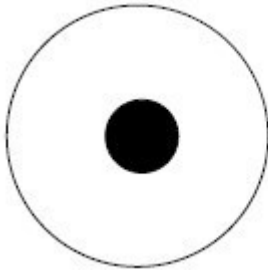
**Trustworthy Digital Transformation
and Data-Driven Societal Resilience**

Guest Editor: Dragan Vukmirović

JULY 2026, VOLUME 22, ISSUE 3

A publication of IPSI Internet Research Society
Petah Tikva, Indianapolis, Tokyo, and Belgrade
ISSN 1820-4503
DOI: 10.58245/ipsi.tir





The IPSI Transactions on Internet Research

Multi-, Inter-, and Trans-disciplinary Issues in Computer Science and Engineering

A publication of IPSI Internet Research Society Petah Tikva, Indianapolis, Tokyo, and Belgrade
July 2026 Volume 22 Number 3 (ISSN 1820-4503)

Trustworthy Digital Transformation and Data-Driven Societal Resilience

Guest Editor: Dragan Vukmirović

Table of Contents:

Editorial

Dragan Vukmirović1

Sustainable Digital Transformation Framework:

A Comprehensive Review of Internal and External Drivers

Fatahi Milasi, Sadegh; Benar, Nooshin; Salam, Maimoona; Radović-Marković, Mirjana;
and Farooq, Muhammad Shoaib4

Validation-Centric AI-Native Decision Systems:

Integrating the GAVA Loop with PRIME–INSPECT Governance

Avramović, Nebojša; Čomić, Tijana; Marković, Aleksandar;
Čavoški, Sava; Zornić, Nikola; and Vujović, Vladimir15

Forward-Secure Blockchain Architecture for National-Scale Electronic Voting:

Enhancing Institutional Sustainability and Democratic Resilience

Stanojević, Nebojša; Vukmirović, Dragan; Čomić, Tijana; Stanojević, Dejan;
and Četković, Miloš30

Generative AI in Audit Oversight: A Checklist-Based Model Evaluation

Kresović, Dejana; Drulović, Sofija; Nebojša, Đoković and Četković, Miloš37

Education and Cardiovascular Diseases: A Case Study of Patients at the Institute of Cardiovascular Diseases of Vojvodina

Kričković, Emina; Kričković, Zoran; Marković, Slobodan B.; Stojšić-Milosavljević,
Anastazija; Cvetković, Vladimir M.; and Lukić, Tin46

The IPSI Internet Research Society

The IPSI Internet Research Society is an association of people with professional interests in the field of the Internet.
All members will receive IPSI Transaction Journals upon registering at the Society.

Member copies of Transactions are for personal use only

IPSI TRANSACTIONS ON ADVANCED RESEARCH

www.ipsitransactions.org

Editorial Board

Veljko Milutinovic Co-Editor-in-Chief	Jakob Salom Co-Editor-in-Chief	Nenad Korolija Co-Editor-in-Chief
Department of Computer Science University of Indiana Bloomington Bloomington, Indiana, USA vm@etf.rs	Department of Computer Science Mathematical Institute of SANU Belgrade, Serbia ipsi.journals@gmail.com	IPSI Internet Research Society Dalmatinska 55 Belgrade, Serbia ipsi.journals@gmail.com
Radović Marković, Mirjana	Gonzalez, Victor	Milligan, Charles
Belgrade Bankers' Academy Belgrade, Serbia	University of Oviedo, Gijon, Spain	StorageTek, Colorado, USA
Filipović, Nenad	Daković, Nevena	Steingartner, William
Faculty of Engineering, University of Kragujevac, Serbia	Faculty of Dramatic Arts Belgrade, Serbia	Faculty of Electrical Engineering and Informatics, Košice, Slovakia
Marinković, Marko	Jutla, Dawn	Neuhold, Erich
Faculty of Civil Engineering, University of Belgrade Serbia	Sant Marry's University, Halifax, Canada	UNIWIE, Vienna, Austria
Domenici, Andrea	Karabeg, Dino	Piccardi, Massimo
University of Pisa, Pisa, Italy	Oslo University, Oslo, Norway	Sydney University of Technology, Sydney, Australia
Flynn, Michael	Kiong, Tan Kok	Miljanic, Scepán
Stanford University, Palo Alto, California, USA	National University of Singapore, Singapore	The School of Physical Chemistry, Belgrade, Serbia
Rishe, Naphtali David	Kovacevic, Branko	Rutledge, Chip
Florida International University Miami, USA	The School of Electrical Engineering, Belgrade, Serbia	Purdue Discovery Park, Indiana, USA
Ganascia, Jean-Luc	Patricelli, Frederic	Mester, Gyula
Paris University, Paris, France	ICTEK Worldwide, Pizzoli, L'Aquila, Italy	Óbuda University, Budapest, Hungary

Editorial

Trustworthy Digital Transformation and Data-Driven Societal Resilience

Guest Editor: Dragan Vukmirović

Special Issue:

Trustworthy Digital Transformation and Data-Driven Societal Resilience

Guest Editor Dragan Vukmirović

This issue's editorial focuses on two core topics: trusted digital transformation, data-driven social resilience.

The research papers featured in this issue of *IPSI Transactions on Internet Research* cover digital technology, artificial intelligence, blockchain architecture, statistical modeling, and data-driven methods. Though they belong to distinct research fields, all these works focus on technology's reshaping of organizations, institutions, and society and center on the shared core of trustworthiness, sustainability, auditability, social value, and strengthened resilience.

From a broader perspective, the importance of responsible digital transformation is growing increasingly prominent. We must correct the one-sided perception that frames digital transformation as nothing more than the deployment of new technologies. Beyond rolling out new technologies, all types of organizations and public institutions need to ensure their digital systems are interpretable, reliable, and resilient to meet social, ethical, and institutional expectations. In this sense, the papers collected in this special issue provide support for existing academic debates across five core areas: sustainable development, AI governance, audit and oversight, democratic infrastructure, and evidence-based public health.

This section sequentially sorts out the core content of four English academic papers, which respectively fall under the fields of sustainable digital transformation, AI-native decision governance, generative AI audit regulation, and national-scale electronic voting blockchain, to string together the research trajectory of the multi-scenario implementation of digital technologies:

The first paper, **Fatahi Milasi et al.'s Sustainable Digital Transformation Framework: A Comprehensive Review of Internal and External Drivers**, focuses on the sustainable digital transformation of small and medium-sized sports enterprises in developing countries. Using the qualitative meta-synthesis method, the study identifies 5 internal drivers, including leadership commitment, and 4 external drivers, including government support. It points out that this transformation is a multi-dimensional process that integrates economic, social, environmental, and ethical dimensions, rather than a simple process of technology adoption;

The paper, **Avramović et al.'s Validation-Centric AI-Native Decision Systems: Integrating the GAVA Loop with PRIME-INSPECT Governance**, centers on the core challenge of real-time output verification for AI-native decision-making. The study puts forward the GAVA (Generate-Act-Validate-Adjust) loop and integrates it with the PRIME-INSPECT governance framework. Through conceptual modeling combined with empirical analysis of samples of IT practitioners and executives, its core contribution is positioning verification as the primary organizational function of AI-native systems;

The third paper, **Stanojević et al.'s Forward-Secure Blockchain Architecture for National-Scale Electronic Voting: Enhancing Institutional Sustainability and Democratic Resilience**, investigates the feasibility of a permissioned blockchain architecture for national-scale electronic voting. The study proposes a model based on slot workflows and one-time key pairs. Prototype tests confirm that this architecture can support auditability, forward security, and predictable finality under controlled conditions. The core research of this paper centers on the design agenda of digital electoral infrastructure, contributing academic value to discussions focused on strengthening institutional trust and democratic resilience.

The paper, **Kresović et al.'s Generative AI in Audit Oversight: A Checklist-Based Model Evaluation**, explores the application of generative AI in the preliminary review of independent audit reports and their supporting financial statements. Based on a corpus of reports from listed companies in Serbia, the study compares the performance of ChatGPT and Gemini 3.1 Pro against the benchmark of professional audit assessment. It finds that generative AI can improve audit consistency and priority-setting capabilities in scenarios that rely on standardized checklists, but scenarios that require professional judgment still need verification by experts;

The fifth paper, cites a study by **Kričković et al: Education and Cardiovascular Diseases**, which took patients receiving treatment at the Institute of Cardiovascular Diseases in Vojvodina, northern Serbia, as its research subjects. The study employed statistical analysis, Poisson regression, and emerging hotspot analysis and found that groups with different educational levels have significantly different hospitalization patterns, that the incidence of cardiovascular diseases presents spatial clustering, and that integrating clinical, socioeconomic, and geospatial data can help interpret public health inequalities and support the formulation of targeted prevention strategies.

As editors of this special issue, we have synthesized the cross-domain research findings in this collection and propose that, beyond technical capabilities, digital transformation requires four categories of non-technical support: verification, governance, institutional trust, and social connectedness. Research spanning five major fields, including sports enterprises, converges on the same core conclusion: digital systems must operate as socio-technical systems and be embedded in organizational, institutional, and social contexts throughout their design and evaluation.

Prof. dr. Dragan Vukmirović is a full professor at the Faculty of Organizational Sciences, University of Belgrade, in the fields of statistics, data science, e-business, and digital transformation. He is a former Director General of the Statistical Office of the Republic of Serbia and has extensive experience in official statistics, statistical methodology, information systems, and evidence-based decision-making. He is the author and co-author of numerous scientific and professional papers and has participated in national and international projects. His research interests include statistics, data science, e-government, digital inclusion, artificial intelligence, and the application of generative AI in education, business, and public administration..

Sustainable Digital Transformation Framework: A Comprehensive Review of Internal and External Drivers

Fatahi Milasi, Sadegh; Benar, Nooshin; Salam, Maimoona; Radović-Marković, Mirjana; and Farooq, Muhammad Shoaib

Abstract: *The present study investigates sustainable digital transformation among sports small and medium-sized enterprises (SMSEs) in developing nations, where organizations must deal with unstable conditions and growing technological demands. The researchers used the qualitative meta-synthesis method of Sandelowski and Barroso (2007) to review previous studies and develop an integrated conceptual framework. The results revealed that internal factors such as leadership commitment, organizational culture, digital skills, technological infrastructure, and dynamic capabilities play a significant role in the success of digital transformation. External factors, such as government support, market competition, technological change and stakeholder expectations, play important roles as well. Sustainable digital transformation is not only about technology adoption, but also a mix of economic, social, environmental and ethical elements that lead towards resilient and data-driven sports organizations.*

Keywords: *Sustainable Digital Transformation, Digital Transformation Framework, Sports SMEs, Digital Capabilities, Organizational Sustainability*

1. INTRODUCTION

Two intertwined strands are shaping the modern world: the digital revolution and the growing demand for sustainability [1]. Technologies enabling computational powers and digital platforms are advancing rapidly [1]. At the same time, Industry 4.0 has produced a volatile, uncertain, complex, and ambiguous (VUCA) environment that has made organizations consider digital transformation as more than a technology choice [2]. It is now a matter of strategic necessity for survival, for remaining competitive and for building long-term value [1, 2]. Here, the notion of “Sustainable Digital Transformation” (SDT) has been introduced as a broad and still evolving paradigm in management and business literature [3]. SDT is a strategic, multidimensional and ethics-driven process [3]. Organizations utilize digital technologies to re-engineer business models to increase economic value, reduce environmental and social impacts,

and respect principles such as protection of privacy, algorithmic fairness and human responsibility [4]. That’s how SDT surpasses basic digitalization [4]. It adds the dimension of “ethical sustainability” to the traditional Triple Bottom Line framework and provides a more integrated and responsible view on how organizations develop [4, 5].

The definition of sustainability has also changed [5]. It now includes economic and social aspects besides environmental concerns [6, 7]. Organizations are being held accountable for Sustainable Development Goals (SDGs) [8]. Digitalization was deeply accelerated by the COVID-19 pandemic [7, 8]. It has pushed the use of digital technologies from a competitive advantage to something almost unavoidable [8]. Thus, SDT is a new and more attractive way for supporting economic growth with less negative social and environmental impacts [9, 10].

These developments influence global sports industry, which is one of the biggest and most active sectors in the world [10]. Already the value has exceeded USD 500 billion and it is expected to keep rising [11, 12]. The industry extends well beyond professional sports [11]. It includes the fields of sports equipment manufacturing, recreational services, sports tourism, event management and related entertainment [12]. All these activities have a real economic, social and cultural importance [13]. Sports is about experiences, about being connected with fans all the time and about the massive use of data and media [14]. These features have made the digital technologies more important in the sector. Moreover, a high amount of investment in sports technology start-ups also indicates the growing importance of this field [14].

Small and medium-sized sports enterprises (SMSEs) represent a significant portion of the sports industry [14]. Their flexibility, agility, lower research and development costs and stronger relationships with local communities have obvious potential for innovation and digital transformation [15]. In developing countries these enterprises boost job creation, economic development and progress on sustainable development goals [16]. But there are many

hurdles in the way of sustainable digital transformation [16, 17]. The lack of capital and digital infrastructure, the lack of digital skills, organization resistance to change, the costs of technology adoption and the lack of consistent supportive policies impede [18-19]. The ethical dimension of digital transformation is one of the most pressing issues today, because of concerns about privacy breaches, algorithmic bias, disappearance of traditional jobs and increasing social inequalities [20, 21]. In the developing world, these conditions increase the risk of failure in digital transformation projects [18]. So many companies either don't even start down this path, or run into problems in the early stages [21].

Although, many studies addressed digital transformation and sustainability as distinct issues [22, 23]. Only a limited amount of work combines them in the digital transformation and business perspectives [22, 23]. Second, the ethical dimension of digital transformation is relatively ignored [22]. These include privacy, algorithmic fairness, implications for the workforce, and social inequalities [20, 21]. Third, the sports industry and especially SMSEs in developing countries are under-represented in the extant literature while these enterprises face unique structural, cultural and institutional limitations [24]. Fourth, there is a lack of extensive qualitative meta-syntheses that bring together scattered findings and offer a conceptual and operational framework for the field [25, 26].

Therefore, this study uses the qualitative meta-synthesis approach proposed by [27] to develop a conceptual framework for the analysis of sustainable digital transformation in small and medium-sized sports enterprises in developing countries [27]. It studies the factors affecting success or failure from the two perspectives of "internal factors" and "external factors" [28, 29]. Moreover, stakeholder theory also supports a multilevel analysis of the roles of athletes, fans, employees, local communities, and policymakers in creating sustainable value [30]. Theoretically, the paper adds to the literature of sustainable digital transformation in sports by aggregating fragmented qualitative results and providing a conceptual framework for future studies [30]. From a practical perspective, the results can help managers and policymakers to identify barriers, recognize the necessary capabilities, and choose effective strategies to carry out a successful sustainable digital transformation [30, 31].

2. LITERATURE REVIEW

2.1 Digital Transformation

Digital Transformation (DT) is a strategic, multi-layered and iterative process, where digital technologies are used not only to improve operational efficiency, but to significantly redefine business models, value propositions, customer experiences, organizational culture and broader stakeholder ecosystems [2, 31]. Within the VUCA dynamics and fast developments of Industry 4.0, digital transformation is much more than a discretionary initiative, it is related to the survival of the organization and its long-term competitiveness [1]. DT is not equivalent to simple digitization or digitalization as it involves deeper changes in organizational identity, strategic orientation, structural arrangements and inter-organizational relationships [32]. This distinction is important, since many organizations still mistakenly view automation alone as evidence of true business model reinvention [33].

The major enabling technologies for DT are also evolving our everyday mobile applications [34, 35]. However, the literature points out repeatedly that technology is an enabling mechanism in and of itself only [34]. The success/failure of digital transformation is highly contingent on the degree to which technological initiatives are aligned with organizational conditions, human capacities and environmental pressures [34, 35]. But technology-organization-environment (TOE) model and other multilevel DT frameworks emphasize the role of top management commitment, organizational culture, and employee capabilities to influence transformation outcomes [28, 29]. In particular, digital transformation in SMEs is often related to a so-called "resource paradox" as termed by researchers [36]. Monetary limitations, lack of practical capability and inadequate human resources can lead to project failure rates between 70% and 90% [36]. However, SMEs also have features that may help to adapt more rapidly, such as flexibility, informal internal structures, and closer ties with customers, all of which can facilitate agile forms of innovation [37].

Study has recognized a number of aspects that have a strong influence on successful DT implementation in SMEs [37]. These include active support from top management, a learning-oriented culture, continuous development of digital competencies and collaboration with external actors and partners [38]. These conditions are even more important for small and medium-sized sports enterprises, which are often characterized by deficiencies in digital capabilities and technological expertise [15]. IOT helps in monitoring of athlete performance, injury

prevention and optimization of game strategies [7, 39]. Technologies and digital gadgets including wearable sensors, AI-powered analytics systems, fan engagement platforms, and customized mobile apps are increasingly being adopted by sport enterprises of small and medium size [40]. The sports industry is highly dynamic and experience-oriented [40]. Therefore, it gradually became a “living laboratory” for DT in the SME sector as described by some scholars [41].

2.2 Sustainability and Digital Transformation

In general, sustainability is understood as the capability of creating long-term economic value, while at the same time meeting social, environmental and ethical obligations [42]. The concept is heavily built on the Triple Bottom Line framework with its core components of economic, social and environmental dimensions [6, 42]. Recently, the SDGs have emerged as the dominant global framework to operationalize sustainability principles at the organizational level [40, 43]. Meanwhile, the Fourth and Fifth Industrial Revolutions have led to a greater focus in academic and managerial discussions on a fourth dimension of sustainability: ethical and human-centered sustainability [41, 44]. This dimension deals with issues such as algorithmic fairness, privacy and data protection, technology-related unemployment and the reduction of digital inequality among different groups and societies [43, 44].

In this sense, digital transformation can be both a value creation mechanism and an ethical risk source [45]. Organizations thus increasingly require a human-centered and systemic approach that can handle this tension in a balanced manner [20]. International organizations like International Olympic Committee, FIFA and UEFA have already established clear targets concerning carbon neutrality, social justice, transparency and governance standards [46]. In this regard, the intersection of DT and sustainability through the notion of “Sustainable Digital Transformation” (SDT) has started to appear as a rather new paradigm in business management research and practice [45, 46]. From this perspective, digital technologies are seen as enabling instruments supporting at the same time economic efficiency, social legitimacy and environmental protection [4]. The relevance of this approach and SDT could open up opportunities to “leapfrog” to greener, more inclusive and more human-centered business models with lower levels of investment and enabling faster organizational adaptations [47].

3. METHODOLOGY

This study applied a meta-synthesis approach grounded in the seven-stage model introduced by [27] in order to systematically combine and interpret findings from previous studies and, subsequently, construct a conceptual framework for digital maturity in small and medium-sized sports enterprises [45, 47]. The central aim of the research was to clarify and describe how internal and external factors connected to digital transformation in SMEs are conceptualized, particularly in relation to sustainability-related considerations and organizational adaptation [48].

From the standpoint of research purpose, the study can be categorized as applied-developmental [49]. Methodologically, because the research relied on secondary sources and did not involve direct intervention within actual organizational environments, it is considered descriptive, non-experimental, and qualitative in character [50]. The meta-synthesis procedure followed the seven-stage framework and included these sequential steps: (1) formulation of the research question, (2) conducting a systematic review of the literature, (3) selection and evaluation of relevant sources, (4) extraction of data and conceptual information, (5) analysis and synthesis of the identified concepts, (6) quality assessment and control of the analytical process, and finally, (7) presentation and development of the final conceptual model [40, 50].

4. FINDINGS

The main research questions guiding this study were formulated as follows: Which dimensions are linked to DT in the sports enterprises when examined from a sustainability perspective? And, what specific components constitute each dimension of sustainable digital transformation within sports enterprises? To answer these questions, the research population included scholarly publications and academic studies published between 2020 and 2026 that were directly related to the topic under investigation. Data collection was conducted through systematic searches in the databases Emerald, SpringerLink, Scopus, and Web of Science. For the retrieval process, different combinations of following keywords were used:

Table 1. Keywords

"Digital Transformation" AND "SMEs"
"Sustainable Digital Transformation" AND "SMEs"
"Digital Transformation" AND "Sustainability" AND "Sports Industry"
"Digital Technologies" AND "Sports Industry" AND "Developing Countries"
"Sustainable Digital Transformation" AND "Sports

Industry"
"Digital Transformation" AND "SMEs" AND "Developing Countries"
<i>Source: Authors</i>

The initial database search resulted in the identification of 405 articles. The screening and selection procedure was then carried out in several stages. During this process, duplicate

records, irrelevant studies, articles lacking sufficient scientific quality, and publications falling outside the defined language and time-period criteria were excluded from further analysis. After completing these screening stages, a smaller group of articles was retained for detailed and in-depth review.

Table 2. Criteria for Acceptance or Rejection of Articles

Acceptance Criteria	Rejection Criteria
Language: English	Languages other than English
Time Frame: 2020 to 2026	Prior to 2020
Study Type: Peer-reviewed journal or conference articles	Personal reports, websites, theses
Methodology: Studies with scientific approach (qualitative or quantitative)	Experimental studies lacking scientific basis
Thematic Focus: Digital transformation in SMEs	Other domains or large enterprises

Source: Authors

Each selected article was assessed using ten separate criteria, with scores assigned on a 1 to 5 scale for every criterion. These criteria covered a range of methodological and reporting aspects, and the overall contribution or value of the study. Based on this assessment process, only articles that scored an average of above 30 (out of a

possible total of 50) were considered eligible to move forward to the final stage of the analysis. Articles below this cut-off were excluded from further consideration. An example of the completed quality assessments is provided below.

Table 3. CASP Assessment of Selected Studies

Article	Research Aims	Methodology Rationale	Research Design	Sampling	Data Collection	Reflexivity	Ethical Considerations	Analysis Rigor	Clarity of Findings	Research Value	Total Score
Article 1	5	3	4	4	4	5	5	4	4	5	43
Article 2	5	5	5	4	4	5	5	5	5	5	49
Article 3	4	5	3	5	4	5	5	5	4	5	45
Article 4	4	5	4	5	5	3	4	4	5	4	43
Article 5	5	4	5	4	5	3	5	4	5	4	44
Article 6	4	5	5	5	4	4	4	5	4	5	45
Article 7	4	5	4	5	5	5	3	5	3	3	42
Article 8	4	4	4	5	5	3	4	5	5	4	46
Article 9	5	4	5	4	5	5	3	5	5	3	44
Article 10	4	5	3	5	4	5	5	5	4	5	45

Source: Authors

Following the CASP quality assessment, 22 studies were excluded from the dataset. Finally,

at the end of the whole screening and appraisal process 39 articles were retained for the final stage of detailed analysis. **Figure 1** gives an overview of the whole process and its results.

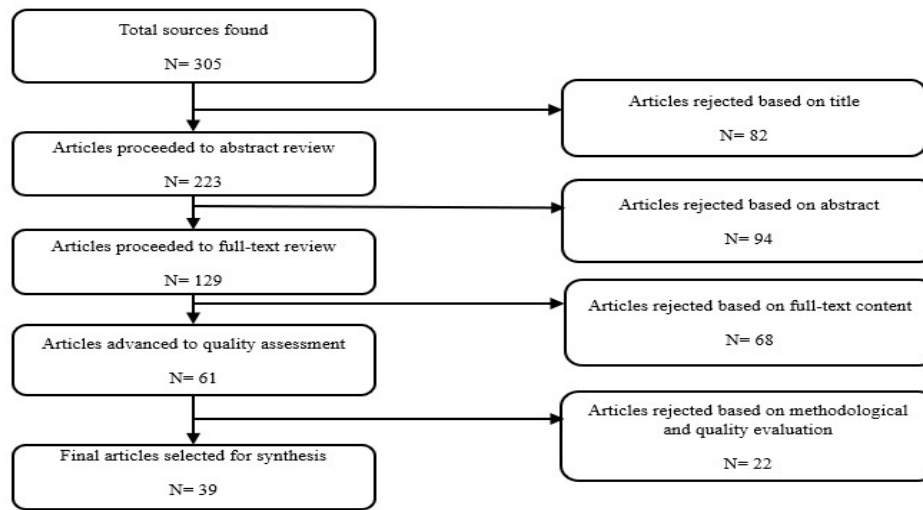


Figure 1. Search Results and Selection of Suitable Articles
Source: Authors

Thematic Analysis and Extraction of Indicators

Data analysis employed thematic analysis, involving multistage coding. The selected articles were read multiple times to extract key themes and indicators. This method was deemed appropriate for meta-synthesis due to its capacity to identify hidden patterns in qualitative data and systematically integrate findings [27]. Thematic

analysis identified dimensions and indicators related to internal and external factors influencing sustainable digital transformation in small and medium-sized sports enterprises [2]. The table below presents a comprehensive classification of the extracted dimensions and indicators, developed through thematic analysis of articles addressing digital transformation and sustainability in the sports industry.

Table 4. Thematic Synthesis of Factors Influencing Sustainable Digital Transformation

Main Category	Subcategory	Key Codes (Indicators)	Brief Description	Reference
Internal Factors	Senior Management Leadership and Commitment	Transformational leadership, top management commitment, creation of digital vision, digital governance, strong executive support, ethical leadership, data-driven strategic decision-making	Main driver of direction-setting, resource allocation, and sustainable digital transformation leadership	[28, 31, 48, 49]
	Organizational Culture	Open and innovative culture, employee technology acceptance, flexibility toward digital changes, strong internal communication, continuous digital learning culture, agile and data-driven culture, organizational readiness for transformation	Main foundation for the adoption, institutionalization, and maturity enhancement of sustainable digital transformation	[2, 36, 50, 51]
	Human Resources and Digital Competencies	Employees' digital skills and competencies, training programs, digital talent acquisition, employee empowerment, employee participation in digital innovation, data and AI literacy, employee digital well-being and resilience	Human capital as the most important driver for moving from digitization toward sustainable transformation	[17, 38, 52-54]
	Financial Resources and Internal Budget	Budget limitations, investment in technology, digital training budget, Digital ROI analysis, green investment, technology investment risk	Financial limitations as one of the major internal barriers to transformation in investment	[52, 55]

	management	SMEs
Internal Capabilities and Infrastructure	Access to IoT, AI, and Cloud technologies, technology-process integration, investment in emerging technologies, infrastructure, platform incorporation capability, cybersecurity	Technical and data [3, 56, 57] infrastructure required for implementing digital transformation and ethical sustainability
	Dynamic Capabilities and Organizational Agility	Continuous learning, resource reconfiguration, organizational agility, strategic flexibility, continuous innovation, rapid adaptation to the VUCA environment [2, 36, 37, 58]
	Government Policies and Support	Government financial and regulatory support, incentive policies for SMEs, digital subsidies and facilities, national digital transformation frameworks, policies supporting green innovation and digital ethics [52, 59, 60]
	Competitive and Market Pressure	Intense competition, changing customer behavior, demand for digital services, competition grounded on digital fan involvement, marketplace stress for transparency and digital responsibility [8, 61]
	Technological Advancements and Accessibility	Advancements in IoT, AI, and AR, technology cost and accessibility, national digital infrastructure, digital innovation ecosystem, cloud platforms, green and energy-efficient technologies, data-driven transformation [48, 62]
External Factors	Stakeholder Expectations	Expectations of fans, sponsors, athletes, and policymakers, demands for digital transparency and accountability, digital trust, stakeholder participation in value co-creation [61, 63]
	Economic Sustainability	Reduction of operational costs, new revenue models, financial transparency, long-term growth, digital economic resilience, sustainable competitive advantage, data-driven productivity [6, 28, 63, 64]
	Social Sustainability	Digital inclusion, gender equality, access for disadvantaged communities, digital fan engagement, enhancement of social capital, improved stakeholder experience, digital social welfare [45, 50, 54, 65]
	Environmental Sustainability	Carbon footprint reduction, energy optimization, smart waste management, Green IT, low-energy digital technologies, reduced resource consumption, development of green sports operations [44, 50, 66, 67]
	Ethical Sustainability	Data protection, responsible use of AI, digital transparency and fairness, ethical governance of data and sustainable digital [20, 21, 68, 69, 70]

		algorithms, algorithmic transparency, transformation Explainable AI, digital trust, prevention of algorithmic bias, AI accountability	
Digital Transform- ation	Digital Transformation Strategy	Process digitization, business model transformation, data-driven decision-making, Digital Maturity Model (DMM), digital transformation alignment between DT and sustainability, integration of sustainability goals with digital strategy, sustainable transformation governance	Roadmap and governance [23, 26, 71, mechanism for sustainable 72]
	Applied Digital Technologies	Sports mobile applications, online platforms, IoT and wearable devices, artificial intelligence and data analytics, implementing sustainable automation, fan engagement platforms, digital transformation in the digital twins, big data analytics, sports industry intelligent fan experience systems	Practical tools and [40, 53, 61]

Source: Authors

Thematic analysis identified dimensions and indicators associated with DT [27]. These dimensions and indicators function as a

groundwork for qualitative synthesis and the progress of an integrated framework (see **Figure 2**) to guide future research and managerial decision-making [2, 27].

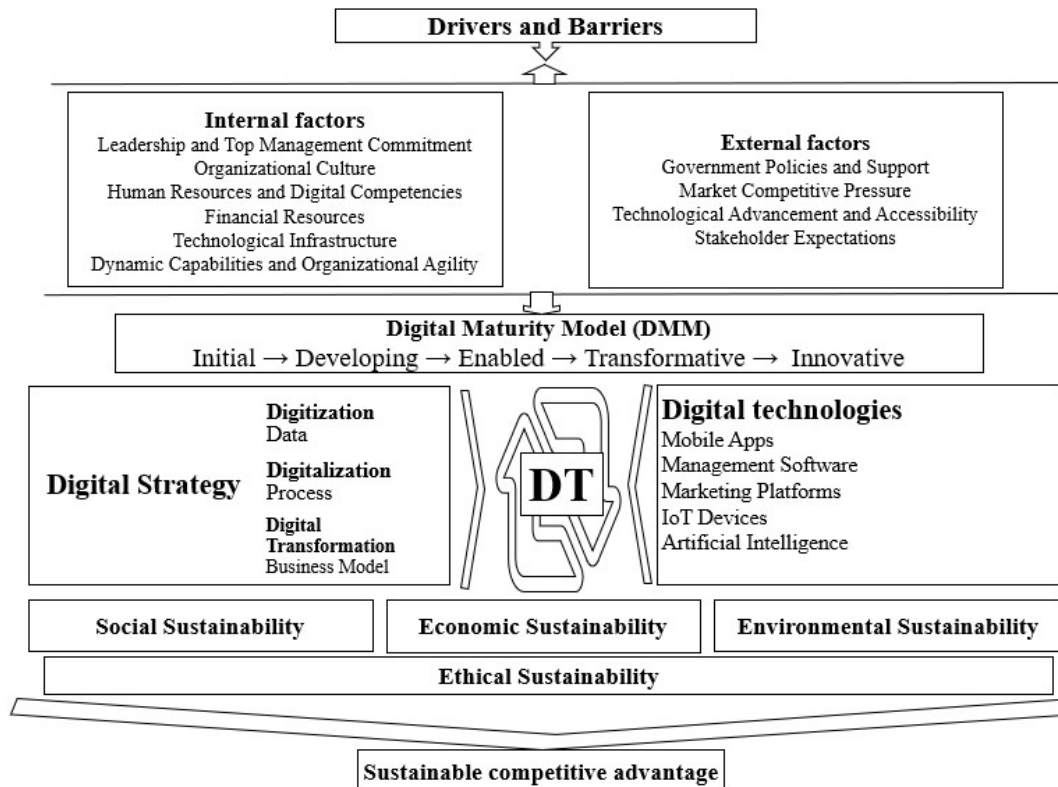


Figure 2. Conceptual Framework for Sustainable Digital Transformation
Source: Authors

5. DISCUSSION

The strategic linking of digital transformation with sustainability is no longer an optional competitive edge in today's complex, fast-moving, technology-shaped environment in which

the VUCA [2, 22] conditions and the demands that come from the Fourth and Fifth Industrial Revolutions are constantly in play [52, 55]. Rather, it is a necessary ingredient for organizational survival, legitimacy, and resilience [2, 28]. The results of this meta-synthesis suggest that DT in SMSEs is not only driven by

new technologies but is a multi-layered and dynamic process that depends on continuous interaction among human, organizational, technological and institutional conditions [58, 59].

A closer look at the results suggests a kind of “systemic interdependence” between internal and external drivers of sustainable digital transformation [66]. Yes, the main driving forces are top management commitment, transformational leadership and having a clear digital vision [73] but their actual effectiveness is heavily shaped by organizational culture, employee readiness, organizational learning processes and how open the organization is to change [74]. In this sense, organizational culture should not be considered only as a background factor [72, 74]. It is much more a living capability that determines how well a firm can adapt to digital changes [74]. Thus, if an organization has no innovation values, trust and continuous learning habits, it is likely to fail in its digital transformation, even if there is external support or infrastructure [75].

