SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 197

EMPIRICAL EXAMINATION OF AI-POWERED DECISION SUPPORT SYSTEMS: ENSURING TRUST AND TRANSPARENCY IN INFORMATION AND KNOWLEDGE SECURITY

Justyna ŻYWIOŁEK

Czestochowa University of Technology, Faculty of Management; justyna.zywiolek@pcz.pl,
ORCID: 0000-0003-0407-0826

Purpose: The aim of this publication is to present the applications of the use of decision support systems using artificial intelligence to ensure trust in information and knowledge.

Design/methodology/approach: Literature analysis based on the Scopus database, taken into account from 2020 and 2024. Using this analysis, research criteria were developed and 11,315 surveys were conducted among owners and managers of small or medium-sized companies. The survey was conducted using the CAWI method, and respondents rated as many as 5 segments on a Likert scale.

Findings: The manuscript presents an empirical analysis of AI-powered decision support systems (DSS) and their role in ensuring trust and transparency in information and knowledge security. By examining 11,315 surveys from small and medium-sized business owners and managers, the study investigates how these AI systems influence trust in information security practices. The research fills a gap in the literature by providing a comprehensive empirical analysis of AI-based DSS, highlighting their technological intricacies and social implications, and proposing solutions to enhance trustworthiness and decision-making processes in AI applications.

Research limitations/implications: The study has several research limitations. Firstly, it was conducted exclusively in Poland, focusing on small and medium-sized enterprises, which may limit the generalizability of the findings to other geographical regions or larger organizations. Additionally, the sampling technique employed was non-random and based on the researcher's subjective judgment, potentially introducing bias and affecting the representativeness of the sample. The survey method used, Computer-Assisted Web Interviewing (CAWI), and the reliance on self-reported data may lead to response bias, impacting the accuracy of the results. In terms of research implications, the study offers practical solutions to enhance trustworthiness and decision-making processes in AI applications, which can be beneficial for organizations looking to implement AI-driven decision support systems. It also highlights various deficiencies in existing studies, suggesting future research directions, such as investigating the long-term effects of AI on organizational frameworks, exploring ethical implications, and developing new theoretical models.

Originality/value: The manuscript fills the gap in the analysis of AI-based decision support systems for the area of trust in the security of information and knowledge resources.

Keywords: Trust, AI, information and knowledge security, AI- decision system.

Category of the paper: research paper.

http://dx.doi.org/10.29119/1641-3466.2024.197.37

Introduction

In recent years, the integration of AI systems into public sectors worldwide has surged, driven by the aim of enhancing decision-making processes for improved efficiency and reliability of public services (2018). This proliferation has underscored the importance of ensuring the trustworthiness of AI systems, characterized by transparency, explicability, legality, ethics, and robustness. The EU's "Ethics Guidelines for Trustworthy Artificial Intelligence" has served as a pivotal reference point in this discourse (Babel et al., 2021).

However, despite extensive discussions on trustworthy AI, a notable gap persists. While there are studies exploring the societal impact of AI and overviews of trustworthy AI principles, there remains a need for comprehensive empirical analyses of real-life scenarios that concurrently examine both the technological intricacies of AI systems and their social implications (Borenstein, Howard, 2021; Klinova, Korinek, 2021). It is imperative to identify these shortcomings and offer practical solutions without necessitating a complete cessation of AI usage or entirely new technical developments (Giuste et al., 2023).

Although progress has been made in optimizing AI use to align with societal principles and values, particularly within the public sector, there is a dearth of tangible case studies that bring together technological and societal considerations. In this paper, we address this gap by examining the case of AI usage by the Swedish Public Employment Service (PES) through the lens of trustworthy AI principles (Patel et al., 2022; Nguyen et al., 2022).

Our examination encompasses various aspects, including the explainability and interpretability of AI systems and their contribution to fair and equal treatment. We uncover numerous challenges, such as the opacity of neural network AI systems, inadequacy in explanations, and difficulties in contesting decisions. In response, we propose potential solutions to enhance decision-making processes and bolster the trustworthiness of AI, including increased stakeholder participation, expansion of professional discretion, and improvement in performance indicators (Yang et al., 2022).

The structure of our paper is as follows: We first outline our main theories, define key concepts, and present a framework for assessing AI system trustworthiness in public decision-making (Section 2). We then detail our research methodology and empirical material (Section 3). Subsequently, we delve into a case study of AI-assisted decision-making in the Swedish Public Employment Service, applying our framework and discussing improvement strategies (Section 4). We analyse the results theoretically (Section 5), discuss theoretical implications, make recommendations, and address limitations (Section 6). Finally, we conclude with a summary of our findings and reflections on the future landscape of trustworthy AI in public decision-making (Section 7).

Ensuring Integrity: The Path to Reliable and Ethical AI Decision-Making

When analysing the integration of Artificial Intelligence (AI) into public decision-making processes, grounding our discussion within robust theoretical frameworks is essential. Here, we draw upon three established theoretical perspectives, as effectively utilized by Di Vaio and colleagues in a similar context (Donins, Behmane, 2023; Emaminejad et al., 2024).

Firstly, Institutional Theory (Gerke et al,, 2020) provides insights into how societal norms, rules, and expectations shape organizational behaviour and decision-making in the public sector. This perspective is particularly relevant for understanding the Swedish Public Employment Service's adoption of new technologies like AI and Big Data (BD), influenced by both external pressures and internal dynamics (Zywiolek et al., 2024).