In addition, the market competition, changing consumer behavior, and increased fan expectations for more personalized digital experiences [76] are compelling sports enterprises to rethink their digital governance structures, data systems, and the overall logic of value creation [77]. In this case it is more about a mechanism to strengthen digital engagement, boost fan participation and develop new revenue pathways that were not really possible before, than a technical upgrade [78]. The analysis also points to what can be called a “resource paradox” in SMSEs [79]. Financial constraints, infrastructure gaps and shortages in digital skills increase the risk of failure in digital transformation projects, with failure rates in many cases often being estimated between 70% and 90% [36]. But on the flip side, the same organizations that are being more flexible, informal and less rigid are often better equipped for rapid innovation and faster technological adaptation [80].

6. CONCLUSION

DT in SMSEs is much more than an IT change [78]. It is rather a deeper modification of the organizational identity, logic of value creation and of the stakeholders' interaction [78]. The results of this study show that a successful sustainable digital transformation in developing countries is dependent on a kind of synergy between internal dynamic capabilities and external institutional and infrastructural support [7, 78]. The study further highlights the complex interplay of factors including transformational leadership, learning-oriented organizational culture, digital skills, data governance structures, policy support, and competitive market pressure towards DT [78]. Organizations that can balance digital innovation

with economic sustainability, social responsibility and data-driven ethics will likely be stronger going forward [78, 80]. The use of such an integrated framework can harness digital technologies to promote sustainability, social trust, and digital transformation.

GENERATIVE AI STATEMENT

The authors confirm that the generative artificial intelligence tool Grok (developed by xAI) was used solely for improving written expression, enhancing readability, and refining the fluency of the text. The prompts provided to this tool were limited to sentence rewriting and improvements in clarity and coherence. All outputs generated by the artificial intelligence were extensively reviewed, edited, and integrated by the authors.

REFERENCES

- [1] N. Bennett and G. J. Lemoine, "What a difference a word makes: Understanding threats to performance in a VUCA world," *Business horizons*, vol. 57, no. 3, pp. 311-317, 2014, doi: <https://doi.org/10.1016/j.bushor.2014.01.001>.
- [2] G. Vial, "Understanding digital transformation: A review and a research agenda," *Managing digital transformation*, pp. 13-66, 2021, doi: <https://doi.org/10.1016/j.jsis.2019.01.003>.
- [3] S. F. Milasi, N. Benar, A. Nazarian, and M. Shahzad, "Unlocking the potential: A comprehensive meta-synthesis of Internet of Things in the sports industry," *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, vol. 240, no. 1, pp. 238-253, 2026, doi: <https://doi.org/10.1177/17543371241229521>.
- [4] A. M. Gomez-Trujillo and M. A. Gonzalez-Perez, "Digital transformation as a strategy to reach sustainability," *Smart and Sustainable Built Environment*, vol. 11, no. 4, pp. 1137-1162, 2022.
- [5] C. Becker, *Sustainability ethics and sustainability research*. Springer Science & Business Media, 2011.
- [6] A. Nazarian, M. Shahzad, X. Ding, and A. Appolloni, "Do TQM Instigate Sustainable Development: Identifying the Key Role of Green Innovation and Knowledge Management," *Journal of the Knowledge Economy*, pp. 1-26, 2023, doi: <https://doi.org/10.1007/s13132-023-01256-2>.
- [7] S. Fatahimilasi, N. Benar, and A. Nazarian, "Sustainable performance of sports business: The role of technological intelligence and the mediating effects of organizational agility and customer agility capabilities," *Sports Marketing Studies*, vol. 4, no. 4, pp. 1-18, 2024, doi: <https://doi.org/10.22034/sms.2024.140139.1271>.
- [8] N. Dobrovolskienė, A. Pozniak, and M. Tvaronavičienė, "Assessment of the sustainability of a real estate project using multi-criteria decision making," *Sustainability*, vol. 13, no. 8, p. 4352, 2021, doi: <https://doi.org/10.3390/su13084352>.
- [9] P. M. Bican and A. Brem, "Digital business model, digital transformation, digital entrepreneurship: Is there a sustainable “digital”?," *Sustainability*, vol. 12, no. 13, p. 5239, 2020, doi: <https://doi.org/10.3390/su12135239>.
- [10] T. Stock and G. Seliger, "Opportunities of sustainable manufacturing in industry 4.0," *procedia CIRP*, vol. 40, pp. 536-541, 2016, doi: <https://doi.org/10.1016/j.procir.2016.01.129>.
- [11] B. Basu, "Perspectives on the Intersection Between Sports and Technology," in *Sports Management in an Uncertain Environment*: Springer, 2023, pp. 143-168.

- [12] P. Cătălin, P. L. Daniel, P. Gabriel, And D. D. Laurențiu, "Entrepreneurship In Sport," *Annals of Constantin Brancusi University of Targu-Jiu. Economy Series/Analele Universității Constantin Brâncuși din Târgu-Jiu Seria Economie*, no. 1, 2023.
- [13] Y. J. Ko, "Sport Management and Marketing: Overview and Recommendation for Future Research.[Gestión y marketing deportivo: Revisión y recomendación para investigaciones futuras]," *RICYDE. Revista Internacional de Ciencias del Deporte*. doi: 10.5232/ricyde, vol. 9, no. 33, pp. 205-207, 2013, doi: <https://doi.org/10.5232/ricyde2013.033>.
- [14] B. Penkert and R. Malhotra, "Global sportstech VC report 2019," *Berlin: Sportechx*, 2019.
- [15] C. Troise, V. Corvello, A. Ghobadian, and N. O'Regan, "How can SMEs successfully navigate VUCA environment: The role of agility in the digital transformation era," *Technological Forecasting and Social Change*, vol. 174, p. 121227, 2022, doi: <https://doi.org/10.1016/j.techfore.2021.121227>.
- [16] S. Ginting, P. Mei, N. Fitriana, and I. Muda, "Analysis of Boiler Operation Workload in Salted Fish Small Medium Enterprises (SMEs)," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 505, no. 1: IOP Publishing, p. 012151, doi: <https://doi.org/10.1088/1757-899X/505/1/012151>.
- [17] R. Rupeika-Apoga and K. Petrovska, "Barriers to Sustainable Digital Transformation in Micro-, Small-, and Medium-Sized Enterprises," *Sustainability*, vol. 14, no. 20, p. 13558, 2022, doi: <https://doi.org/10.3390/su142013558>.
- [18] P. Osterrieder, L. Budde, and T. Friedli, "The smart factory as a key construct of industry 4.0: A systematic literature review," *International Journal of Production Economics*, vol. 221, p. 107476, 2020, doi: <https://doi.org/10.1016/j.ijpe.2019.08.011>.
- [19] Z. Hasib Hussein, N. Benar, and S. Fatahi Milasi, "The Role of Marketing Ambidexterity on the Marketing Performance of the Iraq Basketball Federation Mediated by Managers' Competencies," *Journal of Global Sport Management*, pp. 1-25, 2025, doi: <https://doi.org/10.1080/24704067.2025.2554234>.
- [20] K. R. Memon and S. K. Ooi, "The dark side of industrial revolution 4.0-implications and suggestions," *Academy of Entrepreneurship Journal*, vol. 27, no. S2, pp. 1-18, 2021.
- [21] M. Salam, M. S. Farooq, A. Ikram, S. U. Rehman, Y. H. Mughal, and M. I. Khan, "Integrated collaborative reflection model (ICRM) Part 1: role and process based conceptual study of collaborative reflection and collaborative knowledge building," *Global Knowledge, Memory and Communication*, vol. 0, no. 0, pp. 1-18, 2026, doi: <https://doi.org/10.1108/GKMC-03-2025-0242>.
- [22] A. Roša and L. Lobanova, "Ethical responsibility of a company in the context of digital transformation of work: Conceptual model," *Sustainability*, vol. 14, no. 18, p. 11365, 2022, doi: <https://doi.org/10.3390/su141811365>.
- [23] A. M. Gomez-Trujillo and M. A. Gonzalez-Perez, "Digital transformation as a strategy to reach sustainability," *Smart and Sustainable Built Environment*, vol. 11, no. 4, pp. 1137-1162, 2021, doi: <https://doi.org/10.1108/SASBE-01-2021-0011>.
- [24] C. Lv, Y. Wang, and C. Jin, "The possibility of sports industry business model innovation based on blockchain technology: Evaluation of the innovation efficiency of listed sports companies," *Plos one*, vol. 17, no. 1, p. e0262035, 2022, doi: <https://doi.org/10.1371/journal.pone.0262035>.
- [25] H. Bouwman, S. Nikou, F. J. Molina-Castillo, and M. de Reuver, "The impact of digitalization on business models," *Digital Policy, Regulation and Governance*, vol. 20, no. 2, pp. 105-124, 2018, doi: <https://doi.org/10.1108/dprg-07-2017-0039>.
- [26] S. Mohammadi, A. Heidari, and J. Navkhsi, "Proposing a Framework for the Digital Transformation Maturity of Electronic Sports Businesses in Developing Countries," *Sustainability*, vol. 15, no. 16, p. 12354, 2023, doi: <https://doi.org/10.3390/su151612354>.
- [27] M. Sandelowski and J. Barroso, "Handbook for synthesizing qualitative research," ed: Springer Publishing Company, 2007.
- [28] T. H. Nguyen, X. C. Le, and T. H. L. Vu, "An extended technology-organization-environment (TOE) framework for online retailing utilization in digital transformation: empirical evidence from vietnam," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 8, no. 4, p. 200, 2022, doi: <https://doi.org/10.3390/joitmc8040200>.
- [29] B. Trenerry *et al.*, "Preparing workplaces for digital transformation: An integrative review and framework of multi-level factors," *Frontiers in psychology*. p. 822, 2021, doi: <https://doi.org/10.3389/fpsyg.2021.620766>.
- [30] B. Freudenreich, F. Lüdeke-Freund, and S. Schaltegger, "A stakeholder theory perspective on business models: Value creation for sustainability," *Journal of Business Ethics*, vol. 166, pp. 3-18, 2020, doi: <https://doi.org/10.1007/s10551-019-04112-z>.
- [31] T. T. Le and T. Le Quan Chau, "Drive sustainable corporate performance with digital transformation: the roles of environmental dynamism, sustainable innovation, and circular business model innovation," *Operations Management Research*, vol. 19, no. 1, p. 14, 2026, doi: <https://doi.org/10.1007/s12063-026-00584-1>.
- [32] L. Wessel, A. Baiyere, R. Ologeanu-Taddei, J. Cha, and T. Blegind-Jensen, "Unpacking the difference between digital transformation and IT-enabled organizational transformation," *Journal of the Association for Information Systems*, vol. 22, no. 1, pp. 102-129, 2021, doi: <https://doi.org/10.17705/1jais.00655>.
- [33] D. Horváth and R. Z. Szabó, "Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities?," *Technological forecasting and social change*, vol. 146, pp. 119-132, 2019, doi: <https://doi.org/10.1016/j.techfore.2019.05.021>.
- [34] C. Legner *et al.*, "Digitalization: opportunity and challenge for the business and information systems engineering community," *Business & information systems engineering*, vol. 59, pp. 301-308, 2017, doi: <https://doi.org/10.1007/s12599-017-0484-2>.
- [35] S. Fatahi Milasi, N. Benar, and A. Nazarian, "Economic Sustainability in Sports Ticketing: The Role of Technology Acceptance Factors and Blockchain Capabilities," *Journal of Advanced Sport Technology*, 2025, doi: <https://doi.org/10.22098/jast.2025.17265.1414>.
- [36] S. Kraus, S. Durst, J. J. Ferreira, P. Veiga, N. Kailer, and A. Weinmann, "Digital transformation in business and management research: An overview of the current status quo," *International Journal of Information Management*, vol. 63, p. 102466, 2022, doi: <https://doi.org/10.1016/j.ijinfomgt.2021.102466>.
- [37] C. Blais, "Innovation in SMEs in times of crisis: the ability to reconcile formality, agility and speed," *International Journal of Innovation and Technology Management*, vol. 20, no. 05, p. 2342002, 2023, doi: <https://doi.org/10.1142/S0219877023420026>.
- [38] U. Butolen, I. Vrečko, and I. Palčič, "A Taxonomy of Critical Success Factors for Implementing Corporate Sustainability Strategies: A Systematic Review," *Sustainability*, vol. 17, no. 12, p. 5469, 2025, doi: <https://doi.org/10.3390/su17125469>.
- [39] M. Salam, M. S. Farooq, A. Ikram, M. Shahzad, A. Ali, and N. Jaafar, "Revised artificial intelligence device use acceptance (RAIDUA) model: exploring privacy concerns for socially responsible AI deployment and ethical AI leadership," *Journal of Hospitality and Tourism Insights*, vol. 9, no. 2, pp. 496-518, Sep. 2025, doi: <https://doi.org/10.1108/JHTI-02-2025-0322>.
- [40] I. C. Melo, G. A. Queiroz, P. N. A. Junior, T. B. de Sousa, W. Yushimito, and J. Pereira, "Sustainable digital

- transformation in small and medium enterprises (SMEs): A review on performance," *Heliyon*, 2023.
- [41] P. Michelman, "Why sports is a great proving ground for management ideas," *MIT Sloan Management Review*, 2019.
- [42] T. Hacking and P. Guthrie, "A framework for clarifying the meaning of Triple Bottom-Line, Integrated, and Sustainability Assessment," *Environmental impact assessment review*, vol. 28, no. 2-3, pp. 73-89, 2008, doi: <https://doi.org/10.1016/j.eiar.2007.03.002>.
- [43] L. Floridi *et al.*, "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds and machines*, vol. 28, no. 4, pp. 689-707, 2018, doi: <https://doi.org/10.1007/s11023-018-9482-5>.
- [44] B. C. Stahl, J. Antoniou, M. Ryan, K. Macnish, and T. Jiya, "Organisational responses to the ethical issues of artificial intelligence," *AI & Society*, vol. 37, no. 1, pp. 23-37, 2022, doi: <https://doi.org/10.1007/s00146-021-01148-6>.
- [45] A. Martins, M. C. Branco, P. N. Melo, and C. Machado, "Sustainability in small and medium-sized enterprises: A systematic literature review and future research agenda," *Sustainability*, vol. 14, no. 11, p. 6493, 2022, doi: <https://doi.org/10.3390/su14116493>.
- [46] M. Halonen, "Improving sustainability within sports events via evaluation criteria," 2026.
- [47] H. Li, "Artificial intelligence, green transition and green total factor productivity in enterprises," *Scientific Reports*, vol. 15, no. 1, p. 33248, 2025, doi: <https://doi.org/10.1038/s41598-025-18878-y>.
- [48] R. Wiliandri, "A conceptual approach to identify factors affecting the digital transformation of micro, small and medium-sized enterprises (MSMEs) during COVID-19 pandemic in Indonesia," *Ekonomi Bisnis*, vol. 25, no. 2, p. 66, 2020, doi: <https://doi.org/10.17977/um042v25i2p66-85>.
- [49] C. Van Tonder, S. Hasanefendic, B. Bossink, C. Schachtebeck, and C. Nieuwenhuizen, "Internal organizational factors driving digital transformation for business model innovation in SMEs," *Journal of Business Models (JOBM)*, vol. 11, no. 2, pp. 86-109, 2023, doi: <https://doi.org/10.54337/jbm.v11i2.7403>.
- [50] I. C. Melo, G. A. Queiroz, P. N. A. Junior, T. B. de Sousa, W. F. Yushimito, and J. Pereira, "Sustainable digital transformation in small and medium enterprises (SMEs): A review on performance," *Heliyon*, vol. 9, no. 3, 2023, doi: <https://doi.org/10.1016/j.heliyon.2023.e13908>.
- [51] A. Senadjki, H. N. Au Yong, T. Ganapathy, and S. Ogbeibu, "Unlocking the potential: the impact of digital leadership on firms' performance through digital transformation," *Journal of Business and Socio-Economic Development*, vol. 4, no. 2, pp. 161-177, 2024, doi: <https://doi.org/10.1108/JBSED-06-2023-0050>.
- [52] I. Alyani, I. T. Saptono, and S. Jahroh, "Digital transformation drivers for Indonesian MSMEs," *Indonesian Interdisciplinary Journal of Sharia Economics (IIJSE)*, vol. 7, no. 1, pp. 423-441, 2024, doi: <https://doi.org/10.31538/ijse.v7i1.4374>.
- [53] S. Philbin, R. Viswanathan, and A. Telukdarie, "Understanding how digital transformation can enable SMEs to achieve sustainable development: A systematic literature review," 2022, doi: <https://doi.org/10.26784/sbir.v6i1.473>.
- [54] M. M. A. P. Mick, J. L. Kovaleski, and D. M. d. G. Chirolí, "Sustainable digital transformation roadmaps for SMEs: A systematic literature review," *Sustainability*, vol. 16, no. 19, p. 8551, 2024, doi: <https://doi.org/10.3390/su16198551>.
- [55] A. M. Magaz-González, M. García-Tascón, C. Sahelices-Pinto, A. M. Gallardo, and J. C. Guevara Pérez, "Technology and digital transformation for the structural reform of the sports industry: Building the roadmap," *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, vol. 238, no. 2, pp. 150-158, 2024, doi: <https://doi.org/10.1177/17543371231197323>.
- [56] H. Wang, L. Lu, Y. Fu, and Q. Li, "An empirical assessment of the influence of digital transformation on sports corporate sustainability," *Plos one*, vol. 19, no. 4, p. e0297659, 2024, doi: <https://doi.org/10.1371/journal.pone.0297659>.
- [57] Z. Gong, "The Strategic Transformation Path and Practice of Small and Medium-Sized Sporting Goods Enterprises," *Academic Journal of Business & Management*, vol. 7, no. 4, 2025, doi: <https://doi.org/10.25236/AJBM.2025.070429>.
- [58] Y. Hu *et al.*, "Understanding the digital transformation of SMEs: A systematic review from a dynamic capabilities perspective across business, technology, and societal domains," *Global Review of Humanities, Arts, and Society*, vol. 2, no. 1, pp. 1-17, 2026, doi: <https://doi.org/10.63802/grhas.V2.I1.186>.
- [59] S. Pu, Y. Ou, and O. Bai, "Government public services and regional digital transformation for sustainable development: an innovation ecosystem perspective," *Sustainability*, vol. 17, no. 12, p. 5314, 2025, doi: <https://doi.org/10.3390/su17125314>.
- [60] J. Zhu, J. S. Baker, Z. Song, X.-G. Yue, and W. Li, "Government regulatory policies for digital transformation in small and medium-sized manufacturing enterprises: an evolutionary game analysis," *Humanities and Social Sciences Communications*, vol. 10, no. 1, pp. 1-18, 2023, doi: <https://doi.org/10.1057/s41599-023-02250-4>.
- [61] A. Stroumpoulis and E. Kopanaki, "Theoretical perspectives on sustainable supply chain management and digital transformation: A literature review and a conceptual framework," *Sustainability*, vol. 14, no. 8, p. 4862, 2022, doi: <https://doi.org/10.3390/su14084862>.
- [62] X. Wang, K. Wang, B. Xu, and W. Jin, "Digitalisation and technological innovation: panaceas for sustainability?," *International Journal of Production Research*, vol. 63, no. 16, pp. 6071-6088, 2025, doi: <https://doi.org/10.1080/00207543.2025.2468883>.
- [63] T. Tran, P. Nguyen, N. Ho, N. Hạnh, T. Son, and H. Hang, "Factors affecting the commitment to digital transformation toward achieving sustainable development goals: Evidence from Vietnamese small and medium-sized enterprises," *Journal Of Lifestyle And Sdgs Review Vapeðumenu: Editora Alumniin*, vol. 5, no. 2, p. e04387, 2024, doi: <https://doi.org/10.47172/2965-730x.sdgsreview.v5.n02.pe04387>.
- [64] A. T. Tolossa, M. Singh, and R. K. Gautam, "Linking business sustainable development with entrepreneurial marketing practices: evidence from small and medium enterprises in Ethiopia," *Business Strategy & Development*, vol. 7, no. 3, p. e412, 2024, doi: <https://doi.org/10.1002/bsd2.412>.
- [65] M. Alojail and S. B. Khan, "Impact of digital transformation toward sustainable development," *Sustainability*, vol. 15, no. 20, p. 14697, 2023, doi: <https://doi.org/10.3390/su152014697>.
- [66] B. van Gils and H. Weigand, "Towards sustainable digital transformation," in *2020 IEEE 22nd conference on business informatics (CBI)*, 2020, vol. 1: IEEE, pp. 104-113, doi: <https://doi.org/10.1109/CBI49978.2020.00019>.
- [67] Y. Zhang, W. Zhao, and H. Liu, "Effect of Digital Transformation in Sports Companies on Green Innovation: Evidence from Listed Companies in China," *Sustainability*, vol. 16, no. 19, p. 8346, 2024, doi: <https://doi.org/10.3390/su16198346>.
- [68] F. Russo, "A Relational Ethical Framework for the Digital Society," *Digital Society*, vol. 3, no. 3, p. 65, 2024, doi: <https://doi.org/10.1007/s44206-024-00149-4>.
- [69] Y. Su, E. J. Wang, and P. Berthon, "Ethical marketing AI? A structured literature review of the ethical challenges posed by artificial intelligence in the domains of marketing and consumer behavior," 2023, doi: <https://doi.org/10.24251/HICSS.2023.603>.

- [70] M. Bourban, "Strong sustainability ethics," *Environmental Ethics*, vol. 43, no. 4, pp. 291-314, 2022, doi: <https://doi.org/10.5840/enviroethics202211332>.
- [71] Q. Shao, C. Jiang, J. J. Liou, P. Su, Y. Yuan, and Z. Dan, "Exploring the influencing factors of digital transformation: empirical results from SMEs in China," *Managerial and Decision Economics*, vol. 46, no. 4, pp. 2364-2380, 2025, doi: <https://doi.org/10.1002/mde.4463>.
- [72] I. Šimberová, A. Korauš, D. Schüller, L. Širáňová, J. Straková, and J. Váchal, "Threats and opportunities in digital transformation in smes from the perspective of sustainability: A case study in the Czech Republic," *Sustainability*, vol. 14, no. 6, p. 3628, 2022, doi: <https://doi.org/10.3390/su14063628>.
- [73] S. Fatahi Milasi and N. Benar, "A Conceptual Framework for Analyzing and Assessing Digital Maturity in Sports Enterprises Based on a Systematic Literature Review," *Journal of Sport Management and Development*, pp. -, 2026, doi: <https://doi.org/10.22124/jsmd.2025.31412.3012>.
- [74] S. Kraus, P. Jones, N. Kailer, A. Weinmann, N. Chaparro-Banegas, and N. Roig-Tierno, "Digital transformation: An overview of the current state of the art of research," *Sage Open*, vol. 11, no. 3, p. 21582440211047576, 2021, doi: <https://doi.org/10.1177/21582440211047576>.
- [75] C.-L. Chen, Y.-C. Lin, W.-H. Chen, C.-F. Chao, and H. Pandia, "Role of government to enhance digital transformation in small service business," *Sustainability*, vol. 13, no. 3, p. 1028, 2021, doi: <https://doi.org/10.3390/su13031028>.
- [76] V. Ratten, "Digital transformation in sport and social media," in *Sport startups: new advances in entrepreneurship*: Emerald Publishing Limited, 2020, pp. 89-104.
- [77] P. Pfister and C. Lehmann, "Digital value creation in German SMEs—a return-on-investment analysis," *Journal of Small Business & Entrepreneurship*, vol. 36, no. 4, pp. 548-573, 2024, doi: <https://doi.org/10.1080/08276331.2022.2037065>.
- [78] M. Salam, M. Radović-Marković, M. S. Farooq, D. Marković, and M. Vučković, "Artificial Intelligence Driven Digital Transformation Mediating the Relationship Between Artificial Intelligence Readiness and Business Process Efficiency," *IPSI TIR*, vol. 22, no. 2, pp. 32-41, Apr. 2026, doi: <https://doi.org/10.58245/ipsi.tir.2602.04>.
- [79] von S. Henrik, F. Diez, E. Vukovic, and P. Lalovich, "The Evolving Landscape Towards More Resilient Sports Investments: Institutional Engagement, Tokenization, and Incentives in Professional Football," *Journal of Entrepreneurship and Business Resilience*, vol. 7, no. 2, pp. 111-129, Dec. 2024.
- [80] J. Veselinović, A. Perović, V. Šiljak, and S. Bačevac, "Challenges of modern sport management," *Oditor*, vol. 8, no. 1, pp. 111-134, 2022, doi: <https://doi.org/10.5937/Oditor2201109V>.

Sadegh Fatahi Milasi is a PhD Candidate in Sports Management at the University of Gilan, focusing his research on the impact of digital transformation, sports business strategies, sustainability, and sports diplomacy. His academic pursuits aim to contribute to the understanding and advancement of these key areas within sports management.
E-mail: sadeghfatahi@phd.guilan.ac.ir
ORCID 0009-0004-8648-8787

Associate Prof. Dr. Nooshin Benar (*corresponding author*) is a sport management Associate Professor in the Sports science faculty at University of Guilan. She received her PhD from the Faculty of Sport Sciences at University of Guilan. Her research focuses on sport marketing, Sponsorship and consumer.
E-mail: nooshin.benar@guilan.ac.ir
ORCID 0000-0002-3377-2889

Dr. Maimoona Salam is an Assistant Professor at the University of Engineering and Technology, Lahore. Her research focuses on educational technology, artificial intelligence in learning, conversational AI, intelligent tutoring systems, and software engineering.
E-mail: maimoonasalam@uet.edu.pk
ORCID 0000-0001-7102-274X

Prof. Dr. Mirjana Radović-Marković is a Full Professor and principal research fellow. Professor Radović-Marković is the author of 308 scientific works. Her most famous books are "Entrepreneurship and Work in Gig Economy" (2021), "Globalization and Entrepreneurship in Small Countries", (2019), then "Qualitative methods in economics", (2017), as well as the book "Women and entrepreneurship: female durability, persistence and intuition at work" (2013). The last one in the mentioned series was a bestseller in America in August 2014. The recent book is Globalization and Entrepreneurship: Case of China, Japan and Bangladesh, Routledge, February 2025. She is a Fellow of The European Academy of Sciences and Arts (Salzburg), Academia Europaea (London), WAAS, USA and the National Academies of Sciences of Lithuania, Bangladesh, Zimbabwe, and Botswana.
E-mail: mirjana.radovic.markovic@bba.edu.rs
ORCID 0000-0003-2002-8732

Dr. Muhammad Shoaib Farooq (*corresponding author*) is a faculty member at the University of Engineering and Technology, Lahore. With over a decade of academic and research experience, his expertise spans entrepreneurship, business, management and corporate applications of artificial intelligence.
E-mail: sshoabfarooq@uet.edu.pk
ORCID 0000-0003-3263-2646

Validation-Centric AI-Native Decision Systems: Integrating the GAVA Loop with PRIME–INSPECT Governance

Avramović, Nebojša; Čomić, Tijana; Marković, Aleksandar; Čavoški, Sava;
Zornić, Nikola; and Vujović, Vladimir

Abstract: AI-native decision systems — in which generative artificial intelligence generates, executes, and adapts decisions in real time — are shifting the centre of gravity of analytical effort. Whereas the principal challenge used to be the extraction of insight, it is now becoming the validation of AI outputs while action is still being taken upon them. Existing approaches address this problem only partially: explainable artificial intelligence and governance frameworks treat validation as an *ex post* layer; reinforcement learning reduces it to a single scalar reward; control theory and the broader cybernetic tradition formalise the closed loop but neglect its socio-technical constraints; while the PRIME–INSPECT framework establishes a governance foundation, yet leaves real-time validation implicit. This paper proposes a validation-centric architecture that links the GAVA decision loop (Generate–Act–Validate–Adapt), introduced here, with the PRIME–INSPECT framework and elevates validation to a central analytical function comprising four subdimensions: statistical, operational, cognitive, and governance-related. The proposed architecture is empirically examined on a dual sample of IT professionals and top-management representatives, using descriptive statistics, multiple regression, mediation, moderation, and structural equation modelling. The results confirm the central role of trust, the negative effect of perceived risk, and the importance of top-management support, while robustness checks separated by the IT and TMT subsamples further strengthen the structural model. Taken as a whole, the findings support the view that validation should be treated as a first-order organisational stage in AI-native decision systems.

Index Terms: artificial intelligence governance, AI-native systems, behavioural intention, explainable artificial intelligence (XAI), GAVA loop, generative artificial intelligence, real-time decision-making, structural equation modelling, trust calibration, validation-centric architecture

1. INTRODUCTION

Artificial intelligence is transforming the way organisations make decisions, with generative artificial intelligence and foundation models [1] at the centre of this transformation. For most of their

development, artificial intelligence systems played an auxiliary role: they analysed historical data and provided users with insights on the basis of which those users acted. That relationship is now changing. Large language models, multimodal systems, and autonomous agents [2] have enabled the emergence of a different kind of system — AI-native systems that independently generate, evaluate, and execute decisions in real time [3], [4], [5].

This is a change that goes beyond mere improvement. In data-driven environments, the main constraints were data availability and the ability to extract meaningful insights. In AI-native systems, the situation is reversed: interpretations, recommendations, and actions are produced faster than humans can review them. The scarce resource is no longer insight, but the capacity to validate what the system produces — before an action is taken and while that action is still underway. This shift may be described as a transition from an economy of insight to an economy of validation, in which organisational performance increasingly depends on the reliability, controllability, and trustworthiness of decisions made with the assistance of artificial intelligence.

The consequences of this shift are most pronounced in high-risk and highly automated domains, such as mining, metallurgy, smart manufacturing, and financial systems. In such domains, an AI decision constitutes an action with immediate operational consequences. The speed and autonomy of these systems increase the risk of erroneous, biased, or unvalidated decisions propagating before anyone has time to respond. Under such conditions, validation is not a secondary control activity, but a core analytical function.

Avramović et al. [6] recently proposed the PRIME–INSPECT framework, a socio-technical model intended for governance, ethical alignment, and trust calibration in AI-supported decision-making in the context of Industry 4.0. This framework positions governance as the basic perspective from which AI-supported decisions are assessed. However, real-time validation is not

explicitly developed within it as an operational stage, nor is it situated within a closed decision loop. It is precisely this insufficiently developed aspect that forms the point of departure for the present paper, which directly builds upon the PRIME-INSPECT framework.

In response, this paper introduces an integrated conceptual framework for AI-native decision-making, composed of two parts. The first is the GAVA loop (Generate-Act-Validate-Adapt) — an operational, decision-theoretic model of the flow of AI-native processes, which runs continuously as the system generates decisions, executes actions, validates outcomes, and adapts to new information. The second part links the GAVA loop with the PRIME-INSPECT framework, thereby forming an architecture centred on validation and trust calibration, in which operational decision-making and socio-technical governance function as a single whole.

The contribution of this paper lies in three key respects. First, validation is positioned as the central organisational principle of AI-native systems, thereby shifting the analytical role of data science from the generation of insight to the performance of validation and control. Second, the operational perspective (GAVA) and the governance perspective (PRIME-INSPECT) are combined into a single framework that encompasses the interaction of algorithmic processes, human oversight, and institutional mechanisms. Finally, this integrated framework is empirically supported by inferential methods — multiple regression, mediation, moderation, and structural equation modelling — thereby continuing the future research programme outlined by Avramović et al. [6].

The paper is structured as follows, in accordance with the recommendations of Milutinović [7]. Section 2 presents the problem of real-time validation and the reasons why it is becoming increasingly important. Section 3 reviews existing solutions — explainable artificial intelligence and governance frameworks, reinforcement learning, control theory and cybernetics, and the PRIME-INSPECT framework — with a critical assessment of the limitations of each from the perspective of validation. Section 4 develops the proposed solution: the GAVA decision loop and its integration with the PRIME-INSPECT framework. Section 5 presents the empirical analysis. Section 6 discusses the theoretical, practical, and methodological implications and identifies limitations and directions for future research. Section 7 provides the conclusion, including a discussion of the implementation trade-offs of the architecture and the naming of the creative method applied.

2. PROBLEM DEFINITION

Real-time validation is becoming a core analytical function in AI-native decision systems.

The growing importance of real-time validation can be explained through three interrelated processes. The first is the increasingly rapid deployment of generative artificial intelligence and autonomous agents in operational environments: the volume and frequency of their outputs and decisions far exceed those of earlier rule-based or supervised-learning systems, while the share of decisions made without prior human review continues to increase. The second process arises from the application domains themselves — mining, metallurgy, smart manufacturing, financial trading, medical triage, and energy grids — which are subject to strict time constraints and leave little room for retrospective verification. The third is the increasingly stringent regulatory and ethical environment: the European Union Artificial Intelligence Act, sector-specific governance regimes, and stakeholder expectations require decisions that are explainable, accountable, and auditable under operational conditions.

The scale of this change is visible in practice. Research on the adoption of artificial intelligence in enterprises indicates a growing number of organisations using it not only for analytical support, but also for operational decision-making and task execution. At the same time, autonomous agents are increasingly moving from research prototypes into production environments, including customer support, content moderation, code generation, algorithmic trading, and clinical triage.

The pace of this change is particularly significant. The interval between the emergence of new generative capabilities and their practical deployment is becoming markedly shorter, leaving traditional cycles of verification, testing, and validation with ever less room to mature before AI-supported decisions begin to affect end users.

The regulatory environment is moving towards requirements that, in practical terms, mandate real-time validation. The European Union Artificial Intelligence Act, adopted in 2024, whose obligations for high-risk systems enter into force in 2026, requires continuous post-market monitoring of system performance, documented risk-management processes, and human oversight mechanisms throughout the entire life cycle. The U.S. NIST Artificial Intelligence Risk Management Framework (2023), together with the profile for generative artificial intelligence (2024), defines the functions of governing, mapping, measuring, and managing risk that must operate during system use, and not only at the design stage. Similar obligations are imposed by sector-specific

regulations — financial supervision of algorithmic trading, frameworks for medical devices with artificial-intelligence-based diagnostics, and occupational-safety requirements in autonomous manufacturing. Taken together, these regimes transform validation from an internal quality-control activity into an externally verifiable and continuously demonstrable property of the system, in line with recent calls for extreme AI risks to be regulated at the level of public policy [8].

The cost of an unvalidated decision increases with the degree of autonomy and speed. In supervised analytical environments, errors are most often detected by a human at the very point of decision-making; in AI-native systems, an incorrect decision is executed just as quickly as a correct one, before validation has even had time to occur. Recorded incidents confirm this: sudden market crashes in algorithmic trading, failures of automated content moderation, bias in automated credit approval, and erroneous triage in clinical systems [9], [10] have repeatedly concentrated harm within minutes rather than days. The gap between the speed of decision-making and the slowness of verification represents the operational core of this problem. For this reason, validation must be understood as a function embedded in the decision loop itself, rather than as a subsequent audit mechanism.

Under such conditions, validation cannot remain an *ex post* assessment activity. The interval between the emergence of a decision and its operational consequences is too short for verification outside the decision flow, while the cost of an unvalidated decision is potentially too high. Validation must therefore be continuous, embedded in the decision loop, and capable of assessing, in real time, the statistical accuracy, operational impact, cognitive intelligibility, and governance compliance of decisions.

The challenge can therefore be defined as follows. AI-native decision systems generate decisions at scale and execute them autonomously. Existing analytical paradigms are largely oriented either towards user acceptance of artificial intelligence or towards *ex post* model evaluation outside the actual decision flow, with neither of these approaches fully corresponding to the dynamics of real-time decision-making. What is missing is an integrated decision-theoretic framework in which validation has the status of a constitutive operational stage, embedded in a closed feedback loop and connected with mechanisms of governance and trust calibration. Designing such a framework is the central problem of this paper.

3. EXISTING SOLUTIONS

Four streams of prior research are relevant to the problem of validation presented in Section 2: explainable artificial intelligence and its governance, reinforcement learning, control theory and cybernetics, and the PRIME-INSPECT framework. Each of these captures one aspect of the problem, but none offers an integrated solution centred on validation. In this section, they are critically examined precisely from that perspective.

3.1 Explainable Artificial Intelligence and Governance

Research on explainable artificial intelligence (XAI), trust, and governance has highlighted important dimensions such as transparency, accountability, fairness, and risk management [11], [12], [13], [14]. Its core message is that an artificial intelligence system should be understandable, auditable, and aligned with ethical norms.

Viewed from a validation-centric perspective, however, this body of research still treats validation as a wrapper around a model developed independently. Explanations are, as a rule, generated after the model output, while governance is carried out through external audit rather than through mechanisms embedded in the operational decision loop itself. Validation is thus not formalised as a first-order operational stage, nor is it aligned with the speed and autonomy required by AI-native real-time systems.

3.2 Reinforcement Learning and Closed-Loop Systems

Reinforcement learning (RL) [15], [16] and related closed-loop paradigms are the closest theoretical antecedents of contemporary AI-native decision systems. An RL agent acts within a given environment and learns from feedback signals that shape its future behaviour. In this way, RL shares with these systems the basic logic of a closed structure, in which decisions, outcomes, and adaptation are connected through feedback.

From the standpoint of validation, the problem is that RL compresses feedback into a single scalar reward. Such an abstraction is extremely useful for learning, but it reduces the multidimensional nature of validation — statistical accuracy, operational indicators, cognitive intelligibility, and governance compliance — to a single number. In addition, RL generally operates through the relationship between an agent and an environment, while institutional constraints, regulatory obligations, and ethical alignment do not have a clearly defined place within the decision loop itself. More recent variants partially mitigate these limitations: constrained policy optimisation [17], safe reinforcement learning [18], and multi-

objective reinforcement learning [19] introduce safety constraints and multiple objectives into the learning process. Nevertheless, such approaches remain primarily within the bounds of decision theory and are not, in themselves, connected with broader institutional and governance mechanisms. In this sense, validation in RL is most often implicitly contained in the design of the reward function but is not explicitly separated as a distinct operational stage.

3.3 Control Theory and Cybernetics

Control theory [20] and the broader cybernetic tradition provide the canonical formalisation of feedback systems. A controller observes the state of the system, calculates a control signal, and applies it, after which the resulting state is measured and fed back in order to guide further control. Cybernetics generalises this picture to self-regulating systems that also adapt their own goals.

Here too, the validation-centric perspective reveals a limitation. Control theory and cybernetics formalise the loop, but leave the notion of feedback underspecified, reducing it to measurement and error correction. In a real-time AI-native system, the counterpart of error must include not only statistical conformity but also operational performance, human intelligibility, and institutional permissibility. Contemporary extensions mitigate these limitations to some extent. Model predictive control, with explicit state and input constraints [21], enables the formal encoding of operational and safety constraints. Similarly, supervisory architectures with human-in-the-loop adaptation [22] introduce additional mechanisms for oversight and correction of system behaviour. Even so, these formalisms continue to treat the controlled object primarily as a physical rather than a socio-technical system and do not directly express institutional, regulatory, or ethical conditions. The closed-loop formalism is therefore a necessary but not sufficient condition for validation in AI-native systems.

3.4 The PRIME–INSPECT Framework

Avramović et al. [6] introduced the PRIME–INSPECT framework as a socio-technical model of trustworthy intelligent automation in the context of Industry 4.0. Its PRIME component comprises the execution layer through the phases of prediction, regulation, interpretation, mitigation, and execution, while the INSPECT component comprises the governance layer through data integrity, navigability of explanations, supervisory control, policy and governance maturity, ethical compliance, collaboration, and trust calibration. The framework links algorithmic processes with organisational and institutional controls and was

descriptively tested on a dual sample of IT professionals and top-management representatives. Full formal definitions, including the equations describing the PRIME execution flow and the INSPECT governance conditions, are provided in the original paper and are not repeated here.

From a validation-centric perspective, the PRIME–INSPECT framework establishes the institutional and ethical foundation necessary for trustworthy artificial intelligence, but does not formalise validation as an explicit first-order operational stage within a real-time decision loop. Within that framework, validation is implicitly contained in the Mitigate phase of the PRIME component and in the Supervisory Control component of the INSPECT framework. Significantly, Avramović et al. [6], in their section on future research, themselves emphasise the need to operationalise the dimensions of the PRIME–INSPECT framework on larger samples and to examine their interrelationships by means of inferential methods. The present paper develops the framework precisely in that direction: it embeds an explicit validation stage into a closed decision loop — the GAVA loop presented in Section 4 — and provides the inferential evidence that the original paper left for future research (Section 5).

3.5 MLOps and Operational Monitoring of Artificial Intelligence

The growing literature on MLOps and operational monitoring of artificial intelligence addresses the engineering procedures required for machine-learning systems to remain reliable in production. Sculley et al. [23] point to the hidden technical debt of such systems — the erosion of boundaries between components, implicitly defined consumers of system outputs, hidden feedback loops, and entangled data and model processing pipelines — while warning that maintenance costs may exceed the costs of initial development. Documentation tools, such as Model Cards [24], make visible a model’s intended use, its limitations, and its performance disaggregated by subgroups. This body of research shows that production-grade artificial intelligence requires continuous monitoring of drift, performance degradation, and outcomes across subgroups. From a validation-centric perspective, MLOps already provides operational mechanisms for part of the functions that GAVA elevates to the level of an analytical stage; what it does not, by itself, provide is the integration of the cognitive and governance dimensions of validation, which the proposed framework treats as equal in status.

4. PROPOSED SOLUTION: THE VALIDATION-CENTRIC GAVA-PRIME-INSPECT ARCHITECTURE

This section presents the proposed solution. First, the GAVA decision loop is introduced as an independent theoretical contribution in the field of decision-making, separate from the PRIME-INSPECT framework. The GAVA loop is then formalised in relation to reinforcement learning, control theory, and cybernetics, where it is presented as their generalisation in which validation is isolated as an explicit stage. After that, the GAVA loop is linked with the PRIME-INSPECT framework and the integrated validation-centric architecture is presented (Figure 1, Table 2). Finally, validation is defined as a constitutive analytical function based on four subdimensions.

4.1 The GAVA Decision Loop

GAVA is introduced as an integrative theoretical abstraction for understanding AI-native decision-making. The decision-making process is presented as a continuous four-phase cycle — Generate, Act, Validate, Adapt — unfolding in real time, with each phase conditioning and preparing the next.

Generate (G): The system produces a candidate decision based on a learned representation of the environment. The underlying mechanism may be a large language model, a predictive ensemble, or a hybrid processing pipeline; GAVA is agnostic in this respect and treats the model as a black box that provides both a decision and its rationale.

Act (A): The system executes the candidate decision in the environment, changing the state of the system and producing measurable outcomes.

Validate (V): The system evaluates the decision and its outcomes simultaneously along statistical, operational, cognitive, and governance dimensions, treating validation as a structured, multidimensional assessment rather than as a single scalar signal.

Adapt (A): On the basis of these multidimensional results, the system adapts both its own decision policy and its own validation criteria, thereby establishing the conditions for the next cycle.

Two features distinguish the GAVA loop from earlier closed-loop paradigms. The first is the multidimensionality of validation. Whereas reinforcement learning expresses feedback through a single reward and control theory through a single error signal, GAVA preserves four dimensions throughout the entire cycle — statistical accuracy, operational performance, cognitive intelligibility, and governance compliance — as distinct components. Collapsing them into a single number would discard precisely

the information required for targeted adaptation, since different combinations of failure require different corrective actions. The second feature is that adaptation encompasses the validation criteria themselves, and not only the decision policy. Data drift, distributional shift, and regulation changes require the system to update not only the way it decides, but also the way it assesses whether a decision is good. In this sense, GAVA generalises reinforcement learning, control theory, and cybernetics: it retains their closed-loop structure, but enriches the feedback channel with structured, multidimensional information and allows validation criteria to evolve over time.

GAVA is therefore an abstraction, not a procedural recipe. It does not prescribe a specific model architecture, a specific validation method, or a specific adaptation rule. Its purpose is to provide a conceptual framework within which existing techniques — RL agents, explainable-AI tools, drift detectors, and governance checks — can be aligned into a coherent decision-making system in which validation is an integral operational stage.

4.2 Relationship with Reinforcement Learning, Control Theory, and Cybernetics

GAVA is consistent with established decision-theoretic paradigms and may be read as a generalisation of each of them. From the perspective of reinforcement learning, the Generate phase corresponds to policy generation, Act to action execution, Validate to the reward signal, and Adapt to policy updating. The first generalisation introduced by GAVA is the multidimensionality of the Validate phase instead of a scalar reward, while the second is the possibility that adaptation may modify the reward function itself.

From the perspective of control theory, Generate corresponds to the formulation of the control signal, Act to system actuation, Validate to measurement and error assessment, and Adapt to controller tuning. GAVA generalises classical control by treating measurement as a structured assessment along multiple axes and by allowing the controller's objective to change in accordance with governance constraints.

From a cybernetic perspective, GAVA realises a self-regulating system that observes its own outputs, compares them with a structured set of desired states, and adapts its behaviour

integration yields a validation-centric closed-loop architecture, shown in Figure 1.

Within this architecture, GAVA serves as the decision engine, while PRIME–INSPECT provides

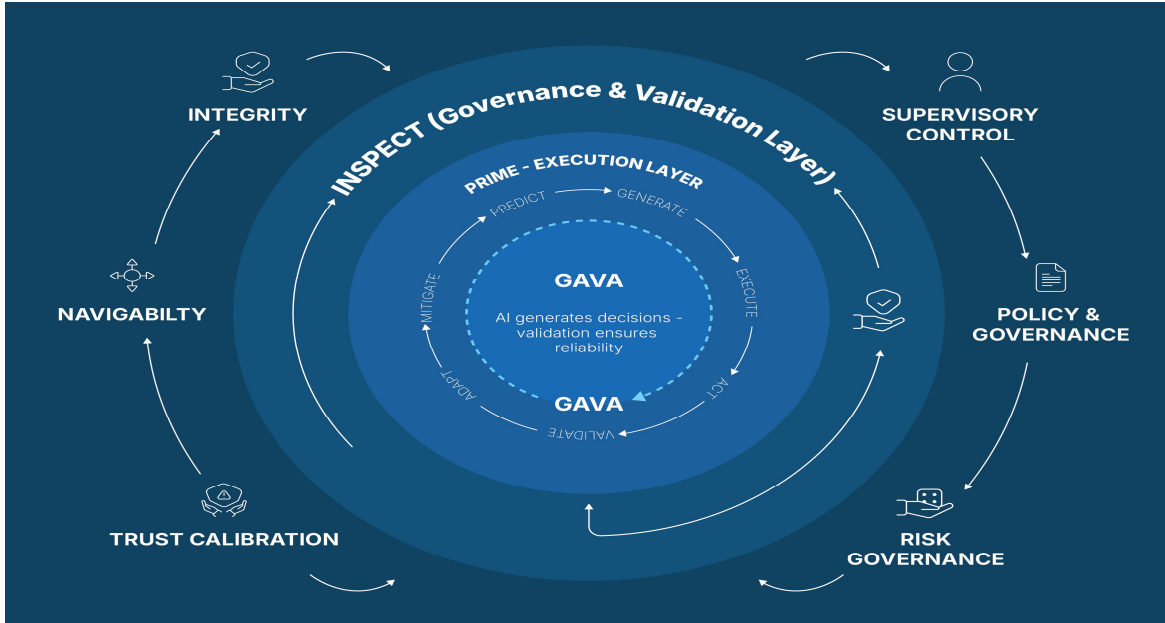


Figure 1. The integrated GAVA–PRIME–INSPECT validation-centric architecture. Legend: central GAVA loop; PRIME execution layer; INSPECT governance layer. Description: validation is embedded within the closed loop. Implication: autonomy and control achieved jointly through integrated operational/governance layers.

accordingly. The four phases correspond to observation, comparison, decision-making, and behavioural adaptation in the classical cybernetic sense.

The relationship of the GAVA loop to the above-mentioned earlier closed-loop paradigms is summarised in Table 1.

Table 1. Mapping between Reinforcement Learning and the GAVA Decision Loop with Validation-Centric Extensions

Reinforcement Learning	GAVA	Extension provided by GAVA
Policy generation	Generate	AI reasoning (LLM-based or hybrid)
Action execution	Act	Autonomous execution under governance constraints
Reward signal (scalar)	Validate	Multi-dimensional validation: statistical, operational, cognitive, governance
Policy update	Adapt	Joint update of both the decision policy and the validation criteria

4.3 Integration with the PRIME–INSPECT Framework

GAVA describes the operational dynamics of decision-making, while PRIME–INSPECT captures the socio-technical governance that constrains and stabilises those dynamics. Their

the execution and governance foundation. The Generate phase in the GAVA loop corresponds to the Predict component of PRIME, whose outputs are grounded in the INSPECT condition of Integrity of Data. The Act phase corresponds to the Execute component, constrained by the condition of Policy and Governance Maturity. The Validate phase corresponds to the Regulate component and relies on the conditions of Supervisory Control, Navigability, and Ethical Compliance — this is precisely where the multidimensional validation function V operates. The Adapt phase corresponds to the feedback loop of the PRIME framework and is stabilised by the conditions of Trust Calibration and Collaboration. This two-axis mapping is formalised in Table 2.

The integration is bidirectional, in accordance with the method of symbiosis [25]. From the PRIME–INSPECT framework, GAVA takes over the institutional and ethical foundation — supervisory control, trust calibration, ethical compliance, navigability, and collaboration — which the operational loop does not determine by itself. In return, from the GAVA loop, the PRIME–INSPECT framework receives an explicit operational closure of the validation feedback loop, which remained implicit in the original framework. Neither paradigm provides this capability on its own; only together do they determine what is validated and how the

institutional context evaluates and constrains the result. This also reveals a key structural property: in the GAVA–PRIME–INSPECT architecture, the closed loop is neither exclusively operational, as in reinforcement learning or control theory, nor exclusively governance-oriented, as in pure AI governance frameworks. It is both at the same time, with validation serving as the link between the two layers.

Table 2. Mapping between the GAVA Decision Loop and the PRIME–INSPECT Framework

GAVA Phase	PRIME Component	INSPECT Condition
Generate	Predict	Integrity of Data
Act	Execute	Policy and Governance
Validate	Regulate	Supervisory Control, Ethical Compliance
Adapt	Feedback loop	Trust Calibration

4.4 Validation as a First-Order Analytical Function

The integrated architecture elevates validation from an implicit, ex post activity to a central analytical function within the loop. Four subdimensions are distinguished — statistical, operational, cognitive, and governance validation — and the system evaluates them in every pass through the loop, with the validation function returning an assessment for each dimension and forwarding it to the Adapt phase. These four subdimensions may be illustrated by the example of a large-scale credit-scoring decision. Statistical validation checks the posterior calibration of the prediction and the range of uncertainty. Operational validation determines whether the decision keeps business indicators within prescribed boundaries — default rates, regulatory capital, and time budgets. Cognitive validation examines whether the explanation presented to the human in the loop is intelligible and usable for the specific operator, drawing on the literature on explainability and trust calibration [14], [26], [22]. Governance validation confirms that the decision is consistent with documented policies, ethical obligations, and audit trails [24]. Failure in any single dimension is sufficient to trigger the Adapt phase or, in the extreme case, a halt: the loop is not closed solely based on prediction error, but on the basis of the full multidimensional validation function.

This redefinition of validation also changes the analytical role of data science. In AI-native systems, the practitioner is no longer responsible only for the predictive model, but also for the validation infrastructure surrounding it — validation functions, drift detectors, explanation generators, governance mechanisms, and

adaptation operators. This is precisely the shift described in Section 1 as the transition from an economy of insight to an economy of validation.

4.5 Research Hypotheses

The integrated GAVA–PRIME–INSPECT architecture implies a set of testable relationships among the socio-technical constructs that govern validation-centric decision-making. Drawing on the validation dimensions from Section 4.4 and on the constructs operationalised in the reused dataset of Avramović et al. [6], seven hypotheses are formulated and empirically tested in Section 5.

H1. Greater explainability of artificial intelligence (XAI) has a positive effect on trust in artificial intelligence.

H2. Greater explainability of artificial intelligence has a positive and direct effect on behavioural intention to adopt artificial intelligence in real-time decision-making.

H3. Trust in artificial intelligence mediates the relationship between its explainability and behavioural intention.

H4. Higher perceived risk has a negative effect on behavioural intention to adopt artificial intelligence in real-time decision-making.

H5. Greater top-management support has a positive effect on behavioural intention to adopt artificial intelligence in real-time decision-making.

H6. Artificial-intelligence literacy positively moderates the relationship between explainability and trust.

H7. Trust in artificial intelligence positively predicts perceived real-time decision-making performance (RTD), either directly or indirectly through behavioural intention. This relationship is treated as exploratory, since it is not estimated in the primary CB-SEM model.

Hypotheses H1–H3 operationalise the Validate phase of the GAVA loop and its cognitive validation subdimension, while hypotheses H4–H6 encompass the organisational conditions under which validation is maintained.

5. EMPIRICAL ANALYSIS

This section provides empirical support for the integrated GAVA–PRIME–INSPECT framework. It examines the relationships between explainability, trust, perceived risk, top-management support, behavioural intention, and real-time decision-making performance, presenting descriptive statistics, Pearson correlations, multiple linear regression, mediation and moderation tests, quadratic regression, group-comparison tests, and structural equation modelling.

5.1 Sample and Data

The analysis in this section is based on the

same dual sample previously presented in descriptive form by Avramović et al. [6]: 151 IT professionals and 85 top-management representatives. Whereas the original paper used this dataset to descriptively establish the recognisability of the PRIME-INSPECT framework dimensions as constructs, the present paper takes a further step by applying inferential techniques — multiple regression, mediation, moderation, and structural equation modelling — to test claims derived from the integrated GAVA-PRIME-INSPECT framework, in accordance with the direction for future research explicitly indicated in the original paper. Items from the original instrument were re-operationalised into seven constructs (XAI, Trust, Perceived Risk, AI Literacy, TMT Support, Behavioural Intention, RTD Performance) so as to reflect the validation-centric perspective. The XAI, Trust, and Behavioural Intention constructs were reconstructed from items previously grouped under broader operational and governance labels in [6], while the remaining constructs — Perceived Risk, AI Literacy, TMT Support, and RTD Performance — rely on partially overlapping sets of items with adapted labels. The two instruments (IT and TMT) differ in the number and wording of items per construct, reflecting the technical and strategic perspectives, respectively; this is acknowledged in the limitations and constitutes the reason for the robustness checks separated by subsample, presented below.

The data were collected from October to December 2025 through a structured online questionnaire distributed via professional networks and organisational contacts, using a combination of convenience and snowball sampling. Respondents came from organisations in which AI-native decision systems had already been introduced or were being actively considered.

The dataset consists of two respondent groups. The group of IT professionals (N≈151) comprises data scientists, machine-learning engineers, and IT managers, representing the technical-implementation perspective. The top-management team (TMT) group (N≈85) comprises senior executives responsible for strategic decision-making, governance, and oversight, representing the strategic perspective. All constructs were measured on a four-point Likert scale, ensuring group comparability and enabling direct comparative analysis. Participation was voluntary and anonymous.

This dual approach reflects the socio-technical nature of AI-native systems, in which technical capabilities and the organisational context jointly shape decision-making.

Table 3. Operationalisation of constructs and allocation of survey items, re-operationalised from [6].

Construct	Representative indicators	IT instr.	TMT instr.
XAI (explainability)	Understanding the reasons behind AI recommendations; verifying the factors that drive them	D1–D2	D1–D3
Trust	Confidence in AI outputs and their compliance with defined rules	D5–D6	D6, D10
Perceived Risk	Erroneous decisions, loss of control, data sensitivity, and ethical exposure	C1–C5	C1–C5
AI Literacy	Self-assessed AI/ML expertise (screening item)	Q3	Q4
TMT Support	Leadership support, accountability, and IT–business collaboration on AI	E1–E4	E1–E6
Behavioural Intention	Intention to expand AI use and embed it in new initiatives	B3–B5	B3–B5
RTD Performance	Perceived gains in decision speed, accuracy, efficiency, and ROI	H1–H4	G1–G5

Since the dataset originally presented by Avramović et al. [6] in descriptive form is re-operationalised here, Table 3 shows how each construct maps onto the corresponding items in both instruments. Explainability and trust were derived from items concerning intelligibility and confidence; perceived risk from the block concerning risk perception; top-management support from leadership-related items; behavioural intention from adoption-intention items; and RTD performance from items concerning perceived performance. AI literacy is captured through respondents' self-assessment of expertise in artificial intelligence and machine learning.

5.2 Descriptive Statistics and Reliability

Table 4 presents the means, standard deviations, and Cronbach's alpha coefficient for

each construct. All α values exceed the conventional threshold of 0.70 [27], confirming satisfactory internal consistency; the highest reliability is observed for the Trust construct ($\alpha = 0.88$). The Trust, Behavioural Intention, and RTD Performance constructs have relatively high mean values, indicating a general openness towards AI-supported decision-making, while Perceived Risk shows the greatest variability, suggesting divergent attitudes towards risks associated with artificial intelligence.

Table 4. Descriptive statistics and reliability of measurement scales

Construct	Mean	SD	Cronbach's α
XAI	3.12	0.54	0.84
Trust	3.08	0.57	0.88
Perceived Risk	2.41	0.71	0.79
AI Literacy	2.95	0.60	0.82
TMT Support	2.87	0.65	0.85
Behavioural Intention (BI)	3.05	0.58	0.87
RTD Performance	3.01	0.55	0.86

5.3 Correlation Analysis

Table 5 presents the Pearson correlations among the seven constructs. The strongest positive relationship appears between Trust and RTD Performance ($r = 0.696$, $p < 0.01$), confirming the central role of trust in effective AI-supported decision-making. The strong correlation between XAI and Trust ($r = 0.645$) supports the role of explainability in building user trust [14]. Perceived Risk shows a significant negative correlation with Behavioural Intention ($r = -0.456$), while TMT Support correlates positively with the same construct ($r = 0.589$), underscoring the importance of organisational support. The sign of the correlations for the XAI construct differs from that of the corresponding aggregate in [6]: here, XAI has been re-operationalised around indicators of explainability and intelligibility — the perceived clarity of explanations — whereas in the earlier paper, related items were grouped into a broader operational-governance construct. This modification is consistent with the validation-centric perspective and is noted in the limitations.

Table 5. Pearson correlation matrix

Variable	1	2	3	4	5	6	7
1. XAI	1						
2. Trust	0.645* *	1					
3. Risk	-0.312* *	-0.421* *	1				
4. AI Literacy	0.511* *	0.533* *	-0.274* *	1			
5. TMT Support	0.402* *	0.512* *	-0.338* *	0.441* *	1		
6. BI	0.478* *	0.601* *	-0.456* *	0.502* *	0.589* *	1	
7. Performance	0.522* *	0.696* *	-0.401* *	0.488* *	0.574* *	0.612* *	1

Note: ** $p < 0.01$.

5.4 Regression Analysis

Model 1 (Table 6) examines XAI as a predictor of Trust, controlling for AI Literacy. XAI has a strong and statistically significant effect on Trust ($\beta = 0.582$, $p < 0.001$), with $R^2 = 0.51$, thereby confirming the hypothesised $XAI \rightarrow Trust$ relationship.

Model 2 (Table 7) examines the determinants of Behavioural Intention. The strongest predictor is Trust ($\beta = 0.421$, $p < 0.001$). Perceived Risk acts in the opposite direction ($\beta = -0.287$, $p < 0.001$), in line with the findings of Glikson and Woolley [26], while TMT Support promotes it ($\beta = 0.318$, $p < 0.001$). The direct effect of XAI on BI is small ($\beta = 0.129$, $p = 0.028$), indicating that its influence operates predominantly indirectly, through trust.

Table 6. Regression results (DV: Trust)

Variable	β	SE	T	P
XAI	0.582	0.061	9.54	<0.001
AI Literacy	0.214	0.055	3.89	<0.001

$R^2 = 0.51$.

Table 7. Regression results (DV: Behavioural Intention)

Variable	β	SE	T	P
Trust	0.421	0.072	5.84	<0.001
Risk	-0.287	0.069	-4.16	<0.001
TMT Support	0.318	0.064	4.97	<0.001
XAI	0.129	0.058	2.22	0.028

$R^2 = 0.63$.

5.5 Mediation and Moderation

The mediation test ($XAI \rightarrow Trust \rightarrow BI$) yields an indirect effect of 0.102, with a 95% bootstrap

confidence interval of [0.021, 0.187] and $p < 0.01$. In combination with the modest direct effect of XAI on BI reported in Section 5.4, this indicates partial mediation: XAI affects behavioural intention primarily through trust.

The moderation test shows that AI Literacy strengthens the XAI $\rightarrow$ Trust relationship ($\beta = 0.167$, $p = 0.012$). Respondents with a higher level of artificial-intelligence literacy translate explainability into trust more effectively, which is consistent with the cognitive subdimension of validation introduced in Section 4.4.

As a robustness check of the linear Trust $\rightarrow$ Performance relationship, a quadratic specification including the Trust² term was examined. In the available data, no stable support was found for an inverted-U pattern of trust calibration. Trust calibration, therefore, remains a theoretically motivated construct that requires larger samples and longitudinal designs for confirmatory testing [22].

5.6 Comparative Analysis: IT and TMT

Welch's t-tests of group differences (Table 8) show that IT respondents report significantly higher trust than TMT respondents (3.21 versus 2.89, $t = 3.42$, $p < 0.001$). TMT respondents, on the other hand, perceive higher risk (2.58 versus

governance layer (PRIME–INSPECT) reflects the priorities of top management.

Table 8. Group differences between IT and TMT respondents (Welch t-test)

Construct	IT Mean	TMT Mean	t	P
Trust	3.21	2.89	3.42	<0.001
Risk	2.31	2.58	-2.77	0.006
Governance	2.68	3.02	-3.11	0.002

5.7 Structural Equation Modelling

The structural equation model (SEM) estimates the direct and indirect relationships among the seven constructs simultaneously. The fit indices are $\chi^2(209) = 396.12$, $p < 0.001$, $\chi^2/df = 1.90$, CFI = 0.94, TLI = 0.93, RMSEA = 0.058, and SRMR = 0.047, exceeding the conventional acceptability threshold of 0.90 according to Hair et al. [27], although slightly below the stricter threshold of 0.95 proposed by Hu and Bentler [28]. To test possible measurement non-invariance between the two instruments, the same model was re-estimated on the IT subsample ($N = 151$), where it showed comparable fit (CFI = 0.94, TLI = 0.93, RMSEA = 0.060), while the TMT subsample ($N = 85$) showed acceptable but weaker fit (CFI = 0.85, RMSEA = 0.11), as expected given its smaller size. The structural paths reported below are stable across both specifications.

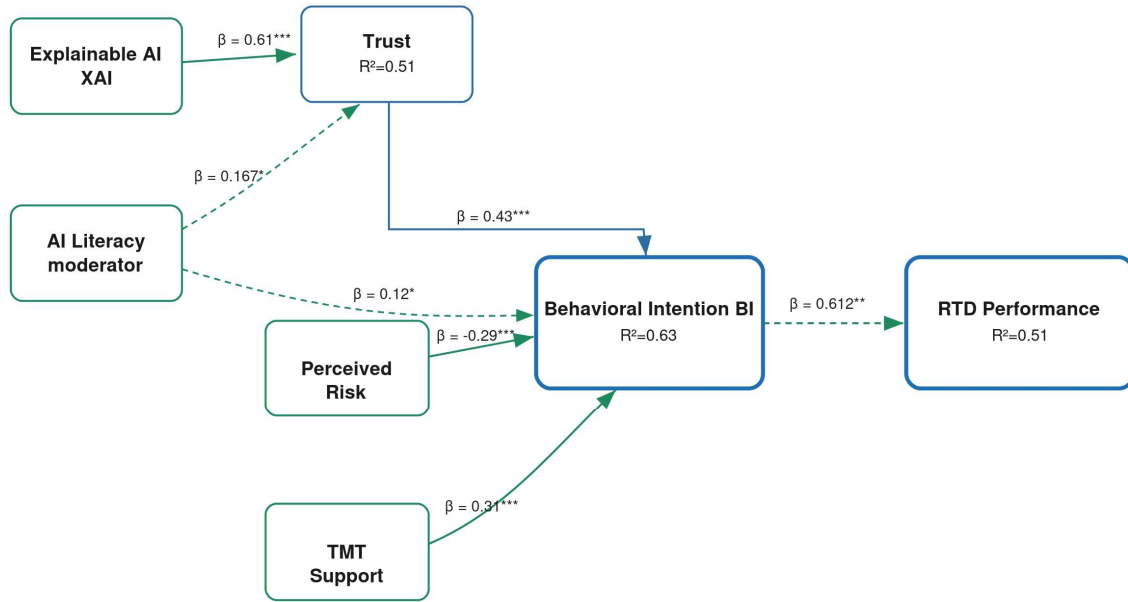


Figure 2. Structural Equation Model and supplementary paths. Solid lines = paths estimated in the primary CB-SEM; dashed lines = moderation or supplementary correlation-based paths reported in Sections 5.4–5.5. Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Fit: CFI = 0.94, TLI = 0.93, RMSEA = 0.058, SRMR = 0.047.

2.31, $t = -2.77$, $p = 0.006$) and attach greater importance to governance (3.02 versus 2.68, $t = -3.11$, $p = 0.002$). This pattern is consistent with the two-layer logic of the proposed architecture: the operational layer (GAVA) is closer to the perspective of IT professionals, while the

The SEM model confirms the findings obtained by regression: XAI $\rightarrow$ Trust ($\beta = 0.61$, $p < 0.001$); Trust $\rightarrow$ BI ($\beta = 0.43$, $p < 0.001$); Risk $\rightarrow$ BI ($\beta = -0.29$, $p < 0.001$); TMT Support $\rightarrow$ BI ($\beta = 0.31$, $p < 0.001$); XAI $\rightarrow$ BI ($\beta = 0.12$, $p = 0.031$). Figure 2 presents the full structural model with

standardised path coefficients.