Secondly, the Resource-Based View (RBV) emphasizes leveraging internal resources, including technological infrastructure, skilled personnel, and organizational knowledge. Understanding how these assets are utilized to harness the capabilities of AI, BD, and Data Intelligence and Analytics (DI&A) for enhancing public sector decision-making processes is crucial (Wu et al., 2024).

Lastly, Ambidexterity Theory explores balancing the exploitation of existing resources with the exploration of new technological opportunities. This theory is essential for understanding how the Swedish Public Employment Service maintains operational efficiency while integrating emerging technologies like DI&A, AI, BD, and Human-Artificial Intelligence (HAI) for decision-making (Schia et al., 2019; Shang et al., 2024).

Building upon these theoretical frameworks, we explore core technological concepts such as DI&A, AI, BD, and HAI within the context of the Swedish Public Employment Service. These technologies are not standalone tools but part of a larger system intertwined with organizational practices and policies. DI&A processes vast amounts of data, AI enhances decision-making accuracy and efficiency, and BD represents the extensive data landscape feeding into these processes (Prieto et al., 2023; Li et al., 2021).

The interplay between theoretical frameworks and technological concepts forms the foundation of our exploration into the Swedish Public Employment Service's application of AI. We critically examine how these elements collectively contribute to understanding the challenges regarding the trustworthiness of AI systems in public decision-making and propose strategies to address them (Lu et al., 2020; Kumar, Suthar, 2024). Our analysis covers aspects such as the explainability and interpretability of AI decisions, alignment with legal and ethical standards, and integration within the broader organizational context.

Next, we delve into six key principles for trustworthy AI. The first principle focuses on performance evaluation at various levels, including system-wide, sub-population specific, and outcomes. We also consider how AI affects human decision-making, termed "augmented performance (Shang et al., 2024)". The second principle, calibration, addresses the system's ability to accurately estimate confidence levels in its decisions. We discuss the significance of well-calibrated systems in providing stakeholders with reliable information about AI-generated decisions (Varriale et al., 2023; Wang et al., 2021).

A trustworthy AI system relies on several key principles to ensure reliability and fairness in public decision-making processes. Firstly, calibration and confidence communication are crucial, allowing users to appropriately trust predictions based on the system's certainty levels. Secondly, interpretability, explainability, and intelligibility are essential for understanding an AI's decision-making logic (Żywiołek, Schiavone, 2021; Wu et al., 2018). While interpretability varies among AI types, explainability methods can approximate complex models' internal logic. Intelligibility ensures that decision-making processes are communicated effectively to stakeholders. Thirdly, fair and equal treatment is paramount, requiring AI systems to treat similar cases consistently and address potential biases based on various demographics. Fourthly, legality, negotiation, and appeal ensure that AI systems operate within legal frameworks, allowing individuals to contest decisions and understand the reasoning behind them (Sibbald et al., 2024). Overall, these principles contribute to building trustworthy AI systems that prioritize transparency, fairness, and legality in public decision-making contexts.

Finally, the last principle revolves around accountability and human oversight. This entails structuring the system in a manner that enables human decision-makers to be accountable for the decisions made with the assistance of AI. For instance, if caseworkers bear formal responsibility for decisions, they must have the capability to oversee the AI and make final decisions, taking into consideration the AI's output as well as other pertinent factors (Petersson et al., 2022).

It's worth noting that while principles such as beneficence, non-maleficence, privacy, and autonomy are not explicitly discussed here, they are either implicitly addressed or deliberately excluded to maintain depth and focus in our discussion. Enhanced efficiency, accuracy, and equitable treatment minimize variability in decision-making, aligning with the principle of beneficence, while non-maleficence is upheld by avoiding harm. Autonomy is preserved through elements such as explainability, interpretability, and negotiation, empowering the subjects' agency within the decision-making process. Privacy is ensured through legality, while technical robustness is omitted due to its complexity within this context (Mantha, García de Soto, 2021b).

In assessing the trustworthiness of an AI system, it's crucial to compare decision-making scenarios with and without the use of AI. If trustworthiness is greater for AI-assisted decision-making, there are reasons to trust it, and vice versa. While some may argue that an AI system only needs to be sufficiently reliable, not fully reliable, determining a precise sufficiency threshold often requires comparisons, typically leading to a comparative state. Stakeholders must have valid reasons to trust an AI system to improve outcomes and procedures regarding the principles listed above, or if some aspects remain unchanged while others are enhanced, it can be considered an instance of what this paper terms "trustworthy AI (Liu, 2022)".

Given the preceding discussion, Table 1 presents the questions that should be posed to ascertain whether there are grounds to believe that an AI system, or the collaboration of human decision-makers and AI, can be deemed trustworthy (Żywiołek et al., 2021; Hlávka, 2020).

Table 1.
Evaluation Framework for AI Systems: Comprehensive Criteria

no	Criteria	Evaluation Questions
1	Performance	a. The accuracy of AI judgments or decisions at all levels is a critical aspect to evaluate. This involves assessing how reliably the AI system predicts outcomes across various scenarios and populations (Amann et al., 2020; Nawshin et al., 2024; Liang et al., 2024). b. Another important consideration is whether human decision-makers achieve greater accuracy when assisted by the AI system. Understanding the comparative performance of human decision-makers with and without AI assistance provides insights into the system's effectiveness (Hagendorff, 2020). c. Communication of the system's performance to stakeholders is essential for transparency and trust. Stakeholders need to be informed about the AI system's performance metrics, including its accuracy, reliability, and limitations, to make informed decisions and assess its impact on decision-making processes (Kosinski et al., 2013; Khan et al., 2023).
2	Calibration	a. Are stakeholders provided with confidence estimates regarding the AI's decisions or judgments (Żywiołek et al., 2024; Kshetri, 2021)? b. If confidence estimates are provided, are they well-calibrated, meaning do they accurately reflect the AI's level of certainty or uncertainty regarding its decisions (Mittelstadt et al., 2019)?
3	Interpretability and Explainability	a. Is the decision-making logic understandable in principle to stakeholders? b. Do the explanations provided accurately represent the actual decision-making logic? (Ibeneme et al., 2021; Jobin et al., 2019).
4	Intelligibility and Availability	a. Is the decision-making logic accessible to stakeholders? b. Are the explanations presented in a way that stakeholders can understand them in practice? (Dwivedi et al., 2021).
5	Equal and Fair Treatment	a. Does the AI consistently make decisions? b. Are relevant aspects of fair treatment fulfilled? (Barredo Arrieta et al., 2020).
6	Legality, Negotiation, and Appeal	a. Does the AI system's usage and functionality comply with the law? b. To what extent does the AI system enable affected individuals to negotiate or appeal unfavorable decisions? (Żywiołek et al., 2022; Raban, Hauptman, 2018).
7	Accountability and Human Oversight	Are human decision-makers able to oversee the operation of the AI and make independent decisions based on the system's output? (Sunarti et al., 2021; Phillips et al., 2021; Batool et al., 2023).

AI for information and knowledge security

Artificial intelligence (AI) poses a significant challenge when it comes to safeguarding information and knowledge, since it involves several technical, ethical, and legal dimensions. The following are the primary areas that require attention.

1. Data security Artificial intelligence frequently processes extensive datasets that may include confidential or sensitive material. Data anonymization is a crucial step that needs to be taken before processing data in order to effectively prevent the identity of persons.

2. Data encryption is essential for ensuring the security of data (García de Soto et al., 2022b). It is important to encrypt data both when it is stored and when it is being transferred to prevent unauthorised individuals from gaining access to it. Data access management involves the implementation of stringent procedures to ensure that only individuals with proper authorization are able to process and analyse the data.
3. Ensuring the security of algorithms Incorporate security through design: AI algorithms should be developed with security as a primary consideration from the beginning. Testing and audits: Regular security testing and source code audits are crucial for identifying and resolving any security vulnerabilities (Żywiołek et al., 2021; Mantha, García de Soto, 2021a).
4. Clarity and comprehensibility Decision comprehensibility: Algorithms should be formulated in a manner that enables the understanding of the process by which decisions are reached. Consequently, it is imperative to create artificial intelligence models that provide enhanced transparency and interpretability (Aloqaily et al., 2022). Documentation is essential for AI systems since it allows for the thorough recording of decision-making processes and data. This documentation enables the system to be audited and facilitates a clear understanding of how the system operates.
5. Mitigation of potential risks Prior to adopting an AI system, it is crucial to conduct a risk assessment in order to detect any potential information security issues. Contingency Planning: Create comprehensive contingency plans to promptly and efficiently address security breaches (García de Soto et al., 2022a).
6. Ethical considerations and adherence to regulations Legal Compliance: AI systems must adhere to relevant data protection standards, such as the General Data Protection Regulation (GDPR) in Europe. Organisations must prioritise the ethical utilisation of data by actively avoiding prejudice and ensuring that AI technologies are employed in a just and transparent manner (Ojha et al., 2024).
7. Ongoing surveillance and revision System upgrades are crucial for maintaining the security of AI systems. Regular updates are necessary to safeguard against emerging threats. Anomaly monitoring involves the ongoing surveillance of AI systems to identify any abnormal activities that could potentially suggest security vulnerabilities (Taddeo et al., 2019).

In conclusion, safeguarding information and knowledge security in the realm of artificial intelligence necessitates a comprehensive strategy that integrates sophisticated data protection methods, transparent processes, risk assessment and mitigation, adherence to regulations, and an ethical framework for the utilisation of AI technology.

After analyzing the literature, the author also analyzed the Scopus database, indicating the keywords AI, security and trust, obtaining 327 articles published in 2020-2024. Based on this database, a map was developed illustrating the keywords indicated by the authors of the articles. Drawing conclusions from them (figure 1), it can be seen how broad this area of research is.