The path from Behavioural Intention to RTD Performance ($\beta = 0.612$, $p < 0.01$) is presented as supplementary evidence, in accordance with the correlation reported in Section 5.3; it was estimated in a separate regression and is shown in Figure 2 as a dashed supplementary path, in line with the indicative support reported in [6]. The moderation effects of artificial-intelligence literacy (trust path $\beta = 0.167$, $p = 0.012$; supplementary BI path $\beta = 0.12$, $p = 0.041$) are also shown in Figure 2 as dashed lines and derive from the moderation specification presented in Section 5.5.

Although Cronbach's α for Trust in the TMT subsample was lower ($\alpha = 0.56$), composite reliability weighted by standardised loadings reached the conventional threshold ($CR = 0.82$), so the construct remains acceptable for structural analysis despite the noise inherent in small samples.

Table 9. Composite reliability (CR) and average variance extracted (AVE) by subsample.

Construct	CR (IT)	AVE (IT)	CR (TMT)	AVE (TMT)
XAI	0.88	0.78	0.91	0.78
Trust	0.93	0.87	0.82	0.70
Perceived Risk	0.90	0.65	0.88	0.61
TMT Support	0.93	0.76	0.92	0.67
Behavioural Intention	0.91	0.77	0.97	0.91
RTD Performance	0.93	0.75	0.94	0.75

To supplement the reliability coefficients reported in Table 4, composite reliability (CR) and average variance extracted (AVE) were calculated for the six multi-item reflective constructs on the basis of the standardised measurement model (Table 9). All values exceed the recommended thresholds ($CR > 0.70$ and $AVE > 0.50$) [27], and standardised factor loadings range from 0.67 to 0.96, indicating satisfactory internal consistency and convergent validity in both subsamples. The AI Literacy construct, measured by a single self-assessment item, is treated as an observed control variable and is therefore omitted from the convergent-validity analysis. The weaker overall fit of the TMT subsample, noted above ($N = 85$; $CFI = 0.85$, $RMSEA = 0.11$), reflects the reduced stability of covariance-based estimation in small samples, rather than a deficiency of the measurement instrument, whose CR and AVE values remain satisfactory in both groups.

5.8 Summary of Empirical Findings

Four main findings emerge from the empirical analysis. First, explainability is a strong predictor of trust ($\beta = 0.61$). Second, trust mediates the

effect of explainability on behavioural intention (indirect effect 0.102, CI [0.021, 0.187]). Third, perceived risk significantly slows adoption ($\beta = -0.29$). Fourth, top-management support significantly promotes it ($\beta = 0.31$).

Taken together, these results provide quantitative support for the central assumptions of the integrated GAVA–PRIME–INSPECT architecture. They show that validation-centric mechanisms — explainability, trust calibration, risk management, and governance more broadly — jointly shape the performance of AI-supported decision-making. At the same time, the results extend the findings of Avramović et al. [6], who reported only descriptive statistics on the same sample, by identifying the structural relationships assumed by the original framework.

These relationships map directly onto the GAVA loop and its validation subdimensions. Explainability (XAI) operationalises the Validate phase by providing intelligible evidence on which validation rests. Trust corresponds to the cognitive subdimension of validation and to the trust-calibration mechanism in the INSPECT layer. Perceived risk reflects governance validation and the regulatory boundaries that constrain autonomous action. Top-management support represents the organisational conditions that enable the Adapt phase and the supervisory function. RTD Performance captures the overall efficiency of the closed GAVA–PRIME–INSPECT loop. The empirical model therefore examines the socio-technical assumptions of validation-centric decision-making, rather than the lower-level operational parameters of the loop — such as validation latency or the dynamic adjustment of validation criteria — which remain the subject of future engineering evaluation.

The mapping of estimates onto the hypotheses is as follows: hypothesis H1 ($XAI \rightarrow Trust$, $\beta = 0.61$) is strongly supported; hypothesis H3 is supported as partial mediation, with trust carrying the largest share of the effect of explainability on behavioural intention ($Trust \rightarrow BI$, $\beta = 0.43$); the direct path in hypothesis H2 ($XAI \rightarrow BI$, $\beta = 0.12$) is supported, but weakly, which is consistent with this mediation; hypotheses H4 ($Risk \rightarrow BI$, $\beta = -0.29$) and H5 ($TMT Support \rightarrow BI$, $\beta = 0.31$) are supported; and hypothesis H6 (AI Literacy moderates $XAI \rightarrow Trust$, $\beta = 0.167$) is supported in the moderation specification from Section 5.5. The exploratory hypothesis H7 ($Trust \rightarrow RTD Performance$) receives only indicative support through the strong correlation between trust and performance ($r = 0.696$) and is not estimated in the primary CB-SEM model, in accordance with [6].

6. DISCUSSION

6.1 Theoretical Implications

This paper contributes to the literature on AI-native systems by treating validation as an integral part of decision-making itself, rather than as a subsequent verification activity. Whereas earlier research has largely relied on insights derived from data and on model performance, the emphasis here shifts towards the question of whether AI outputs can be trusted while decisions are still being made and implemented.

The integration of the GAVA and PRIME-INSPECT frameworks moves the discussion in several important directions. The fundamental shift is conceptual in nature: validation is no longer treated as an external audit layer, but as part of the decision loop itself. In addition, this integration presents decision-making as a closed-loop process driven by feedback, thereby linking AI-native architectures with decision theory, reinforcement learning, and control systems, while placing governance, trust, and ethical constraints inside the loop rather than outside it.

The empirical findings support the socio-technical mechanisms on which this position rests, while the operational components of the closed loop remain the subject of future direct measurement. The significant effect of explainable artificial intelligence on trust ($\beta = 0.61$, $p < 0.001$) and the central role of trust in shaping behavioural intention ($\beta = 0.43$, $p < 0.001$) show that validation and intelligibility are not secondary attributes, but key drivers of artificial-intelligence adoption. The negative effect of perceived risk ($\beta = -0.29$, $p < 0.001$) and the positive influence of top-management support ($\beta = 0.31$, $p < 0.001$) further confirm the socio-technical character of the proposed framework.

In brief, the results show that effective AI-native systems require more than predictive capability: they require integrated mechanisms of validation, control, and trust calibration. This extends existing models of artificial-intelligence adoption and decision support.

6.2 Practical Implications

From a practical standpoint, the findings have direct consequences for how AI-native systems should be designed and introduced in organisations. Above all, organisations should not build such systems exclusively around prediction and automation. Accuracy remains important, but it is not sufficient when decisions are generated and executed in real time; it is equally important whether the system can explain, verify, and constrain its own outputs while they are being used.

This means that governance cannot be added

only after the system has been introduced. Policies, ethical constraints, supervisory controls, and audit trails must be part of the system architecture from the outset.

The observed role of perceived risk and top-management support points to the same conclusion: the adoption of artificial intelligence depends not only on technical performance, but also on the organisation's ability to manage the risks introduced by automation.

Human oversight, too, should be understood dynamically and not as a final approval step. In validation-centric systems, human interpretation and trust calibration remain part of the loop, especially where decisions carry high-risk operational or organisational consequences.

For organisations, the message is clear: the scaling of AI-native decision systems requires investment in validation capabilities — performance monitoring, explainability tools, drift detection, and compliance checks — together with intervention mechanisms when validation fails. Without such investment, the scaling of systems may also mean the scaling of systemic risk.

6.3 Implications for Data Science

For data science, the implication is immediate. In AI-native environments, the work does not end at the moment when a model produces an accurate output; the more difficult task is to create the conditions under which that output can be monitored, explained, questioned, and corrected.

This shift is also confirmed by the empirically established role of trust. Trust mediates the relationship between explainability and behavioural intention, meaning that users respond not only to what the system recommends, but also to whether they can understand and assess that recommendation.

Data scientists are thereby moving increasingly closer to system governance. Their role increasingly includes validation design, real-time monitoring, bias and drift detection, ensuring explainability, supporting compliance, and defining escalation rules. In this sense, data science is becoming less a discipline of pure modelling and more a discipline of controlled decision-making.

6.4 Limitations

Since this paper combines the development of a conceptual framework with empirical analysis, it is subject to several limitations.

First, although the results support the socio-technical mechanisms of the proposed framework, they are based on cross-sectional data, which limits the possibility of drawing conclusions about causal relationships over time.

In addition, the integration of the GAVA and PRIME-INSPECT frameworks is presented at a relatively high level of abstraction. Such an approach is conducive to general applicability, but may not fully capture domain-specific constraints — regulatory requirements, technical architectures, or risk factors characteristic of particular industries.

The sample, although diverse, is limited to two respondent groups (IT and TMT) and may not represent all stakeholders involved in AI-assisted decision-making processes.

Finally, the implementation of validation-centric systems depends on organisational maturity, data quality, and infrastructural capabilities, which may vary considerably across contexts.

6.5 Future Research

The proposed framework opens several directions for future research.

First, longitudinal studies are needed to examine how mechanisms of trust and validation develop in AI-native systems over time.

Second, trust calibration deserves deeper investigation, particularly with respect to longitudinal dynamics and the dynamics of overreliance, which a cross-sectional design cannot capture.

Further, it would be worthwhile to examine applications of the framework in specific domains, especially in high-risk environments such as manufacturing, finance, healthcare, and energy.

Additional work is also needed on the operationalisation of validation-centric architectures, including the development of metrics, tools, and guidelines for implementation in real-time decision-making environments.

Finally, the changing role of data science in AI-native systems requires closer consideration, particularly with regard to organisational structures, required skills, and interdisciplinary collaboration.

7. CONCLUSION

This paper examined decision-making in AI-native systems, in which artificial intelligence increasingly not only supports human judgement but also independently generates and executes decisions. In response to this shift, a validation-centric framework was proposed, linking the GAVA decision loop with the socio-technical PRIME-INSPECT model [6].

The central argument of the paper is that validation should be regarded as an integral part of the decision-making process itself. In traditional data-driven environments, the main challenge was most often to arrive at useful insights; in AI-native systems, the more difficult question is whether the

decisions they generate can be verified, constrained, and trusted while action is being taken upon them. Embedding validation in a closed decision loop and linking it with governance represents one way of addressing this problem.

The integration of the GAVA and PRIME-INSPECT frameworks is useful precisely because the two perspectives cover different parts of the same challenge. GAVA describes the operational cycle of generation, action, validation, and adaptation, while PRIME-INSPECT adds the governance, trust, and oversight conditions that make that cycle accountable. The empirical results support this interpretation, demonstrating the role of explainability, trust, perceived risk, and top-management support in AI-supported decision-making.

The proposed architecture also comes at a cost. It requires additional infrastructure — for multidimensional validation and for the adaptation of both decision policies and validation criteria — in exchange for gains in reliability, intelligibility, and governance compliance. In high-risk domains, such as mining, finance, smart manufacturing, and healthcare, this trade-off will most often be worthwhile.

In methodological terms, the contribution of the paper can be understood through the method of symbiosis [25]: two previously separate paradigms — operational decision loops and socio-technical governance — are connected to achieve a new capability, real-time validation as a first-order analytical function, which neither of them provides on its own. A secondary method is interdisciplinary application, that is, the transfer of abstractions from reinforcement learning and control theory into the field of artificial-intelligence governance and organisational decision-making.

Taken as a whole, the paper offers a validation-centric perspective on AI-native systems, aligning algorithmic decision-making with organisational and societal requirements. As these systems grow in complexity and autonomy, future research should further focus on empirical validation across different domains, the dynamics of trust calibration, and the development of operational validation frameworks for real-time decision-making environments.

DECLARATION ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

The authors confirm that generative artificial-intelligence tools were used solely to improve the written expression, readability, and linguistic fluency of the text. Their use was limited to sentence reformulation and to improving clarity and coherence. All content produced by these tools was carefully reviewed, edited, and

integrated by the authors, who assume full responsibility for the content of the manuscript.

REFERENCES

- [1] Bommasani, R., Hudson, D. A., Adeli, E., et al., "On the opportunities and risks of foundation models," Stanford Center for Research on Foundation Models (CRFM), arXiv:2108.07258, 2021, doi: 10.48550/arXiv.2108.07258.
- [2] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y., "ReAct: Synergizing reasoning and acting in language models," in International Conference on Learning Representations (ICLR), 2023, arXiv:2210.03629.
- [3] Duan, Y., Edwards, J. S., and Dwivedi, Y. K., "Artificial intelligence for decision making in the era of big data — Evolution, challenges and research agenda," International Journal of Information Management, vol. 48, pp. 63–71, 2019, doi: 10.1016/j.ijinfomgt.2019.01.021.
- [4] Dwivedi, Y. K., Hughes, L., Ismagilova, E., et al., "Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," International Journal of Information Management, vol. 57, art. no. 101994, 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.
- [5] Russell, S. and Norvig, P., Artificial Intelligence: A Modern Approach, 4th ed., global ed. Harlow, U.K.: Pearson, 2021.
- [6] Avramović, N., Marković, A., Čomić, T., Čavoški, S., Zornić, N., and Vujović, V., "PRIME-INSPECT: A Socio-Technical Framework for Trustworthy Intelligent Automation and Real-Time Decision-Making in Industry 4.0," Applied Sciences, vol. 16, no. 10, art. no. 4825, 2026, doi: 10.3390/app16104825.
- [7] Milutinović, V., "The best method for presentation of research results," IEEE TCCA Newsletter, no. 9, pp. 1–6, 1996.
- [8] Bengio, Y., Hinton, G., Yao, A., et al., "Managing extreme AI risks amid rapid progress," Science, vol. 384, no. 6698, pp. 842–845, 2024, doi: 10.1126/science.adn0117.
- [9] Kirilenko, A., Kyle, A. S., Samadi, M., and Tuzun, T., "The flash crash: High-frequency trading in an electronic market," Journal of Finance, vol. 72, no. 3, pp. 967–998, 2017, doi: 10.1111/jofi.12498.
- [10] Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S., "Dissecting racial bias in an algorithm used to manage the health of populations," Science, vol. 366, no. 6464, pp. 447–453, 2019, doi: 10.1126/science.aax2342.
- [11] Floridi, L. and Cowls, J., "A unified framework of five principles for AI in society," Harvard Data Science Review, vol. 1, no. 1, 2019, doi: 10.1162/99608f92.8cd550d1.
- [12] Giudici, P., Centurelli, M., and Turchetta, S., "Artificial intelligence risk measurement," Expert Systems with Applications, vol. 235, art. no. 121220, 2024, doi: 10.1016/j.eswa.2023.121220.
- [13] Selbst, A. D. and Barocas, S., "The intuitive appeal of explainable machines," Fordham Law Review, vol. 87, no. 3, pp. 1085–1139, 2018.
- [14] Shin, D., "User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability," Journal of Broadcasting & Electronic Media, vol. 64, no. 4, pp. 541–565, 2020, doi: 10.1080/08838151.2020.1843357.
- [15] Mnih, V., Kavukcuoglu, K., Silver, D., et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, pp. 529–533, 2015, doi: 10.1038/nature14236.
- [16] Sutton, R. S. and Barto, A. G., Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
- [17] Achiam, J., Held, D., Tamar, A., and Abbeel, P., "Constrained policy optimization," in Proceedings of the 34th International Conference on Machine Learning (ICML), vol. 70, 2017, pp. 22–31, arXiv:1705.10528.
- [18] García, J. and Fernández, F., "A comprehensive survey on safe reinforcement learning," Journal of Machine Learning Research, vol. 16, pp. 1437–1480, 2015.
- [19] Hayes, C. F., Rădulescu, R., Bargiacchi, E., et al., "A practical guide to multi-objective reinforcement learning and planning," Autonomous Agents and Multi-Agent Systems, vol. 36, no. 1, art. no. 26, 2022, doi: 10.1007/s10458-022-09552-y.
- [20] Åström, K. J. and Murray, R. M., Feedback Systems: An Introduction for Scientists and Engineers, 2nd ed. Princeton, NJ, USA: Princeton University Press, 2021.
- [21] Bemporad, A., Morari, M., Dua, V., and Pistikopoulos, E. N., "The explicit linear quadratic regulator for constrained systems," Automatica, vol. 38, no. 1, pp. 3–20, 2002, doi: 10.1016/S0005-1098(01)00174-1.
- [22] Lee, J. D. and See, K. A., "Trust in automation: Designing for appropriate reliance," Human Factors, vol. 46, no. 1, pp. 50–80, 2004, doi: 10.1518/hfes.46.1.50.30392.
- [23] Sculley, D., Holt, G., Golovin, D., et al., "Hidden technical debt in machine learning systems," in Advances in Neural Information Processing Systems (NeurIPS), vol. 28, 2015, pp. 2503–2511.
- [24] Mitchell, M., Wu, S., Zaldivar, A., et al., "Model cards for model reporting," in Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (FAT*), 2019, pp. 220–229, doi: 10.1145/3287560.3287596.
- [25] Blagojević, V., Bojić, D., Bojović, M., Cvetanović, M., Đorđević, J., Đurđević, Đ., et al., "A systematic approach to generation of new ideas for PhD research in computing," in Advances in Computers, vol. 104, A. R. Hurson and V. Milutinović, Eds. Cambridge, MA, USA: Academic Press, 2017, pp. 1–31.
- [26] Glikson, E. and Woolley, A. W., "Human trust in artificial intelligence: Review of empirical research," Academy of Management Annals, vol. 14, no. 2, pp. 627–660, 2020, doi: 10.5465/annals.2018.0057.
- [27] Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E., Multivariate Data Analysis, 8th ed. Andover, U.K.: Cengage Learning, 2019.
- [28] Hu, L. and Bentler, P. M., "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," Structural Equation Modeling, vol. 6, no. 1, pp. 1–55, 1999, doi: 10.1080/10705519909540118.

AUTHOR BIOGRAPHIES

Nebojša Avramović

Nebojša Avramović is the founder and CEO of Consulting IT DNA, a company specialised in business and digital transformations through the application of artificial intelligence and modern technologies. He graduated from the Military Technical Academy in Belgrade and is currently a doctoral candidate at the Faculty of Organisational Sciences, University of Belgrade, in the Information Systems and Quantitative Management programme. Throughout his career he has held management positions at Oracle, Ingram Micro, and IBM. He has co-authored papers on agent-based models, artificial intelligence, and digital technologies in business systems. E-mail: nebojsa.avramovic@consultingitdna.rs, ORCID: 0009-0009-2565-7176.

Tijana Čomić

Tijana Čomić, PhD, is an Assistant Professor and researcher specializing in statistics, quantitative analysis, survey methodology, and official statistics. She earned her PhD from the University of Belgrade, Faculty of Organisational Sciences, with a dissertation on improving official statistics through Big Data applications. Her academic work includes publications in international journals and conference proceedings, with research interests spanning sustainable development indicators, survey methodology, disability statistics, artificial intelligence, and data-driven decision-making. She has extensive experience in teaching quantitative subjects and applying advanced statistical methods in international research and development projects. E-mail: tijana.comic@vos.edu.rs, ORCID: 0000-0001-6556-5879.

Aleksandar Marković

Aleksandar Marković is a full professor at the Faculty of Organisational Sciences, University of Belgrade. He teaches Simulation in Business Decision Making, Financial Simulation Modeling, Business Dynamics, and E-business Management. He served as Vice-Dean for Organisation and Finance (2016–2021) and Editor-in-Chief of the international journal Management (2007–2016). His research focuses on business systems modelling, computer simulation, system dynamics, and e-business. He has published more than 70 scientific papers. E-mail: aleksandar.markovic@fon.bg.ac.rs, ORCID: 0000-0002-3976-2788.

Sava Čavoški

Sava Čavoški is an Assistant Professor at the School of Computing, Union University. His professional and research interests focus on computer simulation, agent-based modelling and simulation, machine learning, and AI-supported decision-support tools. He has experience in the design and development of complex enterprise software systems, scalable information systems, and cloud-based applications. His current work focuses on generative agent-based modelling (GABM), the strategic development of information systems, and the integration of artificial intelligence into business decision-making and digital transformation processes. E-mail: scavoski@raf.rs, ORCID: 0000-0003-1163-4466.

Nikola Zornić

Nikola Zornić, PhD, is an Assistant Professor at the Faculty of Organisational Sciences, University of Belgrade, where he received his bachelor's and master's degrees and defended his doctoral dissertation in 2023 on a methodological framework for integrating machine learning algorithms into agent-based simulation models. His research interests span agent-based modelling and simulation, machine learning, scientometrics, and business analytics, with current work focused on generative agent-based modelling (GABM). He has participated in several European and national research projects and has served on the Faculty Council since 2022. E-mail: nikola.zornic@fon.bg.ac.rs, ORCID: 0000-0002-3597-0627.

Vladimir Vujović

Vladimir Vujović is an assistant professor at the Faculty of Electrical Engineering, University of East Sarajevo, in the field of Computer Science. He has authored more than 80 scientific and professional papers, participated in several research projects, and is an active member of professional, scientific, and conference committees. His professional interests focus on software architecture, business analysis, and strategic development of information. E-mail: systems.vladimir.vujovic@etf.ues.rs.ba, ORCID: 0000-0001-8175-5058.

Forward-Secure Blockchain Architecture for National-Scale Electronic Voting: Enhancing Institutional Sustainability and Democratic Resilience

Stanojević, Nebojša; Vukmirović, Dragan; Čomić, Tijana; Stanojević, Dejan; and Četković, Miloš

Abstract: National electronic voting systems require robust auditability, non-repudiation, forward secrecy, and fault tolerance under predictable but highly variable workloads. Existing protocols within permissioned blockchain networks usually favor low latency or long-lasting identities of business entities, which are assumptions that do not fully correspond to adversarial election environments.

This paper presents a feasibility study of a time-slot-based permissioned blockchain architecture designed for state elections involving semi-trusted but mutually adversarial validators. The basic mechanism of the system is a block authentication scheme with advanced security, based on ephemeral key pairs: validators publish the public key for block $N+1$ within block N , after which they use the corresponding private key once to sign the next block and immediately delete it permanently.

As part of the research, a multi-node prototype was implemented as a customized infrastructure of permissioned blockchain in the Node.js environment, with the application of Ed25519 digital signatures, explicit block finalization, peer-to-peer communication between validators and full signature verification by all participants. Empirical evaluation in a simulated wide-area network (WAN) environment shows a stable throughput of approximately 274 transactions per second, with a median finalization (p50) of 10.7 seconds and a 99th percentile (p99) of under 21 seconds.

A projection on an election scenario involving six million voters during the 13:00 voting period indicates an average safety factor load of approximately 2.1. The presented results of the proposed model should be interpreted as a proof of concept obtained in a controlled prototype environment, and not as a complete validation of the production capacity on a geographically distributed election infrastructure. The findings indicate technical feasibility under defined assumptions and demonstrate how domain-specific transaction clustering, ephemeral key authentication, and explicit validator signatures can support an auditable consensus layer for electronic voting. Out of scope for this current study are client-side privacy protection, coercion resistance, and full end-to-end ballot secrecy. Our structural architecture explicitly hardens the consensus and ledger layers instead.

Index Terms: electronic voting; permissioned blockchain; forward security; ephemeral keys;

auditability consensus protocols; blockchain infrastructure; scalability evaluation

1. INTRODUCTION

Electronic voting promises to make public elections both user-friendly and straightforward to audit, yet actual national-scale deployments remain rare. The fundamental hurdle is temporal: an election system must operate flawlessly within a narrow, legally mandated voting window, while its technical records must remain irrefutably convincing long after the polls close. Because legal disputes often surface weeks or months after the fact, the underlying data logs must remain fully verifiable over time. While blockchain technology is frequently proposed to secure these records through append-only structures and distributed validation, standard implementations fail to safeguard the system against long-term vulnerabilities.

To build this unalterable archive, researchers frequently turn to blockchain infrastructure, drawn by its decentralized validation and append-only architecture. However, standard blockchain design is not always suitable for elections. Public blockchains solve different problems because they rely on open participation and economic incentives. On the other hand, private blockchains (with access permissions) are closer to the needs of public institutions. However, they usually rely on well-known validators, whose signing keys are valid for a longer period of time. It is not an ideal solution. Validators can be politically opposed and are also attractive targets for attacks.

This paper addresses a specific problem: the subsequent compromise of validator keys. What if the same key is used for multiple blocks or over a long period? An attacker who gets to it after the election can create a fake history. At the very least, it may cast doubt on the authenticity of the electoral register. In business systems, this is handled by contracts and internal procedures. However, with elections, the consequences for society and politics are incomparably greater.

That is why we study small and controlled consensus layer design. In it, the block signing key lasts for a very short time. Our idea is not to build a complete e-voting system. Voter authentication, secrecy, coercion resistance, and practicality are left to the application layer. Our goal is narrower. We want to examine whether a permissioned blockchain can use simple block authentication with forward secrecy while maintaining a realistic data processing speed at the national level.

The paper has four contributions.

First, it proposes a one-block-ahead ephemeral-key mechanism for block authentication.

Second, it uses a slot-based workflow that fits the bounded and bursty nature of voting.

Third, the architecture retains explicit validator signatures inside the blocks, allowing independent observers to audit the entire chain long after election day.

Fourth, our experimental evaluation transparently outlines the scope of the prototype, distinguishing the fully coded components from the parameters that were simulated.

2. RELATED WORK

2.1. Permissioned Blockchain Consensus

BFT consensus protocols like PBFT, Tendermint, and HotStuff serve as standard foundations for permissioned ledgers, providing deterministic finality under bounded Byzantine failures [5], [6], [17]. Advanced architectures like Narwhal and Bullshark further boost throughput by decoupled processing of data dissemination and ordering [8]. While highly effective, these frameworks fundamentally assume that validator keys remain secure and unaltered under the broader system governance.

That perspective holds true across many enterprise setups where network nodes are run by corporations bound by legal contracts, making key rotation a routine operational task. When applied to public elections, however, this assumption becomes highly problematic. Even if validators represent established public institutions or political entities, they rarely enjoy universal trust from every voting faction. Consequently, if a long-term signing key is compromised, it triggers a widespread crisis of public legitimacy rather than a simple technical failure.

Hyperledger Fabric serves as a prominent enterprise permissioned blockchain framework [2], utilizing an adaptable architecture that uncouples endorsement, ordering, and validation tasks. It is useful in business. However, this also means that public review depends on complex access and configuration rules. In our design, audit proofs are intentionally closer to the block itself. Each block contains explicit validator signatures. It also contains the information necessary to check the transition of the ephemeral (short-lived) key.

2.2. Blockchain-Based Electronic Voting

Electronic voting is not a new phenomenon; it has been the subject of research for many years. The literature in this area generally indicates that the

accuracy of election results is not the only requirement for a reliable electoral system. Equally important are voter privacy, citizen trust, and the ability to subsequently verify the results and the manner in which the electoral process is conducted [4], [7]. One example is Helios, a web-voting system that allows for open audit trails [1]. More recent protocols, such as VERICONDOR, also improve end-to-end verifiability for certain electoral rules [9].

Several authors warn that blockchain is not a complete solution for voting [13]. It makes records resistant to unauthorized changes. However, blockchain alone does not protect the voter's device. It does not prevent coercion, does not hide the vote and does not decide who has the right to vote. We agree with that view. Our work does not solve all voting problems. We focus exclusively on consensus records. The main question is how that record remains reliable if the validator keys are subsequently compromised.

2.3. Forward Security and Key Compromise

Forward-secure signatures have been proposed to reduce the damage caused by key compromise [3], [10]. Fundamentally, exposing the current key material must not grant an attacker the ability to forge valid signatures for historical epochs. While certain formal cryptographic frameworks offer robust security guarantees, they often bring substantial operational complexity to the actual deployment.

Alternatively, simpler strategies achieve practical forward security by restricting key lifetimes and strictly regulating their exposure windows [11]. Our design follows this engineering-oriented path. Rather than introducing a full key-evolving signature scheme, each block is assigned a short-lived signing key whose public component is stored one block in advance.

The transition to post-quantum cryptography makes this topic even more relevant. Measurements of post-quantum signatures in blockchain environments show that signature size and verification costs can affect throughput, propagation, and storage [15]. Our prototype is not post-quantum secure, since it uses Ed25519. The same architecture, however, can be used to study where post-quantum signatures would create additional load in an electoral register.

2.4. Research Gap

The existing literature already discusses developed BFT consensus protocols, verifiable voting protocols, and cryptographic mechanisms that mitigate the consequences of subsequent key discovery. However, it is evident that a permissioned electoral registry model specifically adapted to this environment is missing. Such a system would need to combine predictable block finality, clear evidence for subsequent revision, and limited exposure of validator keys at the block level. This work seeks to fill this gap through the development and testing of a prototype. System and threat model

3. SYSTEM AND THREAT MODEL

3.1. System Model

The proposed system is conceived as a permissioned distributed ledger, designed specifically for the loads characteristic of national elections [13], [14], [16]. The system architecture includes three types of entities: validators, clients and observers. Validators form a fixed set of consensus nodes managed by mutually distrustful institutional actors. These roles may be filled by the electoral commission, political parties, accredited auditors, or civil society groups. Clients submit voting transactions directly, while observers function as read-only nodes that replicate the ledger to verify cryptographic signatures and key transitions.

Our model operates under a partially synchronous network assumption. This means that while individual messages may face temporary transmission delays, their ultimate delivery remains guaranteed within an unspecified, finite time bound. The system assumes a standard BFT threshold at which security is maintained while the number of Byzantine validators satisfies $n \geq 3f + 1$ [6], [12], [17]. The long-term identity keys of the validators are known via the initial public key infrastructure (PKI) but are not used for conventional block signing. Consensus operations rely on short-lived (ephemeral) keys.

3.2. Threat Model and Scope

The threat model assumes that an adversary can control up to f validators, distribute messages within defined partial synchronicity limits, cause individual nodes to crash, or compromise the validator's internal state at time t . The main security goal of the system is to preserve the integrity of the ledger: after a certain block is finalized, later compromise of the validator should not enable retroactive falsification of blocks that were finalized before the very moment of compromise [3], [11].

This paper primarily focuses on the consensus and authentication layer of the ledger. The proposed solution does not include voter anonymity, resistance to coercion, the operation of malicious software (malware) on the client side, the usability of the voting system, or legal identity verification. The above issues are considered orthogonal and must be addressed within the wider application architecture of e-voting, for example, by implementing blind signatures, mix-nets, zero-knowledge proofs or other verifiable voting mechanisms [1], [4], [7], [9].

The temporal execution of these phases is depicted in Figure 2, which emphasizes that the private key esk_{t+1} remains in the 'Exposure Window' only for the duration of a single slot before the Secure Erase operation renders it permanently unrecoverable [10].

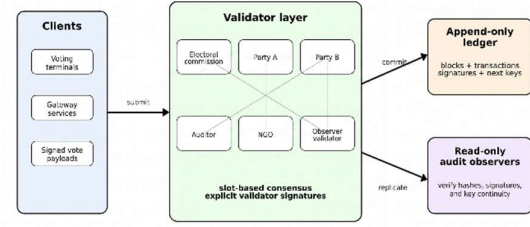


Figure 1. Permissioned blockchain infrastructure for national-scale electronic voting

4.1. Blockchain Infrastructure

The revised implementation is explicitly defined as a custom permissioned blockchain prototype, not as a solution implemented on Hyperledger Fabric [2], Tendermint [5], or any other existing framework. The reason for this design approach lies in the need for experimental control: the developed prototype isolates the proposed authentication mechanism with ephemeral keys and allows the block structure, validator signatures, as well as the life cycle of the keys, to be directly observable and amenable to analysis.

The infrastructure of the subject blockchain consists of an append-only block log, a deterministic slot scheduler, a peer-to-peer communication layer between validators, a transaction pool (mempool), block header verification logic, aggregation of multiple validator signatures, and an ephemeral key cache. Each block structurally contains the height of the block, the hash of the previous block, the group of transactions, the identity of the proposer, the public ephemeral keys for the next block and the explicit signatures of the validators. Validators commit a block only after successfully verifying the previous hash, transaction format, quorum signature and continuity of public ephemeral keys.

4.2. Slot-Based Consensus Model

The time frame of the system is divided into discrete slots S_t . Within each slot, the deterministic leader selection rule assigns exactly one validator with the task of proposing a block B_t . Such a design based on time slots is adequate for application in electronic voting systems, since election traffic is by its nature variable, but legally and time-limited. The primary goal of the proposed solution is not to achieve the maximum theoretical throughput in all operational conditions, but to ensure predictable finalization and stable audit evidence during a defined voting window. Machines within a low-latency institutional network.