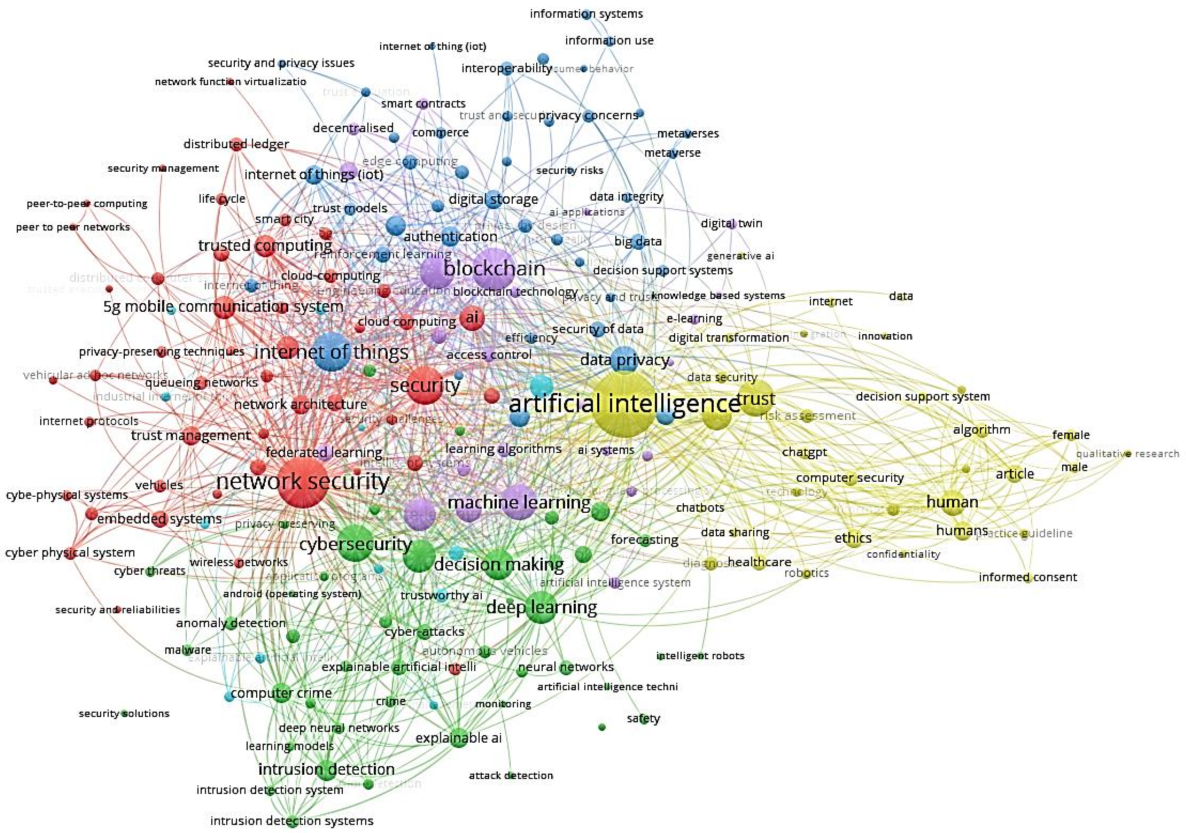

Figure 1. A map showing keywords for the research area of AI, security and trust based on the Scopus database.

Methods, materials and results

The study was carried out in Poland. The acquired database had 11,315 respondents, specifically corporate management or their owners. It was hypothesised that manufacturing organisations would possess the largest amount of data, and this study took into consideration small and medium-sized enterprises. A research study was done from 2020 to 2024. The prolongation of the study period was a consequence of the societal limitations imposed as a result of the COVID-19 epidemic. The sample selection was deliberate, employing a non-random sampling technique where individuals were chosen based on the researcher's subjective judgement. The research was carried out by the primary author of this study. The sole criterion for deliberate selection is that the survey participant fulfils a specified requirement. This requirement is the initial filtering question of the survey, which asks whether the responder possesses knowledge about information security and utilises artificial intelligence in their activities. Following this, a subsequent question was asked to assess the research participant's understanding of the concepts of information and knowledge processing, as well as their awareness of security measures. The research instrument employed was a survey.

The Computer-Assisted Web Interviewing (CAWI) technique was employed. The questionnaire employed a five-point Likert scale, with 1 representing the minimum value and 5 representing the maximum value. The results presented are of a general nature as they do not pertain to specific sequences of operations. Instead, they consist of respondents' evaluations of the influence of AI on the advancement of crucial aspects of information and knowledge security in Poland. Technologies, often known as pillars, are categorised into five segments: BW, BI, BP, AI, and Z (Figure 2).

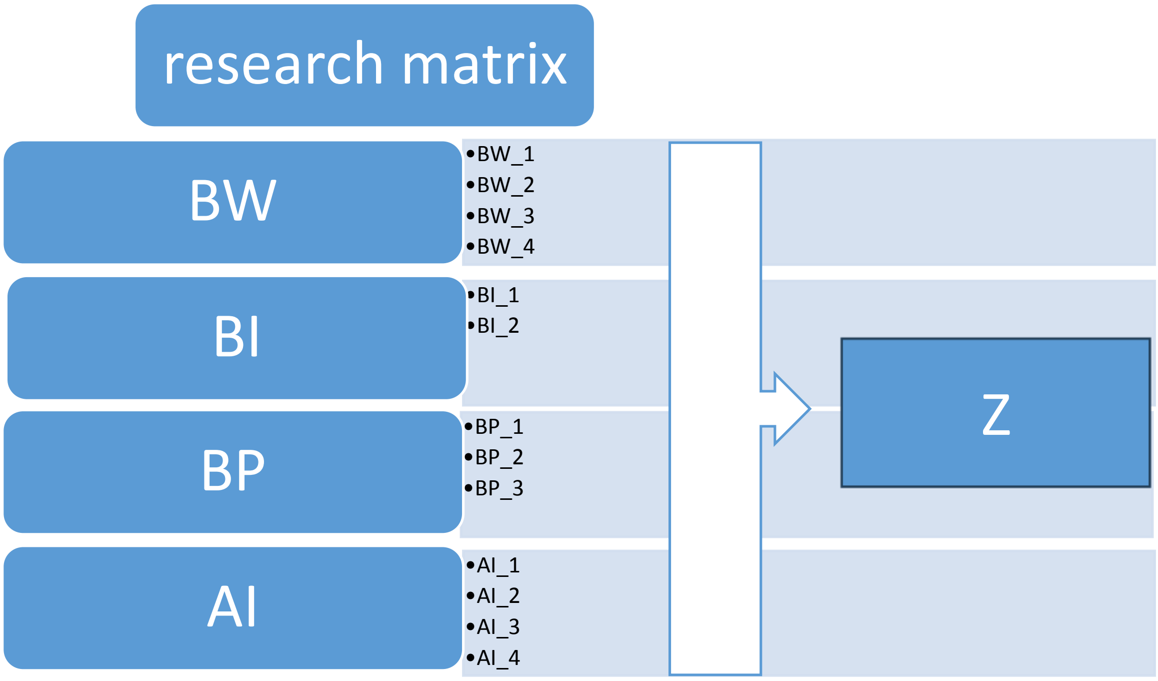

Figure 2. Research matrix.

Knowledge security and consisted of the following solutions, abbreviated as BW in the research:

BW_1: IT system for knowledge management.

BW_2: Automation of knowledge-building data sets.

BW_3: automation of knowledge exchange among employees in various positions.

BW_4: Multi-tasking systems with knowledge management algorithms.

BI was called: information security and consisted of a physical and IT factor:

BI_1: physical security of information processing equipment.

BI_2: IT security of information processing equipment.

BP was called: [Enterprise security] and consisted of:

BP_1: Mobile communication and connectivity technologies.

BP_2: Development and compatibility of supply chain computer systems (SCM)

BP_3: Network and chain integration of enterprise computer systems (end-to-end supply chain).

The AI was called: [AI] and included:

AI_1: Devices and technologies (wireless sensors, RFID sensors) generating real-time data on the operation of machines and technological installations.

AI_2: Extended databases (Big Data), process visualization, augmented reality, simulation calculations.

AI_3: Access to cloud services and the Industrial Internet of Things (IIoT).

AI_4: AI tools for manufacturing enterprises.

Named With: Rated [Security and Cybersecurity]. No sub-segmentation was used in the segment due to the strong connection between the operations (structure) and all the previously mentioned elements and not to indicate specific technological solutions.

Studying the impact of the security of information and knowledge resources on the development of the use of AI in manufacturing companies and trust in AI seems obvious, but the author of the study wanted to determine the degree of change using a Likert scale from 1 to 5.

Table 2.
Factors subject to evaluation

Factor	No response	Likert scale				
		1	2	3	4	5
BW_1	1,28%	0,00%	2,95%	7,67%	36,92%	52,46%
BW_2	0,07%	0,00%	3,11%	9,83%	31,39%	55,67%
BW_3	1,13%	0,00%	2,09%	2,86%	34,20%	60,85%
BW_4	2,08%	0,00%	1,78%	3,27%	33,89%	61,06%
BI_1	1,42%	0,00%	1,62%	2,59%	39,13%	56,67%
BI_2	5,19%	0,00%	1,29%	4,09%	33,43%	53,20%
BP_1	3,42%	0,00%	3,17%	13,51%	38,73%	44,59%
BP_2	0,12%	0,00%	4,25%	15,74%	32,39%	47,62%
BP_3	1,92%	0,00%	5,24%	9,61%	39,81%	45,34%
AI_1	2,43%	0,00%	6,31%	4,26%	40,38%	49,05%
AI_2	1,86%	0,00%	4,16%	7,35%	39,86%	48,63%
AI_3	1,14%	0,00%	3,89%	2,49%	44,17%	49,45%
AI_4	0,72%	0,00%	13,45%	6,11%	29,28%	51,16%

Theoretical implications for empirically examining AI-powered decision support systems in guaranteeing trust and transparency in information and knowledge security. Examining AI-powered decision support systems (DSS) in the context of assuring trust and transparency in information and knowledge security has several important theoretical implications. By drawing on the theoretical insights presented in the discussed articles, we can identify numerous significant implications.

The research enhances current understanding by combining AI technology with ideas of information security and knowledge management. This integration offers novel perspectives on how artificial intelligence might be utilised to augment the security of information systems. This text focuses on the prominent technological aspects that impact the efficiency of artificial intelligence (AI) in safeguarding information and overseeing knowledge within organisations.

This theoretical fusion establishes a basis for future study to delve deeper into the correlation between AI capabilities and information security procedures.

The Role of Artificial Intelligence in Building Trust in Decision-Making: The study expands theoretical understanding of the function of AI in building trust by examining how AI might enhance decision-making processes. User trust is contingent upon the transparency and interpretability of AI systems. AI systems must incorporate explainability and transparency as fundamental elements in their design to enhance user trust and acceptance of AI-driven judgements. This has significant theoretical implications. This is consistent with the wider theoretical frameworks of human-AI interaction and confidence in technology.