4.3. Forward-Structure Block Authentication

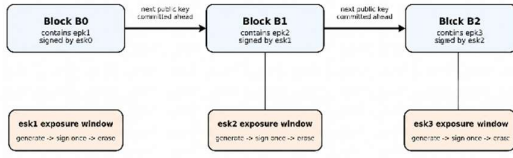


Figure 2. Forward-secure ephemeral-key chaining across consecutive blocks

For consensus operations, each validator maintains a set of ephemeral key pairs. During slot t , validator v_i generates a fresh pair $(epk_{\{i,t+1\}}, esk_{\{i,t+1\}})$ for the subsequent slot. The public key $epk_{\{i,t+1\}}$ is permanently recorded inside block B_t . When the next slot arrives, the validator utilizes $esk_{\{i,t+1\}}$ exactly once to sign the header of B_{t+1} before immediately and securely deleting that specific private key [10].

If a validator is compromised after its private key has been deleted, the attacker can retrieve current or future key material but cannot reconstruct the erased keys from earlier blocks. This approach effectively narrows the vulnerability window and significantly reduces the feasibility of long-range block forgery [3], [11]. While this mechanism provides a highly practical form of forward security suitable for permissioned ledgers, it does not replace formal post-quantum signatures [15] or application-level voting privacy.

4.4. Fault-Tolerant Key Synchronization

This key chain mechanism also introduces a liveness challenge. If a validator misses B_t , it may lack the public key necessary to verify B_{t+1} . To mitigate this risk, validators maintain a short local cache of recently announced public keys. When a required key is missing, the validator requests it from peers and verifies its validity against the relevant block hash before processing it. This bounded retry mechanism is designed to handle moderate network message failures and brief partitions.

4.5. Auditability Through Explicit Signatures

Each committed block carries the explicit cryptographic signatures of a quorum of validators, satisfying the $n \geq 3f + 1$ safety threshold [6]. Observers with read-only access privileges can autonomously initiate the verification process directly from the genesis block and confirm the absolute correctness of the ledger by verifying the hash chains, block heights, validator signatures, and the chronological continuity of the ephemeral keys. This design approach deliberately avoids relying on an unobservable or internal consensus state as the sole proof of validity, making the distributed ledger significantly more amenable to independent public review and post-election auditing [13].

5. IMPLEMENTATION

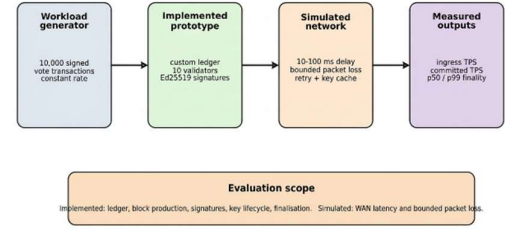


Figure 3. Experimental evaluation workflow and scope

5.1. Software Stack and Deployed Components

The proposed prototype is implemented in the Node.js v20 environment. Its event-driven execution model efficiently supports concurrent transaction dispatch, block propagation, and signature verification without the need to introduce specialized middleware. Digital signatures are implemented using the Ed25519 algorithm via the integrated Node.js crypto module. In terms of memory management, each validator keeps only one active ephemeral private key in working memory and deletes it immediately after successfully signing the next block's header.

Communication between validators relies on HTTP-based peer-to-peer messaging with JSON-encoded packets. Each validator exposes specific endpoints dedicated to submitting transactions, propagating blocks, exchanging signatures, and retrieving public ephemeral keys for the next block. The state of the ledger during the execution of the benchmark test is maintained exclusively in the operational memory and duplicated in the log only for attachment, thus ensuring the reproducibility of the experiment and the possibility of subsequent audit inspection.

Table 1. Implementation scope and evaluation purpose of prototype components

Component	Implemented in prototype		Purpose
Permissioned validator nodes	Yes: 10 independent validator instances		Block proposal, verification, signature exchange, and ledger commit
Append-only ledger	Yes: in-memory state with append-only persistence log		Stores committed blocks and transaction batches
Ephemeral-key lifecycle	Yes: generate, publish, sign once, erase		Enforces bound key exposure and forward security
Network delay/loss	Simulated within		Evaluates behaviour under non-

Component	Implemented in prototype		Purpose
	application logic		ideal wide-area conditions
Client-side ballot secrecy	Out of scope		Assumed to be handled by the voting application layer

5.2. Network Simulation and Benchmark Orchestration

Network effects are simulated within the application logic itself. Each message exchanged between validators was subject to an artificial delay, which was randomly taken from a range of 10 ms to 100 ms, as well as probabilistically bounding packet loss. This segment of the experiment therefore represents a controlled simulation rather than a geographically distributed field application. On the other hand, the blockchain-prototype implementation itself included block creation, transaction grouping, validator signatures, ephemeral key publishing, key erasure, signature verification, and ledger commit — and was fully functional and performed in its native form during the entire benchmark test.

5.3. Hardware Configuration

The experimental evaluation was performed on a standard server configuration equipped with 8 vCPU, 16 GB RAM and Ubuntu Linux 22.04 LTS operating system. We selected this configuration because public-sector systems typically operate with limited infrastructure. While these results should not be interpreted as a guarantee for production deployment, they demonstrate that the proposed mechanism is not inherently too demanding for standard server hardware.

6. EXPERIMENTAL EVALUATION

6.1. Experimental Setup and Measurement Scope

The evaluation utilized ten validator instances and a workload of 10,000 signed voting transactions. Our primary objective was to determine whether the ephemeral-key mechanism and explicit signature verification introduce prohibitive computational overhead. Additionally, the experiment provides an initial estimate of throughput and finality, allowing for a direct comparison against the average load requirements of a national-scale election.

6.2. Scope and Evaluation

The evaluation should be understood as a controlled prototype test. It measures the implemented slot workflow, block creation, signature exchange, ephemeral-key publication, key erasure, signature verification, and ledger commit. It also measures behavior under configured latency and message failure.

At the same time, it does not fully reproduce all the real-world properties of an actual election network.

The validators were not deployed across geographically diverse data centers or operated by independent institutions. Furthermore, the experiment does not account for authentic inter-institutional routing paths, heterogeneous hardware environments, large-scale DDoS attacks, operational monitoring tools, formal legal certification, or live, voter-facing public infrastructure.

For that reason, the results serve as evidence of feasibility rather than full production validation. A real deployment would require multi-site testing, independent operators, more rigorous adversarial testing, and formal certification.

Table 2. Configuration parameters and metrics for experimental evaluation

Parameter	Value/description
Number of validators	10
Fault threshold assumption	$n \geq 3f + 1$; with $n = 10$, the design corresponds to $f = 3$ Byzantine validators
Workload	10,000 signed voting transactions
Network delay	Simulated uniform delay between 10 ms and 100 ms
Primary metrics	Ingress TPS, committed TPS, p50 finality, p99 finality
Implemented blockchain functions	Block production, signature collection, signature verification, ephemeral-key lifecycle, ledger commit
Simulation boundary	Inter-validator latency and bounded message loss

6.3. Throughput and Finality

The prototype reached a peak in ingress throughput of approximately 583 transactions per second, while the sustained committed throughput was about 274 transactions per second. This difference is entirely expected, as the committed throughput factors in the additional overhead of block creation, consensus message propagation, validator signature collection, verification, and the sequential sign-and-erase procedure.

Finality values were also within a range that appears highly practical for voting. The median finality was 10.7 seconds, and p99 finality stayed below 21 seconds within the simulated network setting. While this is not comparable to high-frequency financial systems that target sub-second settlement times, the more critical requirement in an election context is ensuring that each vote is durably recorded and fully auditable at a later date.

Table 3. Summary of experimental results, observed metrics, and interpretation

Metric	Observed value	Interpretation
Peak ingress throughput	Approx. 583 TPS	The rate at which validator APIs accepted valid transactions
Committed throughput	Approx. 274 TPS	The rate at which transactions were finalised in the ledger

Metric	Observed value	Interpretation
p50 finality	10.7 seconds	Median time from submission to irreversible block inclusion
p99 finality	< 21 seconds	Tail latency under simulated network stress
Cryptographic overhead	Acceptable in the evaluated setting	Ed25519 signing and verification did not become a prohibitive bottleneck

6.4. . Election-Scale Projection

To establish a practical reference point, the measured committed throughput was compared against a straightforward national election scenario involving six million voters and a 13-hour voting window. Under a uniform-load approximation, this setup requires a processing capacity of approximately 128 transactions per second. Consequently, the measured sustained capacity of 274 committed transactions per second yields an average-load safety margin of roughly 2.1.

While that projection is useful, it remains a simplified model. Real-world elections inevitably experience pronounced traffic spikes, particularly during the morning rush and right before polling stations close. Future testing should therefore incorporate non-uniform transaction arrival patterns and be executed across true, geographically distributed infrastructure.

Table 4. National-scale election scenario parameters and prototype validation results

Scenario element	Value
Electorate size	6,000,000 voters
Voting window	13 hours = 46,800 seconds
Average required throughput	Approx. 128 TPS
Measured committed throughput	Approx. 274 TPS
Average-load safety factor	Approx. 2.1
Validation status	Feasibility evidence under controlled prototype and simulated network conditions

7. DISCUSSION

The main takeaway is that the proposed block-level forward security mechanism did not render the prototype too slow for the analyzed average workload. The trade-off is higher latency compared to some generic BFT (Byzantine Fault Tolerant) systems. Within a voting context, however, this compromise can be acceptable, as a finality time of several seconds is usually far less problematic than weak or unreliable audit evidence.

The architecture is intentionally narrower in scope than general-purpose blockchain platforms. It is tailored specifically for use cases where the validators are known, the voting period is strictly bounded, and the long-term integrity of the historical record remains paramount long after the event has concluded. This specific focus explains why each block explicitly carries validator signatures alongside

cryptographic information for the subsequent ephemeral public keys.

Ultimately, this evaluation does not prove that the system is ready for a national election; it merely demonstrates that the underlying mechanism is feasible within the tested prototype. Substantiating stronger claims would require side-by-side comparisons with alternative platforms under identical workloads, true distributed deployments, peak-load stress testing, dedicated DDoS experiments, and an independent, third-party security review.

8. LIMITATIONS AND FUTURE WORK

Several open limitations were identified within this research. First, the developed prototype was evaluated on a single regular server running multiple instances of the validator, which implies that a geographically distributed application must be tested separately. Second, network latency and message losses are simulated rather than generated within the actual broadband election network. Third, the projected calculation for six million voters uses the average load model, and in the following stages, it should be expanded with empirical traffic distributions during peak hours. Fourth, the proposed architecture does not solve the issues of vote secrecy, resistance to coercion, or the operation of malicious software (malware) on the client side [13]. Fifth, the system design currently uses the Ed25519 scheme [10] and needs to be evaluated for the integration of post-quantum signature schemes as the technology matures [15].

Future work will therefore primarily focus on distributed multi-site testing, formal ephemeral key machine state verification, comparative analysis with Tendermint [5]/HotStuff-style [17] foundations under equivalent operational loads, integration with ballot privacy protocols [1], [4], [7], [9], as well as detailed evaluation of system behaviour under DDoS attack and node crash recovery scenarios [8].

9. CONCLUSION

This paper presents a feasibility study of a forward-secure (with advanced key security) permissioned blockchain architecture intended for electronic voting at the national level [13], [14]. The proposed architecture successfully combines slot-based block production, explicit multi-validator signatures [6], and a specific ephemeral key mechanism in which public keys are committed one block in advance, while private keys are used once and then permanently deleted [10]. The described design directly addresses a specific risk at the consensus level, which refers to the retroactive falsification of the ledger after the possible compromise of the keys of the validators themselves [3], [11].

The developed custom multi-node prototype demonstrated sustainable commit throughput of approximately 274 TPS and p99 finalization under 21 seconds under simulated broadband network

latency. A projection on an election scenario involving six million voters and a 13-hour voting window provides an average safety factor of approximately 2.1. The stated results provide empirical support for the feasibility of the described approach under the evaluated assumptions, with a simultaneous indication of a clear need for geographically distributed field testing, stronger adversarial evaluation [8], formal security assessment, as well as integration with application-level privacy and verifiability mechanisms [1], [4], [7], [9] before any production deployment in a real environment.

REFERENCES

- [1] B. Adida, "Helios: Web-based open-audit voting," in Proc. 17th USENIX Security Symp., 2008, pp. 335–348.
- [2] E. Androulaki, A. Barger, V. Bortnikov, et al., "Hyperledger Fabric: A distributed operating system for permissioned blockchains," in Proc. 13th EuroSys Conf., 2018. doi: 10.1145/3190508.3190538.
- [3] M. Bellare and S. Miner, "A forward-secure digital signature scheme," in Advances in Cryptology—CRYPTO '99, 1999, pp. 431–448. doi: 10.1007/3-540-48405-1_28.
- [4] J. Benaloh, "Verifiable secret-ballot elections," PhD dissertation, Yale Univ., 1987.
- [5] E. Buchman, "Tendermint: Byzantine fault tolerance in the age of blockchains," M.S. thesis, Univ. of Guelph, 2016.
- [6] M. Castro and B. Liskov, "Practical Byzantine fault tolerance," in Proc. 3rd Symp. Operating Syst. Des. Implementation, 1999, pp. 173–186.
- [7] D. Chaum, P. Y. A. Ryan, and S. Schneider, "A practical voter-verifiable election scheme," in European Symp. Res. Comput. Security, 2005, pp. 118–139. doi: 10.1007/11555827_7.
- [8] G. Danezis, E. Kokoris-Kogias, A. Sonnino, and A. Spiegelman, "Narwhal and Bullshark: A DAG-based mempool and efficient BFT consensus," in Proc. 31st USENIX Security Symp., 2022, pp. 1971–1988.
- [9] L. Harrison, S. Bag, H. Luo, and F. Hao, "VERICONDOR: End-to-end verifiable Condorcet voting with support for strict preference and indifference," ACM Digital Threats: Research and Practice, vol. 5, no. 4, pp. 1–30, 2024. doi: 10.1145/3676267.
- [10] G. Itkis and L. Reyzin, "Forward-secure signatures with optimal signing and verifying," in Advances in Cryptology—CRYPTO 2001, 2001, pp. 332–354. doi: 10.1007/3-540-44647-8_20.
- [11] H. Krawczyk, "Simple forward-secure signatures from any signature scheme," in Proc. 7th ACM Conf. Comput. Commun. Security, 2000, pp. 108–115. doi: 10.1145/352600.352626.
- [12] L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," ACM Trans. Program. Lang. Syst., vol. 4, no. 3, pp. 382–401, 1982. doi: 10.1145/357172.357176.
- [13] S. Park, Y. Kwon, and G. Fuchsbaue, "Is blockchain suitable for electronic voting?" IEEE Security & Privacy, vol. 18, no. 1, pp. 70–77, 2020. doi: 10.1109/MSEC.2019.2955709.
- [14] M. Pilkington, "Blockchain technology: Principles and applications," in Research Handbook on Digital Transformations, X. Olleros and M. Zhegu, Eds. Edward Elgar, 2016, pp. 225–253.
- [15] A. G. Schemitt, H. F. da Silva, R. C. Lunardi, D. Kreutz, R. B. Mansilha, and A. F. Zorzo, "Assessing the impact of post-quantum digital signature algorithms on blockchains," arXiv:2510.09271, 2025.
- [16] M. Swan, Blockchain: Blueprint for a New Economy. Sebastopol, CA: O'Reilly Media, 2015.

- [17] M. Yin, D. Malkhi, M. K. Reiter, G. G. Gueta, and I. Abraham, "HotStuff: BFT consensus with linearity and responsiveness," in Proc. 2019 ACM Symp. Principles Distrib. Comput., 2019, pp. 347–356. doi: 10.1145/3293611.3331591.
- [18] D. Vukmirović, S. Radojičić, N. Stanojević, N. Dragović, and T. Čomić, "Enhancing critical infrastructure resilience: An integrated framework for hazardous material management of electric vehicle batteries," in Critical Infrastructure: Risk and Crisis Management in the New Reality, Z. Keković, S. Kentera, M. Dabetić, and S. Grbović, Eds. Springer, 2026, pp. 81–109. doi: 10.1007/978-3-032-20948-1_5.ž.

Nebojša Stanojević holds a master's degree in Geography, is a PhD student at the Faculty of Organizational Sciences, University of Belgrade, and works as a high school teacher and a data analyst in automotive parts distribution. He is also the founder and software architect at INFOS-N, an IT services agency, where he has extensive experience in designing and implementing numerous domestic and national-scale projects. His research and professional interests are mainly focused on blockchain technology, geographic information systems (GIS), data analytics, and the application of generative artificial intelligence in business and education. E-mail: ns20205003@student.fon.bg.ac.rs, ORCID: 0009-0004-4418-0039.

Dragan Vukmirović is a full professor at the Faculty of Organizational Sciences, University of Belgrade, in the fields of statistics, data science, e-business, and digital transformation. He is a former Director General of the Statistical Office of the Republic of Serbia and has extensive experience in official statistics, statistical methodology, information systems, and evidence-based decision-making. He is the author and co-author of numerous scientific and professional papers and has participated in national and international projects. His research interests include statistics, data science, e-government, digital inclusion, artificial intelligence, and the application of generative AI in education, business, and public administration. E-mail: dragan.vukmirovic@fon.bg.ac.rs, ORCID: 0000-0003-0248-016X.

Tijana Čomić is an assistant professor and statistician with a PhD in Statistics and extensive experience in household surveys, official statistics, and international research projects. Her professional work includes survey methodology, data analysis, Multiple Indicator Cluster Surveys (MICS), EU-SILC, GEDSI indicators, and evidence-based policy support. She has worked as an international consultant and has contributed to research and capacity-building activities in the field of statistics and development. Her recent work also includes the practical application of generative artificial intelligence and educational workshops in this emerging field. Her research interests include statistics, survey methodology, social indicators, Big Data, and generative AI. E-mail: tijana.comic@vos.edu.rs, ORCID: 0000-0001-6556-5879.

Dejan Stanojević is an Information Systems Engineer who graduated from the Faculty of Organizational Sciences, University of Belgrade. He has more than 35 years of professional experience in the development of complex business, geospatial, analytical, and decision-support systems for public administration, utilities, banking, telecommunications, healthcare, pharmaceutical analytics, and automotive parts distribution. He is the software architect at InSist doo and has worked remotely on international projects for European and North American companies and clients, with a particular focus on the architecture and implementation of large-scale information systems.

Miloš Cetković holds a degree in Management and a master's degree in Business Analytics from the Faculty of Organizational Sciences, University of Belgrade. His academic and research interests focus on retail development

strategies, omnichannel business models, digital transformation, and the competitiveness of physical stores in the contemporary e-commerce environment. He has more than 25 years of professional experience in trade, retail, and international business. His career includes founding a company for the import and wholesale of computer equipment and developing a retail network of ten physical stores and four online sales channels in Serbia, Montenegro, and Bosnia and Herzegovina. E-mail: Milos@ekspedicija.rs, ORCID: 0009-0000-9537-4651.

Generative AI in Audit Oversight: A Checklist-Based Model Evaluation

Kresović, Dejana; Drulović, Sofija; Đoković; Nebojša, and Četković, Miloš

Abstract: *This paper examines the use of generative artificial intelligence (GenAI) in audit oversight, with a focus on its role in the preliminary review of the independent auditor's report and accompanying financial statements. Existing audit supervision procedures rely on expert review, which is crucial but also time-consuming when it is necessary to examine large bodies of documents. The proposed solution is a framework supported by GenAI models, based on checklists, for identifying potential formal, structural and substantive problems in reporting that may require further expert review. The empirical analysis is based on 217 sets of documents of public companies in Serbia for the year 2024. ChatGPT and Gemini 3.1 Pro evaluated the same documents using a structured checklist and a risk scale of 0 to 3, while the expert assessment of audit supervision served as a benchmark. Model performance was evaluated using summary risk scores, mean absolute error, bias, Cohen's kappa coefficient, precision, responsiveness, F1-measure, accuracy, and ROC-AUC metrics. The results show that ChatGPT followed a more sensitive review pattern, with higher response and stronger adjusted agreement with the expert benchmark, while Gemini gave more conservative outputs, with higher precision but lower response. Both models performed better in formal and standardized checklist categories than in areas requiring professional judgment, such as key audit issues, business continuity considerations, and internal compliance red flags. The findings suggest that GenAI can support audit oversight by improving the consistency and prioritization of preliminary document review, but only when integrated into a structured framework with expert validation.*

Index Terms: *audit oversight, ChatGPT, expert validation, Gemini, generative AI*

1. INTRODUCTION

The rapid development of generative artificial intelligence (GenAI) has opened up new opportunities for automation and support in professional document analysis [1]–[5]. In accounting and auditing, recent studies have explored the potential of large language models and AI-supported systems in audit documentation, analytical procedures, decision support, and professional work-flows [6]–[10]. In public administration, regulation and supervision, GenAI tools can contribute to faster, more

consistent and more traceable preliminary analytical procedures. However, their use in specialized professional contexts requires caution, especially where model outputs may affect regulatory attention, public trust, or professional judgment [11]–[14].

Audit supervision is one such area. Independent auditor's reports are standardized professional documents, but they also include complex elements that depend on professional judgment [15]–[17], such as modified opinions, key audit matters, disclosures related to business continuity, emphasis of matter paragraphs, and internal consistency between the auditor's report and the accompanying financial statements [17]. These features make audit reports a useful environment for testing GenAI models, as some elements are formal and rule-based, while others require professional interpretation.

This paper does not address whether GenAI can replace auditors or audit oversight bodies. It can't. Rather, it examines whether GenAI models can support the preliminary review of audit reports by identifying potential formal, structural, and substantive issues that may require further expertise. This distinction is important because the value of GenAI in surveillance is not in autonomous decision-making, but in supporting the identification, documentation and prioritization of risks.

This paper evaluates the performance of two GenAI models, ChatGPT and Gemini 3.1 Pro, in the task of reviewing audit reports based on a checklist. Audit reports are used as real-world test documents, while expert assessment of audit oversight serves as a benchmark. The goal is not to assess the quality of individual audit firms, audit engagements or audit opinions. Instead, the study examines how two GenAI models perform in identifying potential issues in audit reporting within the same structured framework.

The contribution of this work is threefold. First, a framework for audit supervision supported by GenAI models, based on a structured checklist review, is proposed. Second, ChatGPT and Gemini are empirically compared on the corpus of audit reports and accompanying financial reports of public companies in Serbia. Third, model performance is evaluated using both concordance measures and binary problem detection metrics, distinguishing between sensitivity, con-

servatism, and concordance with expert judgment.

The rest of the paper is organized as follows. Section 2 presents the theoretical background and proposed framework. Section 3 explains the materials and methods. Section 4 presents and discusses the results. Section 5 concludes the paper and indicates the contribution and limitations.

2. BACKGROUND AND PROPOSED FRAMEWORK

Independent auditor's reports are standardized professional documents used to communicate the auditor's opinion and selected reporting matters relevant to users of financial statements. Their structure and content are primarily determined by international audit reporting standards, including the standard that regulates the formation of opinions and reporting on financial statements [15]. In addition to formal reporting elements, audit reports may also include areas that require a higher level of professional judgment, such as key audit issues and business continuity considerations, which depend on entity-specific interpretation [16], [17]. This combination of standardized structure and judgment-based content makes audit reports suitable real-world documents for evaluating the performance of GenAI models in checklist-based review tasks.

Traditional audit oversight relies on peer review, professional standards, inspection procedures and regulatory assessment. At the same time, recent literature on AI-supported auditing emphasizes that intelligent tools can support but not replace professional judgment [8], [12], [13]. These mechanisms are necessary, but they also require a lot of time and resources. In digital public administration, supervisory bodies increasingly need tools that can support the review of documents and help identify cases that may require more detailed attention from experts. GenAI models can contribute to this process when applied within a structured and controlled framework.

Existing approaches to audit oversight and audit supported by artificial intelligence can be grouped into four broad categories: manual expert inspection supported by standardized checklists, document review based on rules or templates, traditional audit analytics and audit management systems, and new tools supported by artificial intelligence for audit documentation, explanations and decision support. These approaches are useful, but either still rely heavily on manual expert review, or focus mainly on audit execution rather than audit oversight, or provide limited evidence of alignment between GenAI output and an independent expert

benchmark. This study addresses that shortcoming by testing a structured review framework supported by GenAI models on real audit reports and evaluating the model's output against audit oversight expert judgment.

The proposed framework treats GenAI as an assistive tool rather than an autonomous evaluator, which is consistent with previous literature that emphasizes explainability (explainable AI), human supervision, and risk-aware use of AI in auditing [12]–[14]. Its purpose is to support a preliminary review of audit reports using a predefined checklist. The framework consists of five steps: document collection, checklist decomposition, assessment using GenAI models, expert benchmark assessment, and statistical validation.

Figure 1 shows the workflow used in this study. Audit reports and financial statements are first processed through a structured checklist and then independently evaluated by ChatGPT and Gemini. Their risk assessments and findings at the level of individual items are compared with the expert judgment of the audit supervisor and statistically validated. This means that GenAI models can only be used as screening tools when their outputs are compared to expert judgment.

GenAI-Assisted Audit Report Screening Framework

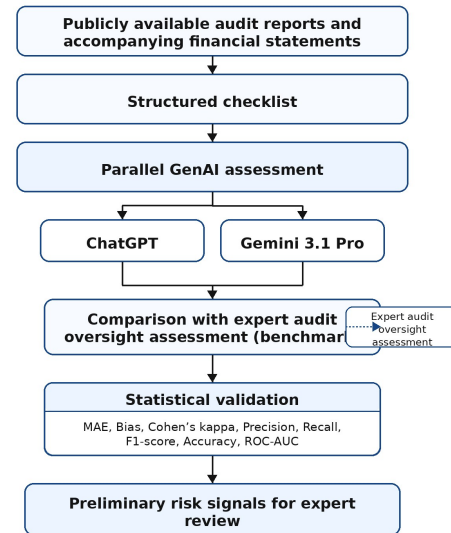


Figure 1. GenAI-Assisted Audit Report Screening Framework.

The framework is designed to be reproducible, transparent and suitable for empirical testing. Both GenAI models receive the same documents, the same checklist and the same rating scale. Their outputs are then compared to expert judgments, allowing the study to examine not only whether the models are detecting potential problems, but also whether their

judgments are aligned with professional supervisory judgment.

In this paper, the focus is on model performance. Audit reports are used as a real-world corpus to test the ChatGPT and Gemini models. The study does not rank audit firms, nor does it make final conclusions about the correctness of audit opinions. Instead, she examines whether GenAI models can support decision-making in the public sector by helping watchdogs identify reports that may require further peer review.

3. MATERIALS AND METHODS

3.1 Data and Audit Report Corpus

The initial research framework included 237 public companies in the Republic of Serbia. The empirical analysis was based on documents that are publicly available through the Agency for Business Registers (APR) [18]. During data collection, complete documentation was not available for 20 companies. These cases were therefore excluded, as they could not provide comparable textual inputs for assessment based on GenAI models.

The final empirical corpus consisted of 217 sets of documents. Each set of documents included one independent auditor's report and accompanying financial statements for the year 2024. The financial statement material included seven standard reporting components for each company, resulting in 1,519 financial statement components, in addition to 217 audit reports. A total of 1,736 documents or document components were used for document review based on GenAI models.

The audit reports were used as real-world test documents to evaluate the performance of the GenAI model. The objective was not to assess the quality of individual audit firms or audit opinions, but to examine how ChatGPT and Gemini identify issues in audit reporting based on the checklist when their outputs are compared to audit oversight expert judgment.

Table 1. Sample and assessment design

Element	Description
Research frame and exclusions	237 public companies; 20 excluded due to incomplete publicly available documentation on APR.
Final analytical sample	217 document sets.
Documents analyzed	Independent auditor's reports and accompanying 2024 financial statements.
Document components	217 audit reports; 1,519 financial statement components; 1,736 total components.
Models and benchmark	ChatGPT (GPT-5.5 Thinking) and Gemini 3.1 Pro, compared against expert audit oversight assessment.
Scoring and purpose	0-3 risk scale; evaluation of GenAI model performance in audit report screening.

3.2 Checklist and Scoring Operationalization

A structured checklist was developed to operationalize the preliminary review of the independent auditor's report and accompanying financial statements. The checklist followed the logic of international audit reporting standards and audit oversight practices, particularly ISA 700, ISA 701 and ISA 570 [15]–[17]. Its purpose was not to determine whether the audit opinion was correct or whether the audit work performed was sufficient, but to identify potential reporting problems that may require further expert review.

The checklist was organized into six categories: formal completeness of the audit report; structural coherence and clarity; consistency of audit opinion; key audit issues; business continuity and highlighting issues; and internal compliance, readability and red flags (risk indicators). Each category included several operational items that could be evaluated based on the audit report and accompanying financial statements.

Table 2. Checklist categories and operational focus

Category	Operational focus and assessment type
A. Formal completeness	Required sections and auditor/date identification; mostly formal.
B. Structure and clarity	Section order, wording clarity, and separation of opinion and basis; limited interpretation.
C. Opinion consistency	Alignment between opinion and basis, including modification wording; mixed formal/judgment-based.
D. Key audit matters	Relevance, entity-specific description, and explanation of auditor response; judgment-based.
E. Going concern/emphasis	Relevant disclosures, consistency with financial statement indicators, and material uncertainty clarity; judgment-based.
F. Internal consistency/red flags	Contradictions, missing explanations, unusual wording, and signals requiring expert review; interpretive.

Each item on the checklist was rated using a four-point ordinal scale of risk. A score of 0 indicated that no problem was identified. A score of 1 referred to a minor problem, such as unclear wording or limited formal inconsistency. A score of 2 indicated a moderate problem requiring further expert review. A score of 3 indicated a significant problem or potential regulatory issue. The same rating scale was applied by ChatGPT, Gemini and an expert rater.

The checklist was deliberately designed to include both standardized and judgment-based elements. Standardized elements are usually clearly visible in the audit report and can be assessed by reviewing the structure and wording of the document. Judgment-based elements require interpretation of the relationship between the auditor's report, financial statement disclosures and potential risk indicators. This design made it possible to examine whether GenAI models show different performance in formal review tasks compared to areas requiring professional judgment.

3.3 Expert Benchmark Assessment

The expert assessment of audit supervision served as a benchmark for evaluating the output of the GenAI model. The expert evaluator was a professional in the field of auditing and financial reporting with practical experience in reviewing independent auditors' reports, assessing compliance with audit reporting requirements and interpreting financial statements in the context of reporting by public interest companies or publicly available company reports.

The expert assessment was conducted independently of the assessment based on the GenAI models. The expert used the same checklist categories and the same rating scale from 0 to 3 as the two GenAI models. The purpose of this assessment was to establish a professional benchmark for preliminary problem detection, not to make final regulatory conclusions or re-evaluate the quality of the audit engagement itself.

To reduce the risk of bias, the expert rater did not use the outputs generated by ChatGPT or Gemini when assigning the reference scores. Therefore, the expert assessment was treated as an independent reference classification. GenAI outputs were compared to this benchmark only after all model evaluations were completed.

When the interpretation of a particular item on the checklist required professional judgment, the expert evaluator took into account both the audit report and the accompanying financial statements. Special attention is paid to modified opinions, key audit matters, disclosures related to business continuity, paragraphs that highlight a particular issue, inconsistencies between the auditor's report and information from the financial statements, as well as other red flags that may require further attention from supervisory bodies.

The quality control procedure was applied to a selected sub-sample of the evaluated document sets. A second auditor independently reviewed part of the sample to check the consistency of the expert assessment. All disagreements were discussed and resolved through professional judgment, with the original reference scores maintained or adjusted where warranted. This procedure was used to increase the reliability of the expert reference value and reduce the risk that the evaluation of the model depends solely on the interpretation of a single evaluator.

3.4 GenAI Assessment Protocol and Reproducibility Measures

An evaluation based on GenAI models was conducted using two general-purpose generative AI systems: ChatGPT, based on OpenAI's GPT-5.5 Thinking model, and Gemini 3.1 Pro, accessed through Google's Gemini environment. For better readability, the models are referred to

as ChatGPT and Gemini in the rest of the paper.

Both models were evaluated under the same evaluation conditions. They were given the same document entries, the same checklist structure, the same rating scale, and the same task instructions. The models were instructed to act as audit oversight assistants and to perform only a preliminary review of the audit report and accompanying financial statements. They are specifically instructed not to determine whether the audit opinion is correct, not to assess the overall quality of the audit engagement, and not to make unsubstantiated regulatory conclusions.

The assessment protocol was designed to enhance comparability and reproducibility. For each set of documents, the relevant audit report and supporting material from the financial statements are made available to the model. The model was then asked to evaluate a set of documents using a predefined checklist and assign a score from 0 to 3 for each item on the checklist. For each non-zero score, the model was also asked to provide a brief rationale and to identify the element of the document that triggered the risk signal.

In order to reduce variability, the same prompt template was used for both models throughout the assessment. The query defined the role of the model, the scope of the task, the checklist categories, the rating scale, and the expected output format. The models were instructed to distinguish confirmed findings in documents from potential problems requiring expert review. They were also instructed to use careful language and to avoid conclusions that would normally require full audit inspection or regulatory authority.