The research highlights the crucial role that advanced information systems play in the effective adoption and application of AI technologies. This highlights the necessity for strong information systems that are capable of managing the intricacies of AI algorithms and data processing. This discovery improves the theoretical comprehension of how information systems facilitate and merge with AI technologies, offering a direction for future research to explore the infrastructure needs for AI in other fields.

The paper establishes a theoretical connection between the integration of AI technology in decision support systems and the attainment of sustainable development goals. This study delves into the ways in which artificial intelligence (AI) can enhance the efficiency of resource utilisation, minimise wastage, and promote sustainable practices. Consequently, it contributes to the advancement of theoretical understanding on the intersection of technology and sustainability. This viewpoint promotes additional investigation into the potential applications of AI in achieving environmental and social goals, in line with wider sustainability ideas.

Key Factors to Consider in Cybersecurity for AI-Enabled Decision Support Systems: A crucial theoretical consequence is the inclusion of cybersecurity in the implementation of AI-driven DSS. The study emphasises the necessity of implementing comprehensive cybersecurity policies in order to safeguard AI systems against vulnerabilities and threats. This contribution enhances the theoretical discussion on cybersecurity by integrating risks and strategies related to artificial intelligence, thereby offering a more nuanced comprehension of security within the realm of advanced technology.

The project aims to enhance the theoretical comprehension of the collaboration between humans and AI by investigating its influence on the efficiency and efficacy of decision-making. This demonstrates the capacity of AI to enhance human decision-making, resulting in decisions that are more informed and more precise. The collaboration between humans and AI is essential for theoretical models that aim to elucidate the dynamics of human-AI collaboration and its impact on organisational decision-making processes.

Future Research Directions: The paper highlights various deficiencies in the existing research and suggests potential avenues for investigating the theoretical elements of artificial intelligence in decision support systems. This entails analysing the enduring effects of AI on organisational frameworks, the ethical implications of AI implementation, and the formulation of novel theoretical frameworks to enhance comprehension of AI's function in intricate decision-making contexts.

The study explores the theoretical implications of AI-powered decision support systems, offering a comprehensive framework for future research. It highlights the significance of trust, transparency, and security in the digital era. These insights are essential for furthering theoretical understanding and directing empirical research into the actual uses of AI in many industries.

Discussion

The empirical examination of AI-powered decision support systems (DSS) for ensuring trust and transparency in information and knowledge security reveals several critical insights that align with broader Industry 4.0 principles. This study builds on the understanding that integrating advanced technologies such as AI into decision support frameworks can significantly enhance the security and transparency of information systems, thereby fostering trust among stakeholders.

The integration of AI technologies in decision support systems emphasizes the role of advanced algorithms and machine learning in identifying, predicting, and mitigating security threats. This aligns with the broader Industry 4.0 focus on leveraging technology to optimize processes and improve operational efficiency. AI's ability to analyse large datasets in real-time and provide predictive insights is crucial for enhancing the security of information systems, as highlighted by various studies on Industry 4.0 technologies in manufacturing and other sectors.

Ensuring trust and transparency in AI-powered DSS is paramount. The study highlights the importance of designing AI systems with explainability and interpretability to gain user trust. This finding is consistent with the theoretical perspectives on human-AI interaction, where transparency in AI decision-making processes is critical for user acceptance and trust. The gradual building of trust in technologies like cloud computing and IIoT, as observed in the steel sector, underscores the need for clear, understandable AI mechanisms.

The study underscores the necessity of robust information systems to support AI technologies, similar to the integration of information systems in Industry 4.0 frameworks. Effective information systems enable seamless data flow and process integration, crucial for the successful deployment of AI in decision support. This is particularly important for managing large datasets and ensuring real-time data processing, which are essential for AI's efficacy in enhancing security and transparency.

The research highlights the role of cybersecurity measures and blockchain technology in protecting sensitive information and ensuring traceability. Blockchain's ability to provide immutable records and enhance transparency is crucial for building trust in digital ecosystems, reflecting the theoretical underpinnings of Industry 4.0's emphasis on secure, transparent operations. The application of blockchain in ensuring the integrity and traceability of data aligns with the broader objectives of Industry 4.0 to enhance supply chain transparency and security.

The study explores the potential of AI technologies to contribute to sustainable development goals (SDGs) by optimizing resource use and reducing waste. This finding expands the theoretical understanding of how AI can support sustainable practices, a key aspect of Industry 4.0's emphasis on sustainable production and resource efficiency.

The study provides a comprehensive framework for integrating AI technologies into decision support systems, emphasizing the importance of transparency, security, and trust. This framework can guide future research and practical implementations of AI in various sectors, ensuring that AI systems are designed with these critical factors in mind.

By empirically examining the impact of AI-powered DSS, the study offers concrete evidence of the benefits of AI in enhancing information security and transparency. This empirical data supports the theoretical claims about the potential of AI to improve decision-making processes and security measures in information systems.

The study extends theoretical knowledge by exploring the application of Industry 4.0 technologies such as blockchain, IIoT, and cloud computing in the context of AI-powered DSS. It provides valuable insights into how these technologies can be leveraged to create secure, transparent, and efficient information systems, contributing to the broader discourse on Industry 4.0.

The research identifies key technological enablers that are critical for the successful implementation of AI-powered DSS. This includes advanced manufacturing technologies, smart sensors, and real-time data analytics, which are essential for building secure and transparent information systems.

Conclusion

The integration of AI-powered decision support systems (DSS) into various sectors offers significant potential for enhancing information and knowledge security. This study aims to explore how these systems can ensure trust and transparency, particularly focusing on the context of information and knowledge security.

The study builds on existing theories of technology and supply chain management, emphasizing the role of AI in improving decision-making processes. It highlights the critical importance of integrating advanced technologies such as AI, blockchain, and IIoT to secure information systems and manage knowledge effectively.

AI technologies can analyse large datasets in real-time to identify, predict, and mitigate security threats. This capability is crucial for enhancing the security of information systems, aligning with the broader principles of Industry 4.0 which focus on leveraging technology to optimize processes and improve operational efficiency.

The study emphasizes the need for AI systems to be designed with transparency and explainability to gain user trust. Ensuring that AI decision-making processes are understandable and interpretable is vital for fostering trust among stakeholders.

Effective implementation of AI technologies requires robust information systems that can handle complex data processing and ensure seamless data flow. This infrastructure supports the successful deployment of AI in enhancing information security and transparency.

Cybersecurity measures and blockchain technology play a significant role in protecting sensitive information and ensuring traceability. Blockchain provides immutable records that enhance transparency, which is crucial for building trust in digital ecosystems.

AI technologies can contribute to sustainable development goals by optimizing resource use, reducing waste, and supporting sustainable practices. This aligns with the broader objectives of Industry 4.0 to enhance sustainability and resource efficiency. Prioritize AI Integration: Managers should prioritize the integration of AI technologies into decision support systems to enhance security and transparency.

Focus on Transparency: Emphasizing the transparency and explainability of AI systems can help build user trust and facilitate broader acceptance of AI-driven decisions. Invest in Robust Information Systems: Ensuring robust and scalable information systems is critical for the effective deployment of AI technologies.

Leverage Blockchain for Security: Utilizing blockchain technology can significantly enhance the traceability and security of information within supply chains.

The empirical examination of AI-powered decision support systems highlights their potential to significantly enhance information and knowledge security by ensuring trust and transparency. The study underscores the importance of integrating advanced technologies, such as AI, blockchain, and IIoT, to create secure and transparent information systems. These findings provide valuable insights for researchers, practitioners, and policymakers, contributing to the broader discourse on the technological transformation of supply chains and other sectors.

By aligning with the principles of Industry 4.0, the study offers a robust framework for future research and practical implementations, emphasizing the need for ongoing technological advancements to enhance the efficiency and resilience of information and knowledge security systems.

References

1. Aloqaily, M., Kanhere, S., Bellavista, P., Nogueira, M. (2022). Special Issue on Cybersecurity Management in the Era of AI. *J. Netw. Sys.t Manage., 30*.
2. Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V.I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making, 20,* 310.
3. Artificial intelligence for the real world (2018). *Harvard Business Review, 96*, 108.
4. Babel, A., Taneja, R., Mondello Malvestiti, F., Monaco, A., Donde, S. (2021). Artificial Intelligence Solutions to Increase Medication Adherence in Patients With Non-communicable Diseases. *Frontiers in Digital Health, 3*, 669869.
5. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R. et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58,* 82-115.
6. Batool, K., Zhao, Z.-Y., Irfan, M., Żywiołek, J. (2023). Assessing the role of sustainable strategies in alleviating energy poverty: an environmental sustainability paradigm. *Environ. Sci. Pollut. Res. Int., 30*, 67109-67130.
7. Borenstein, J., Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics, 1*, 61-65.
8. Donins, U., Behmane, D. (2023). Challenges and Solutions for Artificial Intelligence Adoption in Healthcare – A Literature Review. In: *Innovation in Medicine and Healthcare: Proceedings of 11th KES-InMed 2023,* Chen (ed.), pp. 53-62. Singapore: Springer Singapore.
9. Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A. et al. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994
10. Emaminejad, N., Kath, L., Akhavian, R. (2024). Assessing Trust in Construction AI-Powered Collaborative Robots Using Structural Equation Modeling. *J. Comput. Civ. Eng., 38*.
11. García de Soto, B., Georgescu, A., Mantha, B., Turk, Ž., Maciel, A., Semih Sonkor, M. (2022a). Construction cybersecurity and critical infrastructure protection: new horizons for Construction 4.0. *ITcon. 27*, 571-594.
12. García de Soto, B., Turk, Ž., Maciel, A., Mantha, B., Georgescu, A., Sonkor, M.S. (2022b). Understanding the Significance of Cybersecurity in the Construction Industry: Survey Findings. *Journal of Construction Engineering and Management, 148.*