The output of each model was recorded in a structured format, including checklist item scores, summary risk scores, and brief qualitative explanations. This structured output allowed for comparisons between the ChatGPT, Gemini, and expert benchmark models at both an aggregate and individual item level.

Since commercial GenAI systems may change over time, the model's name, access environment, and evaluation period are documented for transparency and repeatability. This is especially important because the results of document analysis based on GenAI models may depend on query design, model version, document formatting, and interface-level settings that the researcher cannot always fully control.

Assessments were conducted during the same evaluation period using the ChatGPT and Gemini document upload web interfaces. No model output was used to modify the query during the main evaluation phase. The exact query template used in the study is listed in Appendix A.

3.5 Model Evaluation and Statistical Measures

The performance of the model was evaluated by comparing the output of the ChatGPT and Gemini models with an expert reference value (benchmark). The evaluation was conducted on two levels. First, the aggregated risk scores were used to compare the overall rating behavior of both models and the expert rater. Second, ratings at the level of individual checklist items were used to assess compliance and problem detection performance.

The mean absolute error (MAE) was calculated to measure the average distance between the model's rating and the expert's rating. Bias was used to determine whether the model generally overestimated or underestimated the risk relative to the expert reference value. A positive bias indicated that the model on average assigned a higher risk score than the expert, while a negative bias indicated a more conservative rating pattern.

Agreement was evaluated using item-level exact agreement and Cohen's kappa coefficient. Exact compliance measured the proportion of items on the checklist for which the model and the expert assigned the same rating. Cohen's kappa coefficient was applied because simple fit may exaggerate model performance when a large number of checklist items are classified as a problem-free situation [19].

For the purposes of binary problem detection analysis, the ordinal scores from the checklist were converted into two categories. A score of 0 was treated as a situation in which no problem was detected, while all scores greater than 0 were treated as a detected problem. Precision, recall, F1-score and accuracy were calculated using expert classification as a reference value. Accuracy measured the proportion of problems identified by the model that were also identified by the expert. The response measured the proportion of problems identified by the experts that the model was able to detect. The F1-measure was used as a balanced measure of precision and responsiveness, while accuracy measured the total proportion of correctly classified items.

The ROC-AUC metric was used to assess the ability of model scores to discriminate between lower and higher risk cases [20]. In the context of the preliminary review in audit supervision, the response (recall) was considered particularly important because the main purpose of the review is to identify cases that may require further expert analysis. Accuracy was also relevant, since a model that generates too many false positives can increase the workload of expert raters.

The evaluation therefore did not proceed from the assumption that assigning higher risk scores

is automatically better. Instead, the analysis distinguished between sensitivity, conservatism, compliance with expert judgment, and practical utility for the first stage of review in audit oversight.

4. RESULTS AND DISCUSSION

4.1 Overall Risk-Scoring Patterns

The first step of the analysis compared the overall risk assessment behavior of the ChatGPT and Gemini models with an expert benchmark. The results show that the two models did not follow the same pattern. ChatGPT generally assigned higher risk scores, indicating a more sensitive screening approach. Gemini assigned lower scores and more often classified checklist items as non-problem situations, indicating a more conservative review pattern.

These findings should be interpreted in light of the study design itself. The models are not evaluated as autonomous decision-makers in the audit, but as structured preliminary review tools used under the same checklist, rating scale, and expert benchmark. Therefore, the results reflect their utility for risk-based document triage rather than their ability to make definitive conclusions in audit oversight.

Table 3. Overall model performance and agreement with the expert benchmark

Panel A. Overall risk-score distribution

Model/benchmark	Mean	Median	Min.	Max.
ChatGPT	0.428	0.45	0	0.88
Gemini	0.129	0	0	0.60
Expert assessment	0.218	0.22	0	0.73

Panel B. Agreement and error measures relative to expert benchmark

Model/benchmark	MAE	Bias	Kappa	Interpretation
ChatGPT	0.204	0.204	0.532	More sensitive screening behavior
Gemini	0.123	-0.115	0.267	More conservative screening behavior
Expert assessment	-	-	-	Benchmark

Panel A. Overall risk-score distribution

Panel B. Agreement and error measures relative to expert benchmark

Table 3 summarizes the overall risk-score distribution and agreement measures. Panel A presents aggregate risk-score patterns, while Panel B reports model error, bias, and chance-adjusted agreement relative to the expert benchmark.

The results indicate that higher risk scores do not automatically imply better model performance. A model that assigns higher scores may be more sensitive, but may also be more likely to generate false positives. Conversely, a model that assigns lower scores may be more accurate in some cases, but may also miss problems identified by the expert. For this reason, evaluation of GenAI models in audit oversight should consider both sensitivity and accuracy.

The perceived pattern of evaluation is important for decision-making in the public sector. In a preliminary review in audit oversight, missing a relevant issue can be more problematic than flagging an item for further review. Therefore, a model with a higher recall may be more useful in the early stage of screening, while a more conservative model may be preferred only when the main priority is to reduce false alarms.

4.2 Agreement with Expert Assessment

The second step examined the level of agreement between each GenAI model and an expert reference value (benchmark). Exact compliance was calculated at the checklist item level, while Cohen's kappa coefficient was used to adjust for compliance that may have occurred by chance [19].

The results show that both models achieved some level of agreement with expert judgment, but the strength and nature of that agreement differed significantly. ChatGPT showed stronger adjusted concordance when Cohen's kappa coefficient was considered. In contrast, the Gemini model's compliance was partly affected by its frequent zero-risk classification. This suggests that Gemini can perform accurately in categories with a large number of items without problems, while being less reliable in detecting less obvious or judgmental problems.

In audit oversight, compliance with expert judgment is critical because the purpose of the GenAI model is not to make independent final decisions, but to support peer review. The results therefore support the use of GenAI models as an auxiliary review mechanism, provided that expert validation remains an integral part of the process.

4.3 Performance by Checklist Category

Model performance differed by checklist category. Both models scored better in standardized and formal categories than in areas requiring professional judgment. This pattern was expected, since formal elements are usually directly visible in an audit report, while judgmental elements require contextual interpretation.

Table 4. Performance by checklist category

Checklist category	ChatGPT agreement	Gemini agreement	Interpretation
A. Formal completeness	89.40%	97.10%	Strong standardized area
B. Structural compliance and clarity	80.00%	85.00%	Mostly standardized, moderate interpretive demand
C. Consistency of audit opinion	89.20%	91.70%	Relatively strong where wording is explicit
D. Key audit matters	66.20%	60.00%	Weakest area; requires professional judgment
E. Going concern and emphasis of matter	81.50%	78.30%	Context-dependent and judgment-based
F. Internal consistency and red flags	67.70%	66.70%	Requires interpretation and expert review

The best performances were observed in the categories related to formal completeness and report structure. These categories include elements such as title, opinion section, basis for opinion, management's responsibility, auditor's responsibility, date, and auditor identification. Since these elements were mostly explicitly stated in the report, both models identified them with relatively high consistency.

The weakest performance was recorded in key audit issues, business continuity considerations, as well as in internal compliance and red flag detection (risk indicators). These areas require more than just text recognition. They require an understanding of whether a matter is specific to a given entity, whether the auditor's explanation is sufficiently informative, whether the opinion is consistent with the basis for the opinion, and whether signals from the financial statements may require additional attention. The results therefore suggest that GenAI models are more reliable in standardized review tasks than in areas requiring professional skepticism and auditor judgment.

4.4 Issue-Detection Performance

In order to assess whether the models are capable of detecting problems in audit reporting, the ratings at the level of individual items were converted into binary classifications. A score of 0 was treated as a situation in which no problem was detected, while all scores greater than 0 were treated as a detected problem. Expert classification was used as a reference value (benchmark).

Table 5. Binary issue-detection performance

Model	Precision	Recall	F1-score	Accuracy	ROC-AUC
ChatGPT	0.551	0.874	0.676	0.830	0.746
Gemini	0.656	0.273	0.385	0.834	0.746

The results show that ChatGPT and Gemini differed significantly in their problem detection behavior. ChatGPT achieved a higher response (recall), which means that it detected a higher proportion of the problems identified by the expert. Gemini showed a more conservative detection pattern and flagged fewer problems. It achieved higher precision, but a significantly

lower response, indicating that it generated fewer false positives, but also missed a higher proportion of problems identified by the expert.

Interestingly, both models achieved the same ROC-AUC value, although their precision and response profiles differed significantly. This indicates that the models had a similar overall ability to distinguish between lower and higher risk cases, but differed in how they converted risk signals into problem detection. In practical terms, ChatGPT followed a more sensitive detection pattern, while Gemini took a more restrictive approach when flagging problems.

From an audit oversight perspective, responsiveness is particularly important in the preliminary review. The review tool should help identify cases that may require additional human analysis. If the model misses a large number of problems identified by the expert, its practical utility for supervision may be limited. Therefore, a model with a higher response may be more suitable for the first stage of screening, even if it generates more false positives. However, this advantage depends on the available expert review capacity, as high-sensitivity review may increase the number of cases requiring human verification.

The results also show that no model should be used without expert validation. GenAI models can support prioritization and documentation, but cannot determine whether the audit work performed was sufficient or whether the audit opinion was correct. Their role is limited to a structured preliminary examination.

4.5 Interpretation of ChatGPT and Gemini Behavior

Empirical results indicate that ChatGPT and Gemini should not be treated as interchangeable tools. ChatGPT acted as a high-sensitivity screening model. It tended to identify more potential problems, assign a higher risk score, and reveal a higher proportion of findings identified by the expert. This makes it more useful in situations where the main goal is to avoid missing relevant issues, provided there is sufficient expert review capacity to evaluate additional flagged cases.

Gemini behaved as a more conservative model. It produced fewer risk signals and assigned lower ratings more often. This may reduce the number of false positives, but may also increase the risk of underdetection (missing a problem). In audit oversight, this distinction is important because the cost of missing a relevant issue can be higher than the cost of sending a case for expert review.

The results suggest that GenAI tools can be useful for audit oversight only when integrated into a structured framework, which is consistent

with previous research emphasizing human oversight, explainability, and ethical safeguards in AI-assisted auditing [12]–[14]. A checklist gives models a defined scope of work, a rating scale enables comparison, and an expert benchmark prevents undue reliance on AI-generated conclusions. This combination is crucial for the responsible use of GenAI models in public sector decision-making.

Overall, the proposed framework provides a practical way to test and document the performance of GenAI models in a specialized professional context. It does not replace human expertise, but it can improve the efficiency and consistency of preliminary document review. The findings also suggest that the choice of model should depend on the purpose of the review task itself: a more sensitive model may be preferred for early risk identification, while a more conservative model may be useful when reducing false positives is a priority and expert review capacity is limited.

5. CONCLUSION

From an operational point of view, the proposed framework offers a scalable and structured approach to the preliminary review of audit reports within a large corpus of documents. Its main advantage is the ability to consistently process standardized documents and identify cases that may require further expert review. However, the practical value of this framework depends on structured inquiries, pre-defined checklist criteria, transparent assessment, statistical validation and expert supervision.

The results indicate that ChatGPT and Gemini should not be treated as interchangeable tools. ChatGPT demonstrated higher sensitivity and significantly stronger recall, which makes it more suitable for first-stage screening tasks where the main goal is to avoid missing potentially relevant issues. However, this advantage depends on the available expert review capacity, since higher sensitivity may also increase the number of cases requiring human verification. Gemini gave more conservative outputs, with higher precision but lower response, which can be useful when reducing false positives is a priority.

Both models scored better in the formal and standardized categories of the checklist than in areas requiring professional judgment. Key audit issues, business continuity considerations and internal compliance red flags remained a more difficult task for both models. This suggests that GenAI can support audit oversight by improving the efficiency and consistency of the preliminary review, but it cannot replace audit oversight's expert judgment or professional judgment.

For decision-making in the public sector, the

proposed framework can improve the documentation, prioritization and consistency of preliminary reviews of audit reports. Its responsible use requires a human-in-the-loop approach, in which GenAI model outputs are treated as risk signals rather than final conclusions. In this sense, the framework is most useful as a decision support mechanism for identifying documents that deserve more detailed attention by experts.

The contribution of this work can be characterized through the creative methods of Combination and Interdisciplinary Transfer. The study combines generative artificial intelligence, audit supervision, expert validation and statistical evaluation of model performance within a single framework. She also applies document analysis based on GenAI models to the specialized context of audit report review and decision support in the public sector.

The study has several limitations. The analysis was carried out on audit reports and financial reports of public companies in Serbia for one reporting year, and the expert reference value was based on a professional assessment within the defined framework of the checklist. Results may therefore depend on the document corpus, checklist design, query structure, model version, and assessment context. Future research may extend this framework to other jurisdictions, additional GenAI models, multilingual audit reports, and other forms of regulatory documentation. Future studies should also examine whether optimized queries, structured inputs, multiple expert benchmarks, or hybrid review procedures (human-AI) can improve model performance.

Appendix A. Prompt Template Used for Genai Assessment

The following query template was used for both GenAI models during the main evaluation phase:

You act as an assistant in auditing supervision. Your task is to perform a preliminary review of the independent auditor's report and accompanying financial statements. You may not determine whether the audit opinion is correct, nor may you evaluate the overall quality of the audit engagement. Your role is limited to identifying potential issues in audit reporting that may require further expert review.

Use the following checklist categories: A. Formal completeness of the audit report; B. Structural coherence and clarity; C. Consistency of audit opinion; D. Key Audit Issues; E. Continuity of business and highlighting issues; F. Internal compliance, readability and red flags (risk indicators).

For each item on the checklist, assign a score as follows: 0 = no problem identified; 1 = minor

problem; 2 = moderate problem that requires further expert review; 3 = significant problem or potential regulatory issue.

For each score greater than 0, please provide a brief explanation based solely on the documents submitted. Make a clear distinction between confirmed findings in the document and potential problems that require expert review. Do not draw unfounded regulatory conclusions.

REFERENCES

- [1] L. Banh and G. Strobel, "Generative artificial intelligence," *Electronic Markets*, vol. 33, no. 1, article 63, 2023, doi: 10.1007/s12525-023-00680-1.
- [2] J. Achiam et al., "GPT-4 Technical Report," arXiv preprint, arXiv:2303.08774, 2023.
- [3] Gemini Team et al., "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," arXiv preprint, arXiv:2403.05530, 2024.
- [4] W. M. Lim et al., "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators," *The International Journal of Management Education*, vol. 21, no. 2, article 100790, 2023, doi: 10.1016/j.ijme.2023.100790.
- [5] K. B. Ooi et al., "The potential of generative artificial intelligence across disciplines: Perspectives and future directions," *Journal of Computer Information Systems*, vol. 65, no. 1, pp. 76–107, 2025, doi: 10.1080/08874417.2024.2302744.
- [6] M. A. Vasarhelyi, K. C. Moffitt, T. Stewart, and D. Sunderland, "Large language models: An emerging technology in accounting," *Journal of Emerging Technologies in Accounting*, vol. 20, no. 2, pp. 1–10, 2023, doi: 10.2308/JETA-2023-014.
- [7] M. Eulerich, A. Sanatizadeh, H. Vakizadeh, and D. A. Wood, "Is it all hype? ChatGPT's performance and disruptive potential in the accounting and auditing industries," *Review of Accounting Studies*, vol. 29, no. 3, pp. 2318–2349, 2024, doi: 10.1007/s11142-024-09833-9.
- [8] H. Gu, M. Schreyer, K. Moffitt, and M. A. Vasarhelyi, "Artificial intelligence co-piloted auditing," *International Journal of Accounting Information Systems*, vol. 54, article 100698, 2024, doi: 10.1016/j.accinf.2024.100698.
- [9] J. Kokina, S. Blanchette, T. H. Davenport, and D. Pachamanova, "Challenges and opportunities for artificial intelligence in auditing: Evidence from the field," *International Journal of Accounting Information Systems*, vol. 56, article 100734, 2025, doi: 10.1016/j.accinf.2025.100734.
- [10] A. R. Otero and M. Agu, "Benefits and drawbacks of incorporating ChatGPT in financial audits," *Current Issues in Auditing*, 2025, doi: 10.2308/CIIA-2024-015.
- [11] A. Eisikovits, W. C. Johnson, and A. Markelevich, "Should accountants be afraid of AI? Risks and opportunities of incorporating artificial intelligence into accounting and auditing," *Accounting Horizons*, 2024, doi: 10.2308/HORIZONS-2024-001.
- [12] I. Munoko, H. L. Brown-Liburd, and M. A. Vasarhelyi, "The ethical implications of using artificial intelligence in auditing," *Journal of Business Ethics*, vol. 167, no. 2, pp. 209–234, 2020, doi: 10.1007/s10551-019-04407-3.
- [13] C. Zhong and S. Goel, "Transparent AI in auditing through explainable AI," *Current Issues in Auditing*, vol. 18, no. 2, pp. A1–A14, 2024, doi: 10.2308/CIIA-2022-057.
- [14] A. Birhane, R. Steed, V. Ojewale, B. Vecchione, and I. D. Raji, "AI auditing: The broken bus on the road to AI accountability," in *Proc. 2024 IEEE Conference on Secure and Trustworthy Machine Learning*, 2024, pp. 612–643.
- [15] International Auditing and Assurance Standards Board, *International Standard on Auditing 700 (Revised)*,

- Forming an Opinion and Reporting on Financial Statements. New York: IFAC, 2020.
- [16] International Auditing and Assurance Standards Board, International Standard on Auditing 701, Communicating Key Audit Matters in the Independent Auditor's Report. New York: IFAC, 2015.
 - [17] International Auditing and Assurance Standards Board, International Standard on Auditing 570 (Revised), Going Concern. New York: IFAC, 2015.
 - [18] Serbian Business Registers Agency, Financial Statements Database. Belgrade: SBRA/APR, 2024.
 - [19] J. Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46, 1960.
 - [20] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.

Dejana Kresović is a doctoral student at the Faculty of Organizational Sciences, University of Belgrade, and is employed at the Securities Commission of the Republic of Serbia in the field of public audit oversight and audit quality control. Her professional work includes the oversight of audit firms and licensed certified auditors, the application of international auditing standards, and the analysis of audit report quality. She is the author of several research papers and has presented at national and international scientific conferences. Her research interests include auditing, public oversight, financial reporting, and the application of generative artificial intelligence. E-mail: dejana.kresovic@sec.gov.rs, ORCID 0009-0008-9798-0733.

Sofija Drulović is a graduate of the Faculty of Organizational Sciences, University of Belgrade, and works as a Junior System Analyst at the Office for Information Technologies and eGovernment of the Republic of Serbia. Her professional focus includes information systems, business analysis, marketing, and strategic problem-solving. During her studies, she was actively involved in the Case Study Club and participated in local, national, and international case study competitions. Her experience includes marketing research, consulting projects, and the preparation of strategic recommendations for business clients. Her professional interests include digital transformation, e-government, business analysis, market research, and the application of artificial intelligence. E-mail: sofija.vukmirovic@ite.gov.rs, ORCID 0009-0003-7367-3230.

Nebojša Đoković holds a MSc in Economics and is an economic affairs adviser at the Association of Montenegrin Banks. He previously worked in banking supervision at the National Bank of Yugoslavia and the Central Bank of Montenegro, and later as an internal auditor at Hipotekarna banka. His professional experience includes banking regulation, financial supervision, internal audit, deposit protection, and European integration in the field of banking and financial services. He is the author of the book Banking of Montenegro and numerous professional articles. His interests include banking, financial services, economic policy, and European financial integration. E-mail: nebojsa.djokovic@mfa.gov.me, ORCID 0009-0002-9335-1530.

Miloš Cetković holds a degree in Management and a master's degree in Business Analytics from the Faculty of Organizational Sciences, University of Belgrade. His academic and research interests focus on retail development strategies, omnichannel business models, digital transformation, and the competitiveness of physical stores in the contemporary e-commerce environment. He has more than 25 years of professional experience in trade, retail, and international business. His career includes founding a company for the import and wholesale of computer equipment and developing a retail network of ten physical stores and four online sales channels in Serbia, Montenegro, and Bosnia and Herzegovina. E-mail: milos.cetkovic@yahoo.com, ORCID 0009-0000-9537-4651.

Education and Cardiovascular Diseases: A Case Study of Patients at the Institute Of Cardiovascular Diseases of Vojvodina (Northern Serbia)

Kričković, Emina; Kričković, Zoran; Marković, Slobodan B.;
Stojšić-Milosavljević, Anastazija; Cvetković, Vladimir M.; and Lukić, Tin

Abstract: *Investigating the relationship between education level and disease incidence is a critical indicator of population health. This study examines the relationship between educational attainment and cardiovascular diseases (CVDs) among patients treated at the Institute for Cardiovascular Diseases of Vojvodina (ICDV) in Sremska Kamenica, northern Serbia. The study period spans from 2003 to 2023. To perform statistical analysis, data processing, and visualisation, several Python modules were utilised. A chi-squared test was first conducted, yielding a values indicating a statistically significant relationship between education level and the number of admitted CVDs patients. Subsequently, a Poisson regression model was applied to aggregated patient admission data to assess the association between educational attainment and CVDs incidence over time. The results confirmed strong statistical significance for both the main effects and their interactions, with all interaction terms showing p-values smaller than 10^{-18} . Additionally, this study conducted an Emerging Hot Spot (EHS) analysis by the education level of patients. Spatial patterns were identified in 152 settlements. All of these settlements exhibited hot spots only, specifically "Consecutive Hot Spot" and "Sporadic Hot Spot" patterns. These findings indicate that temporal trends in hospitalisation rates vary significantly by education level, emphasising the need for targeted preventive measures among lower-educated populations. Studies of this kind can serve as a model for similar research in other regions, helping to inform evidence-based public health strategies and contribute to a better understanding of the social determinants of health on a broader scale.*

Index Terms: *Education, Cardiovascular Diseases, Institute of Cardiovascular Diseases of Vojvodina, Poisson Regression, Emerging Hot spot analysis.*¹

1. INTRODUCTION

HUMAN health is characterised not by a single indicator, but by a complex of factors [1]. It has long been accepted that the health status of a population is determined by economic and social conditions [1,2]. Living standards, lifestyle, customs, economic development, settlement type, population density, literacy, education, occupation, the development of social and healthcare services, and many other factors shape the health status and prevalence of diseases [1,3]. The number of factors related to socioeconomic conditions is so vast that it is not always easy to determine which specific ones are responsible for differences in disease incidence [1,4].

One of the key social determinants of health is the population's level of education. Individuals with higher levels of education are more likely to be employed, to secure jobs with stable income, and, consequently, to enjoy better health [5]. Education also equips people with the necessary skills to identify and solve both individual and collective problems [5]. People with lower levels of education generally exhibit poorer cardiovascular health, have more comorbidities, and face an overall higher risk of developing cardiovascular diseases [6,7]. The level of education influences both morbidity and mortality from numerous diseases [5]. Limited health literacy and overall literacy can prevent individuals from learning about health, using medications correctly, and taking advantage of preventive health services [8].

CVDs represent a major global public health challenge, contributing significantly to the disease burden across both high- and low-income countries [9]. Their development is influenced by a complex interplay of cardiometabolic, environmental, and social risk factors [7].

In Serbia, the burden of CVD remains substantial, with an incidence rate of 920 per 100,000 inhabitants in 2019, placing Serbia 18th

¹ This study was funded by the Program of Cooperation with the Serbian Scientific Diaspora - Joint Research Projects – DIASPORA 2023, from the Science Fund of the Republic of Serbia, under the project LAMINATION (The Loess Plateau Margins: Towards Innovative Sustainable Conservation), Project number: 17807.

among 54 countries in terms of CVD incidence [10].

According to the results of the 2022 Census, more than half of the population aged 15 and over in the Republic of Serbia had completed secondary education (53.1%) [11]. Additionally, 17.8% of the population had completed primary (eight-year) education, while 22.4% held a diploma from a college or university [11]. Conversely, 6.3% of the population either had no formal education or had not completed primary school [11]. Between the 2011 and 2022 censuses, the share of individuals with higher education increased significantly – from 16.24% in 2011 to 22.44% in 2022 [11,12,13]. The proportion of those with secondary education also rose, from 48.93% to 53.08%, while the share of individuals without formal education or with incomplete primary education declined sharply – from 13.68% in 2011 to just 6.28% in 2022 [11,12,13].

This study aims to examine the association between educational attainment and CVDs among patients treated at the Institute for Cardiovascular Diseases of Vojvodina (ICDV), located in Sremska Kamenica (Northern Serbia). The analysis encompasses the entire territory of Serbia, as patients originate from across the country. Additionally, this study conducted an Emerging Hot Spot (EHS) analysis by the education level of patients. The application of Geographic Information Systems (GIS) is crucial for identifying spatial disparities and clusters. By integrating geospatial and socioeconomic data, GIS enables the visualisation of patterns, guides targeted interventions. Investigating the relationship between education level and disease incidence is a critical indicator of population health. Identifying the most affected population groups can inform targeted public health measures, especially for vulnerable categories. Such findings can also contribute to the revaluation of existing health policies and the development of new strategies aimed at reducing the burden of cardiovascular diseases.

1.1. Literature Review

A growing body of research highlights the significant influence of educational attainment on cardiovascular health, emphasising its role in shaping health outcomes across various populations and contexts. Liu et al. investigated whether higher levels of education causally reduce CVD risk among European populations [7]. Similarly, Dégano et al. demonstrated an inverse association between education and CVD incidence in North-Eastern Spain, showing that individuals with university degrees had a lower risk compared to those with only primary or no

education [14]. In Sweden, Erikson (2001) found that men with only primary education had a mortality rate twice as high as those with doctoral degrees [5,15].

Furthermore, the causal associations of education level with cardiovascular diseases (CVD) have been extensively studied worldwide, reflecting the global recognition of education as a critical social determinant of health.

Mallaiah reviews the critical role of education in cardiovascular disease (CVD) prevention, emphasizing how educational factors influence lifestyle choices, health behaviors, and access to healthcare resources. The paper synthesizes existing research and empirical findings to highlight the strong correlation between educational attainment and cardiovascular outcomes, advocating for public health strategies that target education as a key upstream determinant to reduce CVD burden. [16]

Marzà-Florensa et al. examined educational inequalities in secondary prevention of coronary heart disease (CHD) using data from the multinational SURF CHD II audit involving nearly 14,000 patients across 29 countries. The study found that patients with higher education were more likely to meet guideline-recommended risk factor targets, use cardio protective medications, and participate in cardiac rehabilitation. Despite widespread poor control of risk factors overall, educational disparities persisted across countries of varying income levels, highlighting the need to address socioeconomic barriers in CHD management. [17]

Jones et al. used Mendelian randomization analyses with sibling data from large cohorts to investigate the causal impact of educational attainment on CVD risk and serum lipid levels in the UK Biobank. Their findings show that higher education is associated with a reduced risk of CVD and favourable lipid profiles, effects that persist even after accounting for family-related factors. This suggests that increasing education may directly contribute to improved cardiovascular health independently of shared familial influences. [18]

Hu et al. used two-sample Mendelian randomization to examine the causal effects of educational attainment on CVD, biomarkers, and socioeconomic factors. Analysing genetic data, the study found that higher genetically predicted education significantly reduced the risk of multiple CVD outcomes, improved key cardiovascular biomarkers (e.g., blood pressure, BMI, CRP), and was associated with healthier behaviours and higher income. These findings support a causal link between education and cardiovascular health, emphasizing the role of education equity in CVD prevention. [19]

Tao et al. investigated the relationship

between educational attainment, cardiovascular risk factor control, and the incidence of CVD in China using data from the Kailuan Study. Analysing over 92,000 participants over a 14.5-year follow-up, the study found that lower educational levels were significantly associated with higher CVD risk. This association was partly mediated (14.2%) by poorer control of key risk factors such as blood pressure, glucose, cholesterol, smoking, and BMI. The findings highlight the role of education in shaping health outcomes through its impact on risk factor management. [20]

Magnani et al. explored the link between educational attainment and lifetime CVD risk using pooled data from six U.S. cohort studies (1985-2015) involving over 40,000 adults. The study found that lower education levels were associated with significantly higher lifetime CVD risk and shorter periods of life lived free from CVD. While education was linked to healthier longevity, it contributed minimally to improving traditional CVD risk prediction models. The findings underscore education as a key social determinant influencing long-term cardiovascular health outcomes. [21]

Lynch et al. explored the relationship between socioeconomic statuses – measured through education, income, and occupation – and carotid atherosclerosis among men in Eastern Finland [22]. Davies et al. analysed UK Biobank data to examine the causal effects of education on health outcomes [23], while Rose and Marmot investigated the role of social class in coronary heart disease in England and Wales [24]. In the UK, Floud et al. found that lower education combined with higher area deprivation was associated with increased CHD risk [25]. Galama et al. further examined the links between education, health, and mortality [26].

In the United States, Choi et al. analysed associations between educational attainment, chronic disease, and mortality [27], while Havranek et al., in a statement from the American Heart Association, emphasized that lower education is consistently linked to higher prevalence of cardiovascular risk factors [28]. Mackenbach et al. compared educational differences in mortality from ischemic heart disease, cerebrovascular disease, and overall CVD across the United States and 11 Western European countries, consistently reporting higher mortality among individuals with lower education [29].

Meara et al. examined educational disparities in mortality and life expectancy among non-Hispanic Black and White populations in the 1980s and 1990s [30]. Van Lenthe et al. quantified the contributions of material and behavioural factors to educational differences in

acute myocardial infarction incidence [31]. Loucks et al. confirmed the link between lower education and increased CHD risk [32]. Tillmann et al. employed genetic data to explore the causal relationship between education and coronary heart disease [33]. Carter et al. assessed the mediating roles of body mass index (BMI), systolic blood pressure, and smoking behaviour in the relationship between education and CVD outcomes [34]. Gill et al. investigated cardio metabolic factors underlying the clustering of osteoarthritis with CVD, as well as how these mediate the effects of educational attainment [35]. Huang et al. examined spatial patterns of educational differentiation in Guangdong Province and their alignment with objective indicators from a spatial equity perspective [36].

Various studies have applied geographic information systems (GIS) and spatial statistical methods in geographical-medical research to analyse spatial patterns and temporal dynamics of health outcomes [37,38,39,40,41,42].

Janković et al. examined the association between socioeconomic statuses – measured by educational level – and cardiovascular health in the adult population of the Republic of Srpska [43].

In Serbia, geographic-medical investigations into CVDs are still relatively scarce. Maksimović et al. investigated the association between socioeconomic status – measured through education – and risk factors for carotid atherosclerosis in Belgrade during 2006-2007 [44]. Janković et al. assessed the prevalence of ideal cardiovascular health (CVH) across sex and age groups in Serbia and examined how demographic and socioeconomic factors relate to CVH metrics [45]. Šašić et al. explored socio-demographic differences in CVD prevalence and its association with risk factors [46], while Kričković et al. provided a geographic-medical overview of non-communicable diseases, including CVDs in the Autonomous Province of Vojvodina (AP Vojvodina) [47]. Markov analysed changes in the educational structure of the population of AP Vojvodina during the second half of the 20th century, noting a significant rise in the number of individuals with vocational and specialized education and a decline in illiteracy and incomplete primary education [48]. Tomašević et al. examined the growing public health challenges of obesity and hypertension in AP Vojvodina [49], and Petrović and Čanković studied obesity prevalence and socio-demographic influences among adolescents aged 15-19 in the same region [50]. Kričković et al. conducted a spatio-temporal analysis of hypertension among patients treated at the ICDV from 2000 to 2023 [51]. The level of grammar school students' knowledge in Novi Sad about

cardiovascular disease risk factors was investigated by Jaraković et al. [52].

2. MATERIALS AND METHODS

This section presents the dataset used in the study, along with the chi-square test, Poisson regression and Emerging Hot spot (EHS) analyses. It offers a detailed description of the data and the methodological approaches applied.

2.1 Data Set

For this study, data were sourced from the reference database of the ICDV [51,53]. The database is organised according to the 10th Revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), as established by the World Health Organization (WHO) [51,54]. Although the WHO introduced the 11th Revision (ICD-11) at the 72nd World Health Assembly in 2019, which became effective on January 1, 2022 [51,55], the ICD-10 remains the standard classification system used in the Republic of Serbia. In Figure 1 recorded patients through the settlements in the Republic of Serbia in the ICDV are presented.

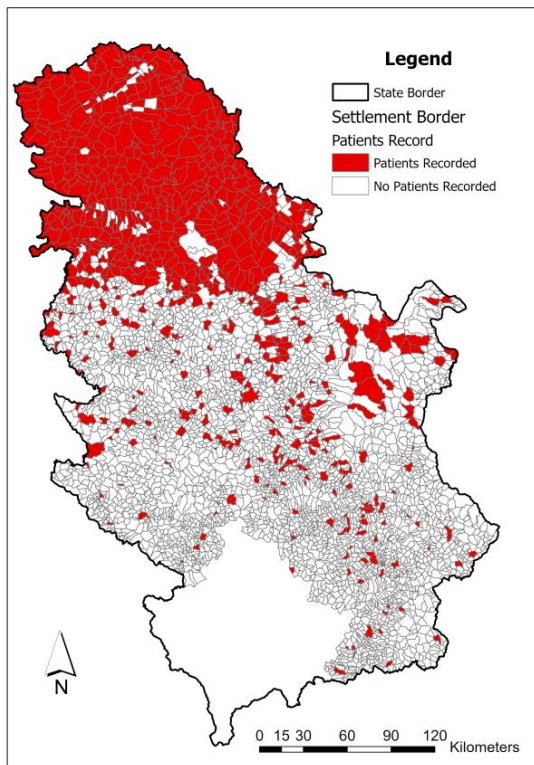


Figure 1: Patients recorded in the ICDV

Although patients from settlements in the Autonomous Province of Kosovo and Metohija (AP KiM) were recorded in the ICDV database, they were excluded from this study due to the unavailability of settlement-level data from the open-access sources of the Republic Geodetic Authority (RGA) of the Republic of Serbia. As illustrated in Figure 1, settlement boundaries for the AP KiM region are not publicly available.

Research process diagram is presented in the Figure 2. Each record in the dataset represented an individual hospital admission, with the patient's education level, and the year of admission extracted from the admission date. During this process, it was observed that the column containing information on patients' education levels included values that did not align with the database description. Out of a total of 1,387,598 recorded cases, 363,093 entries either contained invalid values or were missing education level data. Consequently, 1,024,505 cases with valid information on education level were retained for further analysis.

The education level variable comprised five categories:

- Value 1 represented patients with no formal education;
- Value 2 corresponded to those with primary education;
- Value 3 denoted patients with secondary (high school) education;
- Value 4 indicated patients with post-secondary (higher) education, and
- Value 5 included those with university-level education or higher

The ICDV's database was in *.CSV format, and the observed period was from 2003 to 2023. Following data cleansing, Figure 3 presents the total number of patients stratified by education level.

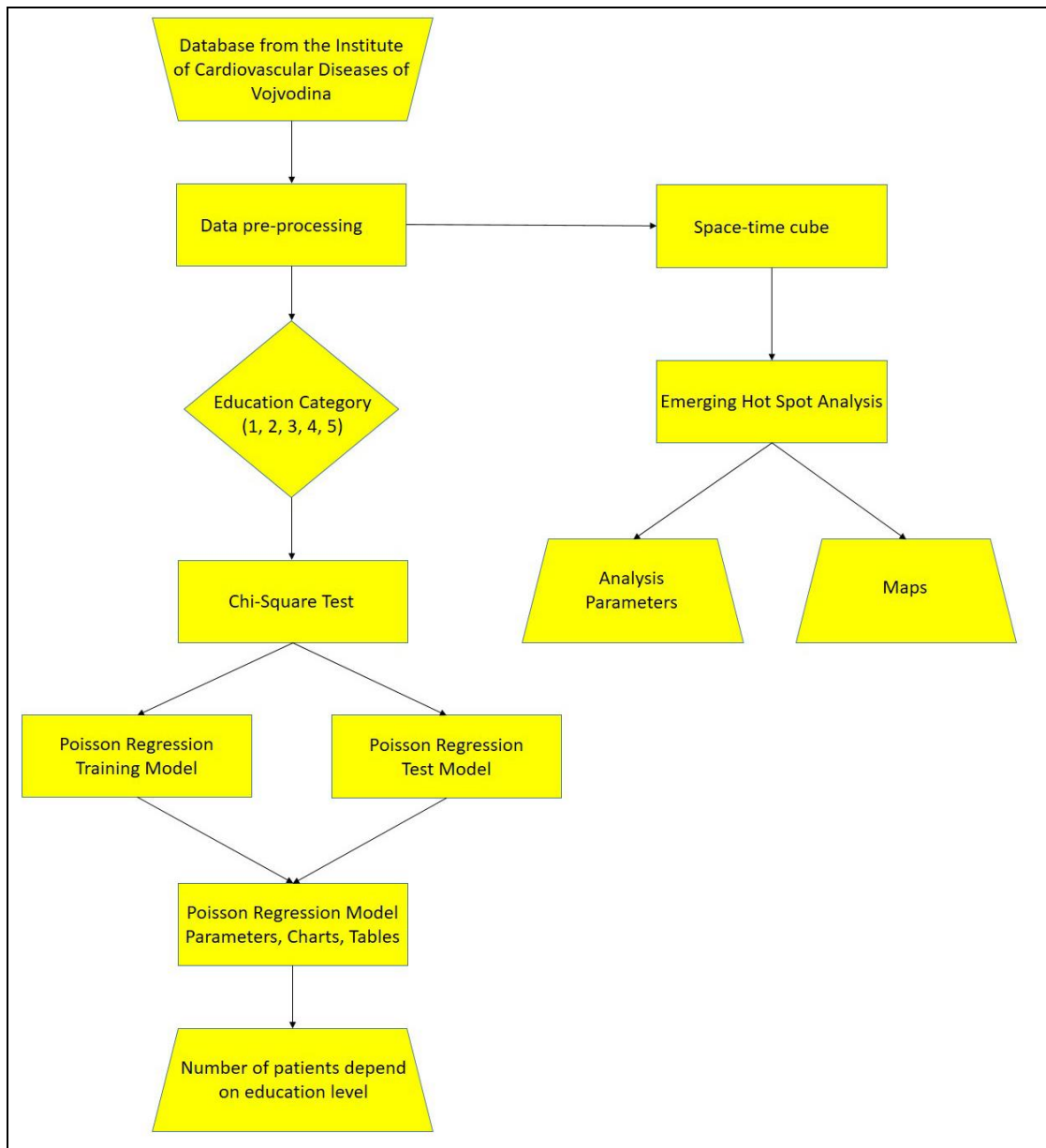


Figure 2: Schematic representation of the research process.

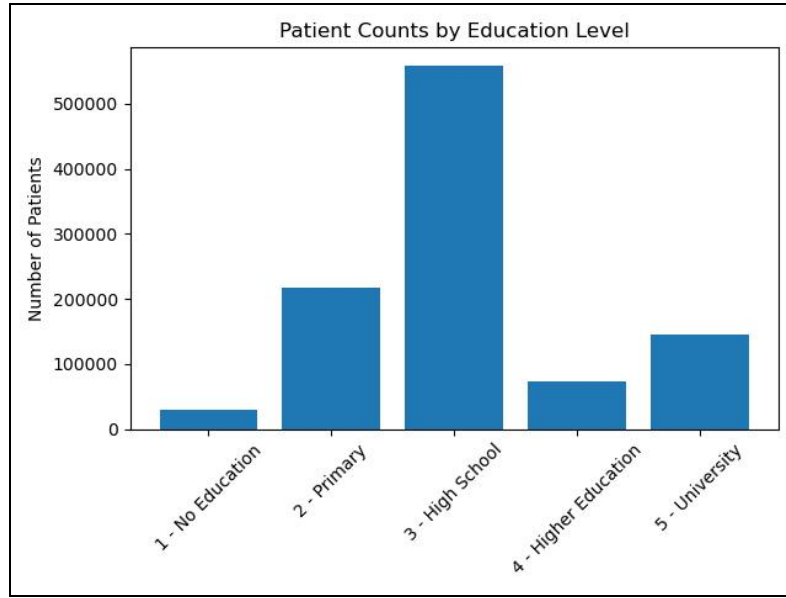


Figure 3: Total number of patients in the ICDV by education level.

For this analysis, several Python modules were utilized to perform statistical operations, data processing, and visualization. This included libraries such as “scipy”, “pandas”, “statsmodels”, “numpy”, and “matplotlib” [56,57,58,59,60].

2.2 Chi-Square Test

The chi-square test is a non-parametric statistical method used to assess the difference or association between observed and expected categorical variables in a dataset [61,62]. Developed by Karl Pearson (1857-1936), it has become a key tool in hypothesis testing, particularly when dealing with categorical data [62,63]. One of the main advantages of Pearson's chi-square test is its ability to use statistical methods that do not assume a normal distribution [62]. The significance of the chi-square value is determined by consulting a chi-square table, taking into account the appropriate degrees of freedom and the chosen level of significance [62,63]. The chi-square test serves two primary purposes: determining whether there is an association between two or more groups, populations, or criteria, and evaluating how well the observed data distribution fits the expected distribution [62,64].

Both of Pearson's chi-square tests use the same equation to calculate the test statistic, chi-square (χ^2) [61] expressed in equation (1):

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad (1)$$

where χ^2 is the chi-square test statistic, Σ is the summation operator, O is the observed frequency, and E is the expected frequency.

The larger the difference between the observations and the expectations ($O - E$ in the equation), the bigger the chi-square will be. To decide whether the difference is big enough to be statistically significant, you compare the chi-square value to a critical value [61].

If the p-value is less than the significance level (typically 0.05), reject the null hypothesis, indicating a significant relationship between the variables. If the p-value is greater than 0.05, fail to reject the null hypothesis, meaning no significant relationship was found. [65]

2.3 Poisson Regression

In this study, a Poisson regression model was employed to examine the relationship between the level of education and the incidence of CVDs. Given that the dependent variable (the number of recorded CVD cases) represents count data and is assumed to follow a Poisson distribution, this statistical method is well-suited for the analysis.

Poisson regression is the simplest count regression model. Coefficients are exponentiated, since counts must be 0 or greater. Poisson regression assumes a Poisson distribution, often characterised by a substantial positive skew (with most cases falling at the low end of the dependent variable's distribution) and a variance that equals the mean. Because count data distributions (e.g., visit counts) often have a Poisson distribution, Poisson regression tends to fit these data better than linear regression does (which assumes a normal distribution). [66]

This allows the analysis of dependent variables, which follow Poisson distribution, i.e.,

a distribution frequently encountered when counting a number of events, or concerning the rate of occurrence with nonnegative integers [67,68]. The Poisson regression model, applied in both analyses, assumes that the logarithm of the expected count is a linear function of the predictor variables. The general form of the model is shown in equation (2):

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (2)$$

where μ is the expected count (the number of hospital admissions per year), $X_1, X_2, \dots, X_n$ are the predictor variables (education level), β_0 is the intercept, $\beta_1, \beta_2, \dots, \beta_n$ are the regression coefficients for each predictor. Since a Poisson variable is assumed to have a mean > 0 , $\log$ transformation ensures that the model-based predictions of rates are constrained to be greater than, or equal to, zero. A P-value of less than 0.05 was considered statistically significant in this study [68].

2.4 EHS Analysis

The Getis-Ord G_i^* statistic, implemented through the EHS Analysis (Space-Time Pattern Mining) tool, was utilized to identify areas classified as "hotspots" and "coldspots" [47,51,69,70,71] concerning education level. This method categorizes spatial patterns into clusters and outliers, with clusters representing either positive patterns (hotspots) or negative patterns (cold spots) [47,51,69,72]. Outliers, on the other hand, are spatial features with attribute values that significantly differ from those of their surrounding neighbours [47,39,69]. This hotspot analysis is based on the G_i^* statistic, which is calculated using the following equation (3) [47,51,39,69,72]:

$$G_i^* = \frac{\sum_{j=1}^n \omega_{i,j} x_j - \bar{X} \sum_{j=1}^n \omega_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n \omega_{i,j}^2 - (\sum_{j=1}^n \omega_{i,j})^2}{n-1}}} \quad (3)$$

where G_i^* is the spatial autocorrelation (spatial dependency) statistics of an event i over the n events, the term x_j defines the magnitude of the variable x at the events j over all n , and the term $\omega_{i,j}$ defines the weight value between the events i and j that represents their spatial interrelationship and in more detail, $\bar{X}$ shown in equation (4):

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (4)$$

and S in equation (5):

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (5)$$

The G_i^* statistic evaluates the magnitude of each feature in relation to the values of its neighbouring features. The local sum of a feature and its neighbours is compared to the sum of all features. If the local sum differs significantly from the expected sum, with the difference being too large to attribute to random variation, a statistically significant z-score is obtained [47,51,39,69,70,71,72].

3. RESULTS

The results of this study are presented across separate subsections, providing detailed findings from the chi-square test, Poisson regression, and EHS analyses.

3.1. Results of Chi-square test and Poisson regression

As previously mentioned, a chi-squared test was conducted to assess the statistical significance of the differences in patient numbers across education levels. The test yielded a chi-squared value of 729,405.92 with a p-value less than 0.005, indicating a statistically significant association between education level and the number of admitted patients.

To assess whether the number of hospitalized patients over time is associated with their level of education, a Poisson regression model was employed using aggregated patient admission data. Following data cleaning and grouping by education level and year, a train/test split based on time was conducted. The training set covered the period from 2003 to 2018 (80 grouped observations), while the test set spanned 2019 to 2023 (25 grouped observations), ensuring that all education levels were present in both subsets. This allowed both for baseline differences between education levels and for interaction terms to model education-specific temporal trends. Education level 5 (university degree) was used as the reference category, enabling direct comparisons with all other education levels.

The full model results are shown in Table 1, which includes estimated coefficients, standard errors, z-values, 95% confidence intervals, and associated p-values. All main effects and interactions were highly statistically significant ($p < 0.01$), indicating that education level significantly influences both the number of hospitalizations and their temporal trends.

To aid interpretation, the coefficients and their 95% confidence intervals are visualized in Figure 4. All coefficients are significantly different from zero (as shown by intervals that do not cross the vertical line at zero), especially the interaction terms involving education levels 2 (primary school) and 3 (high school), which exhibit steeper

increases over time compared to the reference group.

Table 1. Poisson Regression Results for ICDV Admissions by Education Level (2003-2018), with coefficient values, standard errors, z-values, p-values and confidence intervals.

Term	Coefficient	Std. Error	z-value	p-value	95% CI (Lower)	95% CI (Upper)
Intercept	-8.7451	3.292	-2.657	7.89×10^{-3}	-15.197	-2.293
C(education)[T.2.0]	-39.3916	3.542	-11.122	9.80×10^{-29}	-46.333	-32.450
C(education)[T.3.0]	-56.4781	3.406	-16.581	9.51×10^{-62}	-63.154	-49.802
C(education)[T.4.0]	-35.0770	4.043	-8.676	4.11×10^{-18}	-43.001	-27.153
C(education)[T.5.0]	-54.2657	3.730	-14.548	5.97×10^{-48}	-61.576	-46.955
year	0.0078	0.002	4.784	1.72×10^{-6}	0.005	0.011
C(education)[T.2.0]:year	0.0205	0.002	11.646	2.42×10^{-31}	0.017	0.024
C(education)[T.3.0]:year	0.0294	0.002	17.362	1.60×10^{-67}	0.026	0.033
C(education)[T.4.0]:year	0.0178	0.002	8.845	9.18×10^{-19}	0.014	0.022
C(education)[T.5.0]:year	0.0276	0.002	14.889	3.88×10^{-50}	0.024	0.031

Predictions were generated on the test set (2019-2023) to assess model generalization. Figure 5 presents a comparison of predicted vs. observed patient counts across years and education levels. The model generally aligns with the observed trends, though it slightly under

predicts counts for lower education levels – particularly for patients with primary and high school education – suggesting potential structural changes or external factors not captured in the model.

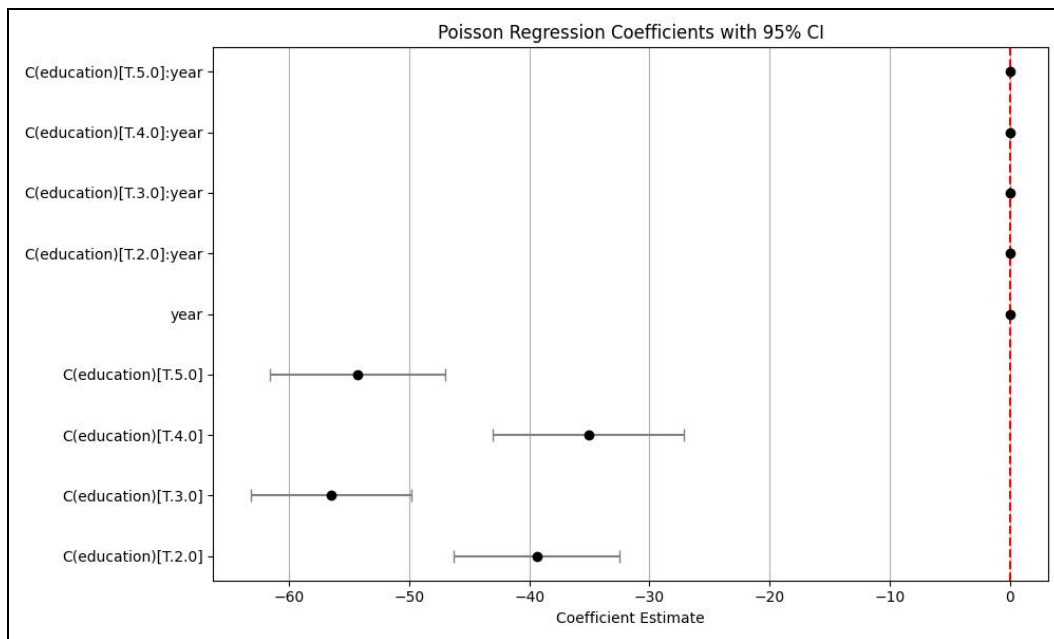


Figure 4: Poisson Regression Coefficients with 95% Confidence Intervals.

Despite this under prediction, residuals remained within statistically interpretable bounds and did not exhibit signs of systematic bias (see Table 2 for a selection of test set residuals). The

university-educated group (education level 5) had the most accurate predictions, further reinforcing the stratified trend differences identified in the regression analysis.

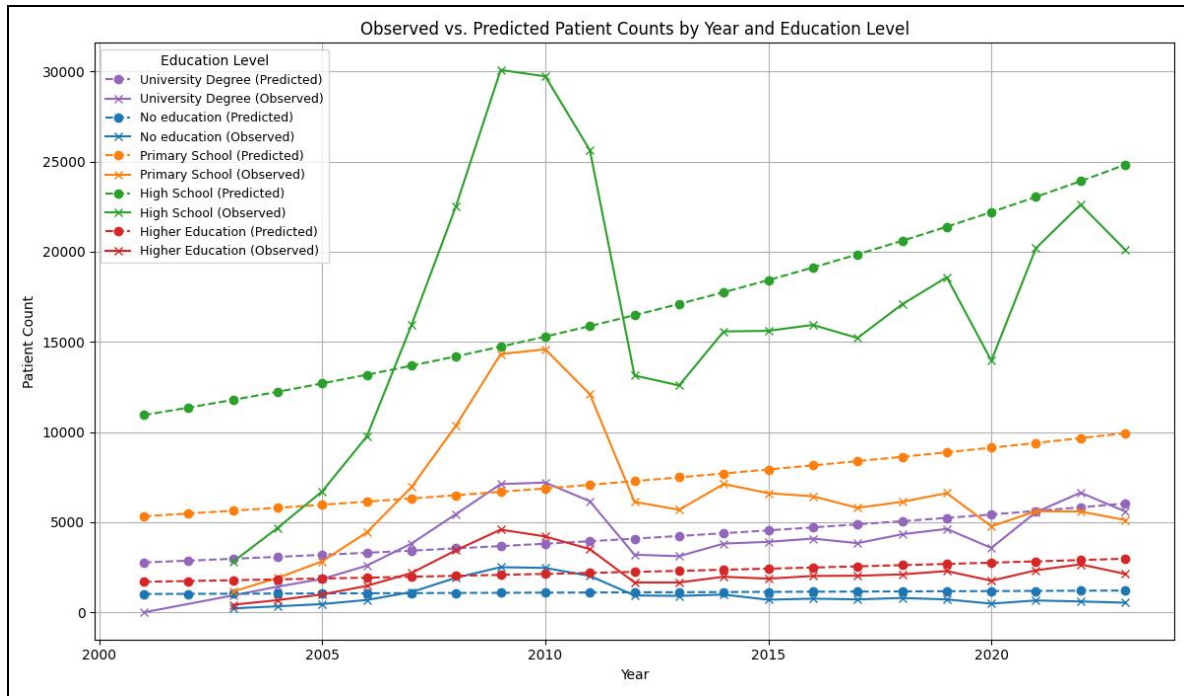


Figure 5: Observed vs. Predicted Patient Counts by Year and Education Level.

Table 2. Example Residuals from Test Set (2019-2023).

Education Level	Year	Observed Count	Predicted Count	Residual
No Education	2019	718	1,173.02	-455.02
No Education	2020	486	1,182.24	-696.24
No Education	2021	659	1,191.54	-532.54
No Education	2022	605	1,200.90	-595.90
No Education	2023	537	1,210.35	-673.35
Primary School	2019	6,614	8,877.57	-2,263.57
Primary School	2020	4,775	9,132.80	-4,357.80
Primary School	2021	5,612	9,395.37	-3,783.37
Primary School	2022	5,595	9,665.49	-4,070.49
Primary School	2023	5,121	9,943.38	-4,822.38
High School	2019	18,589	21,390.65	-2,801.65
High School	2020	13,961	22,202.33	-8,241.33
High School	2021	20,203	23,044.80	-2,841.80
High School	2022	22,617	23,919.24	-1,302.24
High School	2023	20,116	24,826.87	-4,710.87
Higher Education	2019	2,291	2,687.33	-396.33
Higher Education	2020	1,756	2,757.06	-1,001.06
Higher Education	2021	2,335	2,828.59	-493.59
Higher Education	2022	2,664	2,901.99	-237.99
Higher Education	2023	2,132	2,977.28	-845.28
University Degree	2019	4,629	5,240.99	-611.99
University Degree	2020	3,575	5,430.12	-1,855.12
University Degree	2021	5,559	5,626.07	-67.07
University Degree	2022	6,638	5,829.10	+808.90
University Degree	2023	5,599	6,039.45	-440.45

P-values for all regression coefficients are provided in Table 1 and reaffirm the strong

statistical support for both main and interaction effects. All interaction terms had p-values

smaller than 10^{-18} , confirming that temporal trends in hospitalization counts differ significantly by education level.

3.2. Results of the EHS Analysis by Education Level

The EHS analysis was conducted by converting settlement polygons into points. This step was necessary because the analysis required the creation of a Space Time Cube through point aggregation. In this way, all admitted patients were geolocated by merging data from the CSV database with spatial data from the RGA. Using the admission date from the database, a temporal parameter was defined. This procedure enabled the construction of the Space Time Cube and the identification of potential cold or hot spots through EHS analysis.

The analysis was performed separately for each education level. For all cubes created, a uniform configuration was applied: the time field was set, the time-step interval was fixed at one year, the time-step alignment was set to the end time, and a fishnet grid was used as the aggregation shape. The coordinate system applied was WGS 1984 UTM Zone 34N.

According to the supplementary data file ([IPSI Journal Education and Cardiovascular Diseases a Case Study - Supplementary data](#)), spatial patterns were identified in 152 settlements. All of these settlements exhibited hot spots only, specifically “Consecutive Hot Spot” and “Sporadic Hot Spot” patterns. This is attributed to an increasing trend detected in the Space Time Cube, generated from aggregated settlement points, with a high level of statistical confidence. In 45 settlements, hot spots were observed across all education levels. This indicates that in approximately one-third of the settlements, EHS analysis consistently identified hot spots regardless of education level. In 27 settlements, at least one pattern per education level was recorded. The most frequent pattern detected was the “Consecutive Hot Spot.” Figure 6 presents the EHS analysis across education levels: no education, primary education, high school education, higher education, and university/higher education.

When examining the number of settlements identified for each education level, the lowest count was recorded for the no-education category (50 settlements), while the highest was found in the primary education category (127 settlements). High school education was associated with 126 settlements, university and higher education with 115 settlements, and higher education alone with 97 settlements. Consistent with the previous analyses, the lowest number of detected settlements in the EHS analysis corresponded to the lowest education

level.

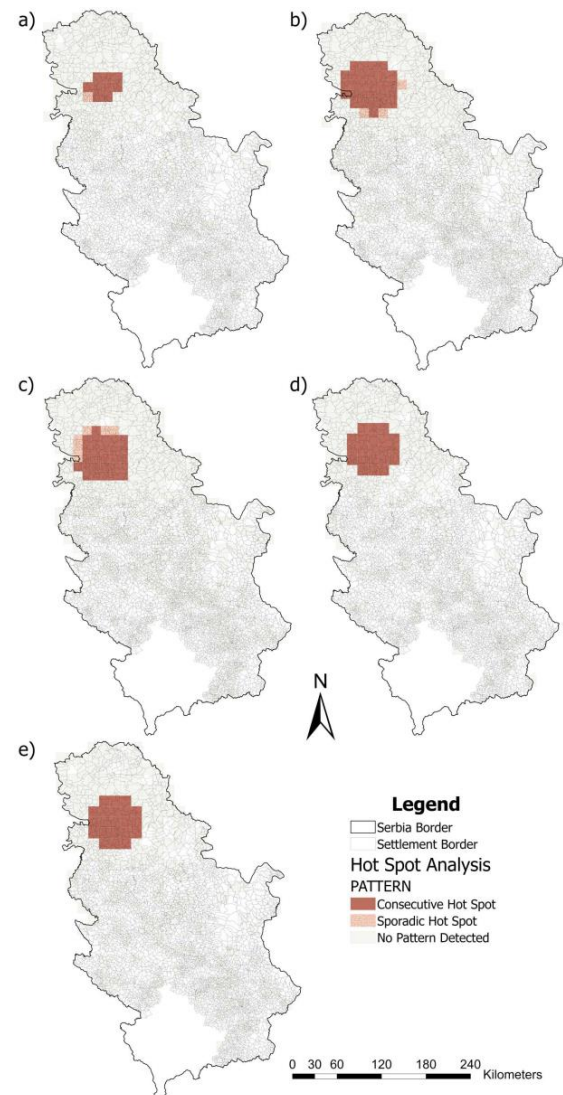


Figure 6: EHS analysis by education level: a) No education, b) Primary education, c) High School education, d) Higher Education, e) University and higher education.

4. DISCUSSION

Education plays a crucial role in shaping health outcomes, particularly in the context of CVDs [7]. This study confirmed a significant association between education level and the number of hospitalised patients. The model also showed that hospitalisation rates and their temporal trends vary across education groups. Most patients at the ICDV in Sremska Kamenica had completed secondary education, followed by those with primary education. A study conducted by Markov revealed that during the second half of the 20th century in AP Vojvodina, there was a notable rise in literacy levels and an improvement in the educational structure by level of education [48]. This was largely

attributed to the expansion of the educational network, which significantly increased the number of individuals with vocational and professional qualifications, while the number of illiterate persons and those with incomplete primary education declined at a slower pace, highlighting the importance of understanding this relationship for guiding targeted prevention strategies, especially for vulnerable populations with lower levels of education [48].

Our findings are consistent with international research highlighting the protective effects of higher educational attainment. A large-scale Swedish study [73] found that completing compulsory education was significantly associated with reduced mortality. Liu et al. further demonstrated, through genetic and epidemiological analysis, that individuals with higher education had a lower incidence of ten distinct cardiovascular conditions [7]. Sulo et al. and Tillmann et al. also identified education as a key determinant of disease progression and prognosis [33,74].

In Serbia and the broader region, similar results have been observed. Janković et al. (2014), Janković et al. (2019) and Maksimović et al. reported that individuals with higher education had more ideal CVH metrics and fewer metabolic risk factors [43,44,45,]. Šašić et al. using data from the 2019 National Health Survey, confirmed strong links between CVDs prevalence and preventable risk factors – such as obesity, hypertension, alcohol use, and smoking – closely tied to socio-economic inequalities [46]. These findings underscore the role of education not only as a social determinant but also as a modifiable factor influencing behavioural and biological risk pathways [34,35].

This study is based on data from a single healthcare institution, the ICDV, which treats patients from across the Republic of Serbia. However, most patients come from AP Vojvodina, which reflects the institution's location. Figure 7 illustrates the distribution of ICDV patients by their dominant education level. As shown in the figure, the majority of patients have attained a high school level of education; however, it is also evident that university and higher education levels are dominant in some urban areas, e.g. university centres.

While the data provide important insights, a more comprehensive and nationally representative picture requires data from multiple institutions across different regions. Such efforts demand financial and logistical resources and strong inter-institutional collaboration.

Moreover, this study focused on education as a predictor of CVDs but did not explore lifestyle behaviours. Future research should incorporate behavioural, environmental, and genetic data to

better understand the mechanisms through which education influences cardiovascular health.

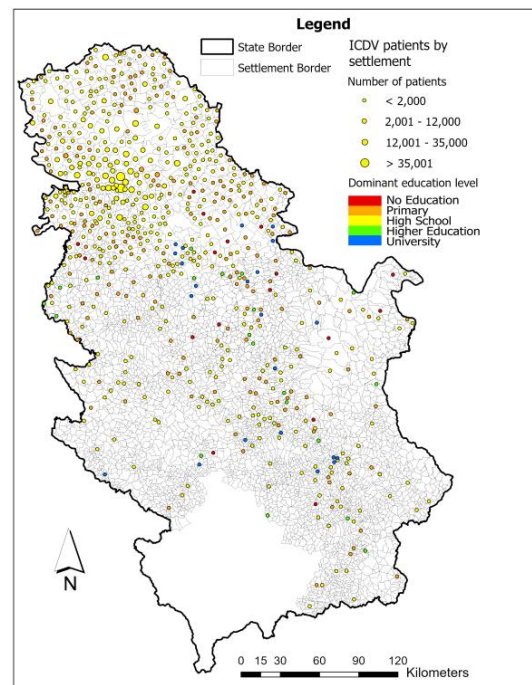


Figure 7: ICDV patients by settlement.

Improving access to education, especially for socially and economically disadvantaged populations, remains a vital public health objective. Expanding educational infrastructure, reducing tuition fees, increasing scholarships, and supporting adult learning are all measures that can improve health outcomes indirectly by promoting healthier lifestyles and improving health literacy [1]. In addition, continuous professional development and adequate support for teachers are essential to maintaining high-quality education across regions. Public awareness campaigns highlighting the benefits of education for personal growth, employability, and health can further reinforce these efforts, particularly when local communities and parents are actively engaged. Integrating technology through digital tools, internet access, and interactive learning platforms will also help bridge educational gaps and modernise the system. It is also important to monitor how changes in the education system over time – such as curriculum reforms or digitalisation – affect population health in the long term.

Strengthening the collaboration between the health and education sectors – particularly through school-based health education programs – can foster healthier behaviours from an early age and reduce the lifetime risk of CVDs. Embedding cardiovascular health literacy into formal curricula may be a long-term strategy for breaking intergenerational cycles of

disadvantage.

However, education alone cannot address all health disparities. CVH is shaped by a complex interplay of social, economic, and environmental factors, as well as access to healthcare. National health policies must align with the Sustainable Development Goals and account for local needs and constraints [46]. As Lakić et al. showed, CVDs poses a significant economic burden in Serbia, underscoring the urgency of strategic investment in prevention [75].

Analysing the relationship between educational attainment and disease prevalence provides a critical lens for understanding population health dynamics and social determinants that drive health disparities. Education not only influences individual health behaviours and access to healthcare but also reflects broader socioeconomic conditions that shape disease risk across communities. By uncovering how education levels correlate with cardiovascular disease incidence, researchers and policymakers can identify high-risk groups more accurately and target interventions where they are most needed. These insights enable the design of more equitable, interdisciplinary, and evidence-based public health strategies that address both the medical and social drivers of disease. Ultimately, such approaches are essential to reducing the overall burden of cardiovascular diseases and advancing health equity throughout Serbia, ensuring that vulnerable populations receive tailored support to achieve better health outcomes.

5. CONCLUSION

This study is among the first geo-medical investigations in AP Vojvodina and the Republic of Serbia to explore the complex link between educational attainment and CVD incidence, revealing a significant association between lower education levels and higher CVD risk. By using patient data from the ICDV in Sremska Kamenica and incorporating GIS and geospatial analysis, the study effectively captures socio-spatial disparities that influence disease patterns, demonstrating the suitability of these tools in medical geography to identify vulnerable populations and spatial clusters of disease. GIS enables the integration of epidemiological data with environmental and socioeconomic variables, offering a powerful framework to visualize and analyse how education and other social determinants interact with geographic factors to affect cardiovascular health. While the study's strengths lie in its innovative combination of spatial and clinical data, its cross-sectional design and focus on a single centre limit causal inference and generalizability. Additionally,

educational attainment, though a valuable proxy, does not fully encompass the complexity of socioeconomic influences. Future research should expand to longitudinal, multi-regional studies incorporating broader social and environmental determinants and apply advanced spatial-temporal modelling to better understand dynamic CVD patterns. These insights reinforce the importance of targeted public health interventions focused on education, health literacy, and equitable healthcare access, highlighting how GIS-based geospatial analysis can guide precision public health strategies to reduce health inequalities and improve cardiovascular outcomes in Serbia and beyond.

ACKNOWLEDGEMENT

This study was supported by the Program of Cooperation with the Serbian Scientific Diaspora-Joint Research Projects – DIASPORA 2023, from the Science Fund of the Republic of Serbia, under the project LAMINATION (The Loess Plateau Margins: Towards Innovative Sustainable Conservation), Project number: 17807 and the support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Grants No 451-03-33/2026-03/200125 & 451-03-34/2026-03/200125 and 451-03-33/2026-03/200091). Furthermore, this study was supported by the Scientific-Professional Society for Disaster Risk Management and ProSafeNet – The Global Hub for Safety, Security, Risk & Emergency Professionals and Scientists (<https://prosafenet.com/>). Also, this study was supported by the project of the Ministry of Defence of the Republic of Serbia and the Serbian Armed Forces – Application of Digital Geotopographic Data in Multi-Criteria Analysis for Identifying Safe Helicopter Landing Zones in the Republic of Serbia (Project number: 1.18-2026).

REFERENCES

- [1] Kričković, E., "The influence of geographical and medical factors on the health of the population of AP Vojvodina," Doctoral dissertation, University of Belgrade, Faculty of Geography, Belgrade, Serbia, 2022.
- [2] Sokić, S., General epidemiology, in S. Krajinović & Z. Radovanović (Eds.), Faculty of Medicine, University of Belgrade, 1985.
- [3] Obradović-Arsić, D., Gledović, Z., Medical geography, Faculty of Geography, University of Belgrade, 2012.
- [4] Vlajinac, H., Descriptive epidemiology, in S. Krajinović & Z. Radovanović (Eds.), General Epidemiology, University of Belgrade, Faculty of Medicine, 1985.
- [5] Nadj, I., Dujmović, F., Plavša, J., Lukić T., Medical geography, Faculty of Sciences, Department of Geography, Tourism and Hotel Management, Novi Sad, 2016. ISBN 978-86-7031-426-9
- [6] Di Chiara, T., Scaglione, A., Corrao, S., Argano, C., Pinto, A., Scaglione, R., "Association between low education and

- higher global cardiovascular risk," Wiley, J. Clin. Hypertens., vol. 17, no. 5, pp. 332–337, 2015. <https://doi.org/10.1111/jch.12506>
- [7] Liu, W., Lin, Q., Fan, Z., Cui, J., Wu, Y., "Education and cardiovascular diseases: a Mendelian randomization study," *Frontiers in Cardiovascular Medicine*, vol. 11, p. 1320205, 2024. <https://doi.org/10.3389/fcvm.2024.1320205>
 - [8] Tulane University School of Public Health, "Education as a social determinant of health," 2021. Accessed: 28 July 2025. [Online]. Available: <https://publichealth.tulane.edu/blog/social-determinant-of-health-education-is-crucial/>
 - [9] Rudasingwa, G., Cho, S., Ishimwe, N. C., "Regional prevalence and spatial modeling of hypertension in South Africa," *Open Health*, vol. 5, p. 20230030, 2024. <https://doi.org/10.1515/ohe-2023-0030>
 - [10] Timmis, A., Vardas, P., Townsend, N., Torbica, A., Katus, H., De Smedt, D., Gale, C. P., Maggioni, A. P., Petersen, S. E., Huculeci, R., "European Society of Cardiology: Cardiovascular disease statistics 2021," *Eur. Heart J.*, vol. 43, pp. 716–799, 2022. <https://doi.org/10.1093/eurheartj/ehab892>
 - [11] Statistical Office of the Republic of Serbia, "2022 Census of population, households and dwellings: Ethnicity. Data by municipalities and cities," 2023. [Online]. Available: <https://publikacije.stat.gov.rs/G2023/Pdf/G20234001.pdf>
 - [12] Statistical Office of the Republic of Serbia, "Census 2011 – Educational attainment and literacy," 2023. [Online]. Available: <https://popis2022.stat.gov.rs/sr-Cyrl/5-vestisaopstenja/news-events/20230731-skolska-sprema-pismenost>
 - [13] Statistical Office of the Republic of Serbia, "Census 2022 – Educational attainment and literacy," 2023. [Online]. Available: <https://popis2022.stat.gov.rs/sr-Cyrl/5-vestisaopstenja/news-events/20230731-skolska-sprema-pismenost>
 - [14] Dégano, I. R., Marrugat, J., Grau, M., Salvador-González, B., Ramos, R., Zamora, A., Martí, R., Elosua, R., "The association between education and cardiovascular disease incidence is mediated by hypertension, diabetes, and body mass index," *Nature Publishing Group, Sci. Rep.*, vol. 7, no. 1, p. 12370, 2017. <https://doi.org/10.1038/s41598-017-10775-3>
 - [15] Erikson, R., "Why do graduates live longer? Education, occupation, family and mortality during the 1990s," in *Cradle to Grave: Life-Course Change in Modern Sweden*, J. O. Jonsson and C. Mills, Eds., Sociology Press, Durham, UK, 2002, pp. 125–147. ISBN 9781315074610
 - [16] Mallaiiah, P., "How quality of education impacts cardiovascular disease," *Rising Researchers*, 2024. Accessed: 09 October 2025. [Online]. Available: <https://risingresearchers.com/journal/how-quality-of-education-impacts-cardiovascular-disease/>
 - [17] Marzà-Florensa, A., Vaartjes, I., Graham, I., Klipstein-Grobusch, K., Grobbee, D. E., "A Global perspective on cardiovascular risk factors by educational level in CHD patients: SURF CHD II," *Glob. Heart*, vol. 19, p. 60, 2024. <https://doi.org/10.5334/gh.1340>
 - [18] Jones, P., Bhatta, L., Howe, L., Vinuela Veloz, M. F., Smith, G. D., Naess, Ø. E., Brumpton, B. M., "Education and cardiovascular disease: a within-family Mendelian randomisation analysis," *Cold Spring Harbor Laboratory, medRxiv*, 2025. [Online]. Available: <https://doi.org/10.1101/2025.07.02.25330770>
 - [19] Hu, M., Yang, T., Yang, Y., "Causal associations of education level with cardiovascular diseases, cardiovascular biomarkers, and socioeconomic factors," *Elsevier, Am. J. Cardiol.*, vol. 213, pp. 76–85, 2024. <https://doi.org/10.1016/j.amicard.2023.06.044>
 - [20] Tao, J., Zhao, X., Li, B., Sun, H., Hu, Y., Chen, S., Wu, S., Wu, Y., "Associations of educational attainment and traditional risk factor control with cardiovascular disease," *Am. J. Prev. Cardiol.*, vol. 23, p. 101031, 2025. <https://doi.org/10.1016/j.ajpc.2025.101031>
 - [21] Magnani, J. W., Ning, H., Wilkins, J. T., Lloyd-Jones, D. M., Allen, N. B., "Educational attainment and lifetime risk of cardiovascular disease," *JAMA Cardiology*, vol. 9, p. 45, 2024. <https://doi.org/10.1001/jamacardio.2023.3990>
 - [22] Lynch, J., Kaplan, G. A., Salonen, R., Cohen, R. D., Salonen, J. T., "Socioeconomic status and carotid atherosclerosis," *Circulation*, vol. 92, p. 1786, 1995. <https://doi.org/10.1161/01.CIR.92.7.1786>
 - [23] Davies, N. M., Dickson, M., Davey Smith, G., van den Berg, G. J., Windmeijer, F., "The causal effects of education on health outcomes in the UK Biobank," *Nature Publishing Group, Nat. Hum. Behav.*, vol. 2, no. 2, pp. 117–125, 2018. <https://doi.org/10.1038/s41562-017-0289-7>
 - [24] Rose, G., Marmot, M. G., "Social class and coronary heart disease," *Br. Heart J.*, vol. 45, pp. 13–19, 1981. <https://doi.org/10.1136/hrt.45.1.13>
 - [25] Floud, S., Balkwill, A., Moser, K., Reeves, G. K., Green, J., Beral, V., Cairns, B. J., "The role of health-related behavioural factors in accounting for inequalities in coronary heart disease risk by education and area deprivation: Prospective study of 1.2 million UK women," *BioMed Central, BMC Med.*, vol. 14, p. 145, 2016. <https://doi.org/10.1186/s12916-016-0687-2>
 - [26] Galama, T. J., Lleras-Muney, A., van Kippersluis, H., "The effect of education on health and mortality: A review of experimental and quasi-experimental evidence," *National Bureau of Economic Research, NBER Working Paper No. 24225*, Cambridge, MA, USA, 2018. [Online]. Available: <https://www.nber.org/papers/w24225>
 - [27] Choi, A. I., Weekley, C. C., Chen, S. C., Li, S., Kurella Tamura, M., Norris, K. C., Shlipak, M. G., "Association of educational attainment with chronic disease and mortality: the Kidney Early Evaluation Program (KEEP)," *Elsevier, Am. J. Kidney Dis.*, vol. 58, no. 2, pp. 228–234, 2011. <https://doi.org/10.1053/j.ajkd.2011.02.388>
 - [28] Havranek, E. P., Mujahid, M. S., Barr, D. A., Blair, I. V., Cohen, M. S., Cruz-Flores, S., Davey Smith, G., Dennison Himmelfarb, C. R., Lauer, M. S., Lockwood, D. W., Rosal, M., Yancy, C. W., "Social determinants of risk and outcomes for cardiovascular disease: A scientific statement from the American Heart Association," *American Heart Association, Circulation*, vol. 132, no. 9, pp. 873–898, 2015. <https://doi.org/10.1161/CIR.0000000000000228>
 - [29] Mackenbach, J. P., Cavelaars, A. E., Kunst, A. E., Groenhouf, F., "Socioeconomic inequalities in cardiovascular disease mortality: an international study," *European Heart Journal*, vol. 21, pp. 1141–1151, 2000. <https://doi.org/10.1053/ehj.1999.1990>
 - [30] Meara, E. R., Richards, S., Cutler, D. M., "The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000," *Health Aff.*, vol. 27, pp. 350–360, 2008. <https://doi.org/10.1377/hlthaff.27.2.350>
 - [31] Van Lenthe, F. J., Gevers, E., Joung, I. M. A., Bosma, H., Mackenbach, J. P., "Material and behavioral factors in the explanation of educational differences in incidence of acute myocardial infarction: The Globe study," *Ann. Epidemiol.*, vol. 12, pp. 535–542, 2002. [https://doi.org/10.1016/S1047-2797\(01\)00279-4](https://doi.org/10.1016/S1047-2797(01)00279-4)
 - [32] Loucks, E. B., Buka, S. L., Rogers, M. L., Liu, T., Kawachi, I., Kubzansky, L. D., Martin, L. T., Gilman, S. E., "Education and coronary heart disease risk associations may be affected by early-life common prior causes: a propensity score matching analysis," *Annals of Epidemiology*, vol. 22, pp. 221–232, 2012. <https://doi.org/10.1016/j.annepidem.2012.02.005>
 - [33] Tillmann, T., Vaucher, J., Okbay, A., Pikhart, H., Peasey, A., Kubinova, R., Pajak, A., Tamosiunas, A., Malutina, S., Hartwig, F. S., Fischer, K., Veronesi, G., Bowden, J., Smith, G. D., Bobak, M., Holmes, M. V., "Education and coronary heart disease: Mendelian randomisation study," *BMJ*, vol. 358, p. j3542, 2017. <https://doi.org/10.1136/bmj.j3542>
 - [34] Carter, A. R., Gill, D., Davies, N. M., Taylor, A. E., Tillmann, T., Vaucher, J., Wootton, R. E., Munafò, M. R., Hemani, G., Malik, R., Seshadri, S., Wood, D., Burgess, S., Davey Smith, G., Holmes, M. V., Tzoulaki, I., Howe, L. D., Dehghan, A., "Understanding the consequences of

- education inequality on cardiovascular disease: Mendelian randomisation study," *BMJ Publishing Group, BMJ*, vol. 365, p. 11855, 2019. <https://doi.org/10.1136/bmj.11855>
- [35] Gill, D., Karhunen, V., Malik, R., Dichgans, M., Sofat, N., "Cardiometabolic traits mediating the effect of education on osteoarthritis risk: a Mendelian randomization study," *Elsevier, Osteoarthritis Cartilage*, vol. 29, no. 3, pp. 365–371, 2021. <https://doi.org/10.1016/j.joca.2020.12.015>
- [36] Huang, Z., Huang, S., Ling, H., Lin, Y., Liu, H., Wang, M., "Subjective perception and spatial equity: Geographic characteristics of educational divergence in Guangdong Province," *Springer, Appl. Spat. Anal. Policy*, vol. 18, pp. 98–115, 2025. <https://doi.org/10.1007/s12061-025-09704-z>
- [37] McEntee, J. C., Ogneva-Himmelberger, Y., "Diesel particulate matter, lung cancer, and asthma incidences along major traffic corridors in MA, USA: A GIS analysis," *Health Place*, vol. 14, pp. 817–828, 2008. <https://doi.org/10.1016/j.healthplace.2008.01.002>
- [38] Bahri, M. A. S., Sabri, S., Johar, F., Kararbassi, Z., Majid, M. R., Ludin, A. N. M., "Comparison of spatial autocorrelation analysis methods for distribution pattern of diabetes type II patients in Iskandar Malaysia neighbourhoods," *Universiti Teknologi Malaysia, Proc. Int. Conf. Urban and Regional Planning*, pp. 1–18, 2014.
- [39] Said, S. N. B. M., Zahran, E. S. M. M., Shams, S., "Forest fire risk assessment using hotspot analysis in GIS," *Open J. Civ. Eng.*, vol. 11, pp. 786–801, 2017. <https://doi.org/10.2174/1874149501711010786>
- [40] Mahara, G., Karki, M., Yang, K., Chen, S., Wang, W., Guo, X., "Space-time cluster analysis of tuberculosis incidence in Beijing, China," *Journal of Tuberculosis Research*, vol. 6, pp. 302–319, 2018. <https://doi.org/10.4236/jtr.2018.64027>
- [41] Shariati, M., Mesgari, T., Kasraee, M., Jahangiri-Rad, M., "Spatiotemporal analysis and hotspots detection of COVID-19 using geographic information system," *J. Environ. Health Sci. Eng.*, vol. 18, pp. 1499–1507, 2020. <https://doi.org/10.1007/s40201-020-00547-2>
- [42] Kuznetsov, A., Sadovskaya, V., "Spatial variation and hotspot detection of COVID-19 cases in Kazakhstan, 2020," *Spatial and Spatiotemporal Epidemiology*, vol. 39, p. 100430, 2021. <https://doi.org/10.1016/j.sste.2021.100430>
- [43] Janković, S., Stojisavljević, D., Janković, J., Erić, M., Marinković, J., "Association of socioeconomic status measured by education, and cardiovascular health: a population-based cross-sectional study," *BMJ Publishing Group, BMJ Open*, vol. 4, no. 5, p. e005222, 2014. <https://doi.org/10.1136/bmjopen-2014-005222>
- [44] Maksimović, M. Z., Vlajinac, H. D., Radak, D. J., Maksimović, J. M., Marinković, J. M., Jorga, J. B., "Association of socioeconomic status measured by education and risk factors for carotid atherosclerosis: cross-sectional study," *Croat. Med. J.*, vol. 49, pp. 824–831, 2008. <https://doi.org/10.3325/cmj.2008.49.824>
- [45] Janković, J., Davidović, M., Bjegović Mikanović, V., Janković, S., "Status of cardiovascular health in the Republic of Serbia: Results from the National Health Survey," *Public Library of Science, PLoS ONE*, vol. 14, no. 4, p. e0214505, 2019. <https://doi.org/10.1371/journal.pone.0214505>
- [46] Šašić, B. M., Čanković, D. K., Mijatović Jovanović, V. P., Ukropina, S. N., Ostojić, Z. Z., Milijašević, D. S., Stamenković, K. D., Čanković, S. I., "Socio-demographic differences in the prevalence of cardiovascular diseases and the association with risk factors in the population of Serbia," *Hospital Pharmacology*, vol. 11, pp. 1555–1564, 2024. <https://doi.org/10.5937/hpimj2403555S>
- [47] Kričković, E., Lukić, T., Jovanović-Popović, D., "Geographic medical overview of noncommunicable diseases (cardiovascular diseases and diabetes) in the territory of the AP Vojvodina (Northern Serbia)," *MDPI, Healthcare*, vol. 11, no. 1, p. 48, 2023. <https://doi.org/10.3390/healthcare11010048>
- [48] Markov, S., "Changes in the Educational Structure of the Population of Vojvodina," 2007. [Online]. Available: <https://doiserbia.nb.rs/img/doi/0352-5732/2007/0352-57320722087M.pdf>
- [49] Tomasević, T., Milijašević, D., Ac Nikolic, E., Ukropina, S., Topalov, Z., "Obesity and hypertension as a growing public health problem in Province of Vojvodina, Serbia," *Eur. J. Public Health*, vol. 25, Suppl. 3, p. ckv175.248, 2015. <https://doi.org/10.1093/eurpub/ckv175.248>
- [50] Petrović, V., Čanković, S., "Prevalence of obesity and impact of socio-demographic factors on obesity in 15–19 years old adolescents in Vojvodina, Serbia," *Cent. Eur. J. Public Health*, vol. 31, pp. 287–295, 2023. <https://doi.org/10.21101/ceiph.a7683>
- [51] Kričković, E., Lukić, T., Kričković, Z., Stojić-Milosavljević, A., "Hypertension in AP Vojvodina (Northern Serbia): A spatio-temporal analysis of patients at the Institute for Cardiovascular Diseases of Vojvodina," *Open Geosciences*, vol. 17, p. 200250843, 2025a. <https://doi.org/10.1515/geo-2025-0843>
- [52] Jaraković, M., Mihajlović, B., Čemerlić, S., Adić, F., Sladojević, M., Mihajlović, B., "The level of grammar school students' knowledge on cardiovascular disease risk factors," *Medical Faculty of Novi Sad, Medicinski pregled*, vol. 68, no. 3–4, pp. 98–102, 2015. <https://doi.org/10.2298/MPNS1504098J>
- [53] Institute for Cardiovascular Diseases of Vojvodina, "Database (1995–2023)," Institute for Cardiovascular Diseases of Vojvodina, Sremska Kamenica, Serbia, 2025.
- [54] World Health Organization, *International statistical classification of diseases and related health problems: 10th revision (ICD-10), Volume 1 (2010 ed.)*, Geneva, Switzerland, World Health Organization, 2011. [Online]. Available: <https://www.who.int/standards/classifications/classification-of-diseases>
- [55] World Health Organization, *Classification of diseases*, Geneva, Switzerland. [Online]. Available: <https://www.who.int/standards/classifications/classification-of-diseases>
- [56] SciPy Developers, "scipy.stats – SciPy v1.11.1 Reference Guide," 2025. Accessed: 10 July 2025. [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/stats.html>
- [57] Pandas development team, "pandas: Python Data Analysis Library," 2025. Accessed: 10 July 2025. [Online]. Available: <https://pypi.org/project/pandas/>
- [58] Statsmodels Developers, "Statsmodels: Statistical modeling and econometrics in Python," 2025. Accessed: 10 July 2025. [Online]. Available: <https://www.statsmodels.org/>
- [59] "NumPy" – The fundamental package for numerical computations in Python. Accessed: 10 July 2025. [Online]. Available: <https://numpy.org/>
- [60] Matplotlib Developers, "Matplotlib: Visualization with Python," 2025. Accessed: 10 July 2025. [Online]. Available: <https://matplotlib.org/>
- [61] Turney, S., "Chi-Square (χ^2) Tests | Types, Formula & Examples," 2023. Accessed: 15 July 2025. [Online]. Available: <https://www.scribbr.com/statistics/chi-square-tests/>
- [62] Valarmathi, S., Hemapriya, A. S., Sundar, J. S., "Chi-square tests: A quick guide for health researchers," *IJAR*, vol. 12, pp. 1214–1222, 2024. <https://doi.org/10.21474/IJAR01/19746>
- [63] Nihan, S. T., "Karl Pearson's chi-square tests," *Educ. Res. Rev.*, vol. 15, pp. 575–580, 2020. <https://eric.ed.gov/?id=EJ1267545>
- [64] Rana, R., Singhal, R., "Chi-square test and its application in hypothesis testing," *J. Pract. Cardiovasc. Sci.*, vol. 1, p. 69, 2015. doi:10.4103/2395-5414.157577
- [65] Biswal, A., "Chi-Square Test: Formula, Types, & Examples," *Simplelearn*, Aug. 28, 2025. [Online]. Available: <https://www.simplelearn.com/authors/aviieet-biswal>
- [66] Elhai, J. D., Calhoun, P. S., Ford, J. D., "Statistical procedures for analyzing mental health services data," *Elsevier, Psychiatry Res.*, vol. 160, no. 2, pp. 129–136, 2008. <https://doi.org/10.1016/j.psychres.2007.07.003>
- [67] Khan, R., Phillips, D., Fernando, D., Fowles, J., Lea, R., "Environmental health indicators in New Zealand: Drinking

- water – A case study,” Springer, *EcoHealth*, vol. 4, no. 1, pp. 63–71, 2007. <https://doi.org/10.1007/s10393-007-0089-1>
- [68] Kričković, E., Novković, I., Lukić, T., Kričković, Z., “Using DPSEEA framework to investigate the relation between arsenic concentration values in waters and cancer rates in AP Vojvodina (North Serbia),” *Polish Academy of Sciences, Pol. J. Environ. Stud.*, vol. 31, no. 3, pp. 2117–2129, 2022. <https://doi.org/10.15244/pjoes/143918>
- [69] Kričković, E., Lukić, T., Kričković, Z., Stojić Milosavljević, A., Živanović, M., Srejić, T., “Spatiotemporal and trend analysis of common cancers in men in Central Serbia (1999–2021),” *Open Geosciences*, vol. 17, p. 20250802, 2025b. <https://doi.org/10.1515/geo-2025-0802>
- [70] Getis, A., Ord, J. K., “The analysis of spatial association by use of distance statistics,” *Wiley, Geogr. Anal.*, vol. 24, no. 3, pp. 189–206, 1992.
- [71] Ord, J. K., Getis, A., “Local spatial autocorrelation statistics: Distributional issues and an application,” *Geogr. Anal.*, vol. 27, pp. 286–306, 1995. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x>
- [72] Kričković, E., Lukić, T., Srejić, T., Stojić Milosavljević, A., Stojanović, V., Kričković, Z., “Spatial-temporal and trend analysis of traffic accidents in AP Vojvodina (North Serbia),” *Open Geosciences*, vol. 16, p. 20220630, 2024. <https://doi.org/10.1515/geo-2022-0630>
- [73] Lager, A. C., Torssander, J., “Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 8461–8466, 2012. <https://doi.org/10.1073/pnas.1105839109>
- [74] Sulo, G., Nygård, O., Vollset, S. E., Igland, J., Ebbing, M., Sulo, E., Egeland, G. M., Tell, G. S., “Higher education is associated with reduced risk of heart failure among patients with acute myocardial infarction: a nationwide analysis using data from the CVDNOR project,” *Eur. J. Prev. Cardiol.*, vol. 23, pp. 1743–1750, 2016. <https://doi.org/10.1177/2047487316655910>
- [75] Lakić, D., Tasić, Lj., Kos, M., “Economic burden of cardiovascular diseases in Serbia,” *VSP*, vol. 71, pp. 137–143, 2014. <https://doi.org/10.2298/VSP1402137L>

Emina Kričković is a Research Associate at the University of Belgrade, Faculty of Geography, Serbia. She received a PhD degree in Geosciences with the thesis “Influence of geographic-medical factors on population health in the Republic of Serbia’s Autonomous Province of Vojvodina (AP Vojvodina)” in 2022 at the University of Belgrade, Faculty of Geography. In the same faculty, she received the BSc degree in Geospatial and Environmental Science and the MSc degree in Geospatial and Environmental Science. Her research interests include medical geography and environmental sciences. ORCID: 0000-0002-0737-5636
e-mail: emina.krickovic@gef.bg.ac.rs

Institute “General Stevan Bošković” and a PhD student at the Faculty of Geography, University of Belgrade. His research interests include GIS, spatial analysis and photogrammetry. ORCID:0009-0006-1509-5008
e-mail: zoran.krickovic@vs.rs

Slobodan B. Marković is a Full Professor at the Department of Geography, Tourism and Hotel Management, University of Novi Sad. He has been a corresponding member of the Serbian Academy of Sciences and Arts since 2015. His research focuses on the investigation of loess-paleosol sequences, Quaternary stratigraphy, and the reconstruction of Quaternary climatic and environmental dynamics, as well as paleopedology, geomorphology, and hydrology. ORCID: 0000-0002-4977-634X
e-mail: slobodan.markovic@dgt.uns.ac.rs

Anastazija Stojić Milosavljević is employed at the Institute for Cardiovascular Diseases of Vojvodina, at the Cardiology Clinic. He is a full professor at the Department of Internal Medicine, Faculty of Medicine, University of Novi Sad. She specialized in Internal Medicine and subspecialized in Cardiology in the field of heart muscle diseases. It was the first in the country to introduce trans pulmonary contrast echocardiography into routine use. ORCID: 0000-0003-2092-4289
e-mail: anastazija.stojic@mf.uns.ac.rs

Vladimir M. Cvetković is a Professor in Disaster Risk Management at the University of Belgrade, Faculty of Security Studies. Engaged in disaster studies research for 15 years. Founder and President of the Society for Disaster Risk Management, Director of the International Institute for Disaster Research, and Editor-in-Chief of the International Journal of Disaster Risk Management. Visiting Professor at Technical University of Leoben, Safety and Disaster Man., Austria. ORCID: 0000-0002-3450-0658
e-mail: vmc@fb.bg.ac.rs

Tin Lukić is a Full Professor at the Chair of Physical Geography at the Department of Geography, Tourism, and Hotel Management, Faculty of Sciences, University of Novi Sad. He is also engaged at the Faculty of Geography, University of Belgrade and the International Private School Rudjer Bošković. His research is focused on the investigation of physical geography, geomorphology, natural hazards, climate change studies, and environmental science. ORCID: 0000-0001-5398-0928
e-mail: tin.lukic@dgt.uns.ac.rs

Zoran Kričković is working in the Serbian Ministry of Defence. Additionally, as a researcher at the Military Geographical

Reviewers:

Australia

Abramov, Vyacheslav; Monash University
Begg, Rezaul; Victoria University
Bem, Derek; University of Western Sydney
Betts, Christopher; Pegacat Computing Pty. Ltd.
Buyya, Rajkumar; The University of Melbourne
Chapman, Judith; Australian University Limited
Chen, Yi-Ping Phoebe; Deakin University
Hammond, Mark; Flinders University
Henman, Paul; University of Queensland
Palmisano, Stephen; University of Wollongong
Ristic, Branko; Science and Technology Organisation
Sajjanhar, Atul; Deakin University
Sidhu, Amandeep; University of Technology, Sydney
Sudweeks, Fay; Murdoch University

Austria

Derntl, Michael; University of Vienna
Hug, Theo; University of Innsbruck
Loidl, Susanne; Johannes Kepler University
Linz Stockinger, Heinz; University of Vienna
Sutter, Matthias; University of Innsbruck

Brazil

Parracho, Annibal; Universidad Federal
Fluminense Traina, Agma; University of Sao Paulo
Traina, Caetano; University of Sao Paulo
Vicari, Rosa; Federal University of Rio Grande

Belgium

Huang, Ping; European Commission

Canada

Fung, Benjamin; Simon Fraser University
Grayson, Paul; York University Gray,
Bette; Alberta Education
Memmi, Daniel; UQAM
Neti, Sangeeta; University of Victoria
Nickull, Duane; Adobe Systems, Inc. Ollivier-Gooch,
Carl; The University of British Columbia Paulin,
Michele; Concordia University
Plaisent, Michel; University of Quebec Reid, Keith;
Ontario Ministry of Agriculture Shewchenko,
Nicholas; Biokinetics and Associates Steffan,
Gregory; University of Toronto Vandenbergh,
Christian; HEC Montreal

Czech Republic

Kala, Zdenek; Brno University of Technology
Korab, Vojtech; Brno University of technology
Lhotska, Lenka; Czech Technical University

Finland

Lahdelma, Risto; University of Turku
Salminen, Pekka; University of
Jyvaskyla

France

Cardey, Sylviane; University of Franche-Comte
Klinger, Evelyne; LTCI – ENST, Paris
Roche, Christophe; University of Savoie
Valette, Robert; LAAS - CNRS

Germany

Accorsi, Rafael; University of Freiburg
Glatzer, Wolfgang; Goethe-University
Gradmann, Stefan; Universitat Hamburg
Groll, Andre; University of Siegen
Klamma, Ralf; RWTH Aachen University
Wurtz, Rolf P.; Ruhr-Universitat Bochum

Hungary

Mester, Gyula; Óbuda University, Budapest

India

Pareek, Deepak; Technology4Development
Scaria, Vinod; Institute of Integrative Biology
Shah, Mugdha; Mansukhlal Svayam

Ireland

Eisenberg, Jacob; University College Dublin

Israel

Feintuch, Uri; Hadassah-Hebrew University

Italy

Badia, Leonardo; IMT Institute for Advanced
Studies Berrittella, Maria; University of Palermo
Carpaneto, Enrico; Politecnico di Torino

Japan

Hattori, Yasunao; Shimane University
Livingston, Paisley; Linghan University

Srinivas, Hari; Global Development Research Center
Obayashi, Shigeru; Institute of Fluid Science,
Tohoku University

Netherlands

Mills, Melinda C.; University of Groningen
Pires, Luis Ferreira; University of Twente

New Zealand

Anderson, Tim; Van Der Veer Institute

Portugal

Cardoso, Jorge; University of Madeira
Natividade, Eduardo; Polytechnic Institute of
Coimbra Oliveira, Eugenio; University of Porto

Singapore

Tan, Fock-Lai; Nanyang Technological University

South Korea

Kwon, Wook Hyun; Seoul National University

Spain

Barrera, Juan Pablo Soto; University of Castilla
Gonzalez, Evelio J.; University of La Laguna
Perez, Juan Mendez; Universidad de La
Laguna Royuela, Vicente; Universidad de
Barcelona Vizcaino, Aurora; University of
Castilla-La Mancha Vilarrasa, Clelia Colombo;
Open University of Catalonia

Sweden

Johansson, Mats; Royal Institute of Technology

Switzerland

Niinimäki, Marko; Helsinki Institute of Physics
Pletka, Roman; AdNovum Informatik AG
Rizzotti, Sven; University of Basel
Specht, Matthias; University of Zurich

Taiwan

Lin, Hsiung Cheng; Chienkuo Technology University
Shyu, Yuh-Huei; Tamkang University
Sue, Chuan-Ching; National Cheng Kung
University

United Kingdom

Ariwa, Ezendu; London Metropolitan University
Biggam, John; Glasgow Caledonian University
Coleman, Shirley; University of Newcastle
Conole, Grainne; University of Southampton
Dorfler, Viktor; Strathclyde University
Engelmann, Dirk; University of London
Eze, Emmanuel; University of Hull
Forrester, John; Stockholm Environment Institute
Jensen, Jens; STFC Rutherford Appleton
Laboratory Kolovos, Dimitrios S.; The University
of York McBurney, Peter; University of Liverpool
Vetta, Atam; Oxford Brookes University
WHYTE, William Stewart; University of Leeds
Xie, Changwen; Wicks and Wilson Limited

USA

Bach, Eric; University of Wisconsin Bolzendahl,
Catherine; University of California Bussler,
Christoph; Cisco Systems, Inc. Charpentier,
Michel; University of New Hampshire Chong,
Stephen; Cornell University
Collison, George; The Concord Consortium
DeWeaver, Eric; University of Wisconsin -
Madison Gans, Eric; University of California
Gill, Sam; San Francisco State University
Hunter, Lynette; University of California Davis
Iceland, John; University of Maryland
Kaplan, Samantha W.; University of Wisconsin
Langou, Julien; The University of Tennessee
Liu, Yuliang; Southern Illinois University
Edwardsville Lok, Benjamin; University of Florida
Minh, Chi Cao; Stanford University
Morrissey, Robert; The University of Chicago
Mui, Lik; Google, Inc
Rizzo, Albert; University of Southern
California Rosenberg, Jonathan M.; University
of Maryland Shaffer, Cliff; Virginia Tech
Sherman, Elaine; Hofstra University
Snyder, David F.; Texas State University
Song, Zhe; University of Iowa
Wei, Chen; Intelligent Automation, Inc.
Yu, Zhiyi; University of California

Welcome to IPSI Conferences and Journals!

<http://tir.ipsitransactions.org>

<http://www.ipsitransactions.org>

**CIP – Katalogizacija u publikaciji
Narodna biblioteka Srbije, Beograd**

ISSN 1820 – 4503

**The IPSI Transactions
on Internet Research**

COBISS.SR - ID 119127052