13. Gerke, S., Minssen, T., Cohen, G., Gerke, S., Minssen, T., Cohen, G. (2020). *Artificial intelligence in healthcare*. London, San Diego, CA: Academic Press, imprint of Elsevier.
14. Giuste, F., Shi, W., Zhu, Y., Naren, T., Isgut, M., Sha, Y., Tong, L., Gupte, M., Wang, M.D. (2023). Explainable Artificial Intelligence Methods in Combating Pandemics: A Systematic Review. *IEEE Reviews in Biomedical Engineering, 16*, 5-21.
15. Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines, 30*, 99-120.
16. Hlávka, J.P., Hlávka, J.P. (2020). *Artificial intelligence in healthcare.* London, San Diego, CA: Academic Press, imprint of Elsevier.
17. Ibeneme, S., Okeibunor, J., Muneene, D., Husain, I., Bento, P., Gaju, C., Housseynou, B., Chibi, M., Karamagi, H., Makubalo, L. (2021). Data revolution, health status transformation and the role of artificial intelligence for health and pandemic preparedness in the African context. *BMC Proceedings, 15*, 22.
18. Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat. Mach. Intell., 1,* 389-399.
19. Khan, M.A., Kumar, N., Mohsan, S.A.H. Khan, W.U., Nasralla, M.M., Alsharif, M.H., Żywiołek, J., Ullah, I. (2023). Swarm of UAVs for Network Management in 6G: A Technical Review. *IEEE Trans. Netw. Serv. Manage., 20,* 741-761.
20. Klinova, K., Korinek, A. (2021). *AI and Shared Prosperity,* ACM.
21. Kosinski, M., Stillwell, D., Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, 110,* 5802-5805.
22. Kshetri, N. (2021). Economics of Artificial Intelligence in Cybersecurity. *IT Prof., 23*, 73-77.
23. Kumar, D., Suthar, N. (2024). Ethical and legal challenges of AI in marketing: an exploration of solutions. *JICES.*
24. Li, J.-P.O., Liu, H., Ting, D.S.J., Jeon, S., Chan, R.V.P., Kim, J.E., Sim, D.A., Thomas, P.B.M., Lin, H., Chen, Y. et al. (2021). Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. *Progress in Retinal and Eye Research, 82, 100900*.
25. Liang, C.-J., Le, T.-H., Ham, Y., Mantha, B.R., Cheng, M.H., Lin, J.J. (2024). Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry. *Automation in Construction, 162, 105369.*
26. Liu (2022). Trustworthy AI: a computational perspective. *ACM Transactions on Intelligent Systems and Technology, 14, 1.*
27. Lu, Y., Huang, X., Dai, Y., Maharjan, S., Zhang, Y. (2020). Federated Learning for Data Privacy Preservation in Vehicular Cyber-Physical Systems. *IEEE Network, 34,* 50-56.
28. Mantha, B.R., García de Soto, B. (2021a). Assessment of the cybersecurity vulnerability of construction networks. *ECAM, 28,* 3078-3105.

29. Mantha, B.R.K., García de Soto, B. (2021b). Cybersecurity in Construction: Where Do We Stand and How Do We Get Better Prepared. *Front. Built Environ., 7.*
30. Mittelstadt, B., Russell, C., Wachter, S. (2019). E*xplaining Explanations in AI*, ACM.
31. Nawshin, F., Unal, D., Hammoudeh, M., Suganthan, P.N. (2024). AI-powered malware detection with Differential Privacy for zero trust security in Internet of Things networks. *Ad Hoc Networks, 161*, 103523.
32. Nguyen, T.V., Dakka, M.A., Diakiw, S.M., VerMilyea, M.D., Perugini, M., Hall, J.M.M., Perugini, D (2022) A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical data. *Scientific Reports* 12: 8888.
33. Ojha, N.K., Pandita, A., Ramkumar, J. (2024). Cyber Security Challenges and Dark Side of AI. In: *Demystifying the Dark Side of AI in Business* (pp. 117-137). Dadwal, S. (ed.). IGI Global.
34. Patel, V.A., Bhattacharya, P., Tanwar, S., Gupta, R., Sharma, G., Bokoro, P.N., Sharma, R. (2022). Adoption of Federated Learning for Healthcare Informatics: Emerging Applications and Future Directions. *IEEE Access, 10*, 90792-90826.
35. Petersson, L., Larsson, I., Nygren, J.M., Nilsen, P., Neher, M., Reed, J.E., Tyskbo, D., Svedberg, P. (2022). Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. *BMC Health Services Research, 22,* 850.
36. Phillips, P.J., Hahn, C.A., Fontana, P.C., Yates, A.N., Greene, K., Broniatowski, D.A., Przybocki, M.A. (2021). *Four principles of explainable artificial intelligence.* Gaithersburg, MD: National Institute of Standards and Technology (U.S.)
37. Prieto, S.A., Mengiste, E.T., García de Soto, B. (2023). Investigating the Use of ChatGPT for the Scheduling of Construction Projects. *Buildings, 13,* 857.
38. Raban, Y., Hauptman, A. (2018). Foresight of cyber security threat drivers and affecting technologies. *FS, 20,* 353-363.
39. Schia, M.H., Trollsås, B.C., Fyhn, H., Lædre, O. (2019). *The Introduction of AI in the Construction Industry and Its Impact on Human Behavior*. International Group for Lean Construction.
40. Shang, Y., Zhou, S., Zhuang, D., Żywiołek, J., Dincer, H. (2024). The impact of artificial intelligence application on enterprise environmental performance: Evidence from microenterprises. *Gondwana Research, 131,* 181-195.
41. Sibbald, M., Zwaan, L., Yilmaz, Y., Lal, S. (2024). Incorporating artificial intelligence in medical diagnosis: A case for an invisible and (un)disruptive approach. *Journal of Evaluation in Clinical Practice, 30,* 3-8.
42. Sunarti, S., Fadzlul Rahman, F., Naufal, M., Risky, M., Febriyanto, K., Masnina, R. (2021). Artificial intelligence in healthcare: opportunities and risk for future. *Gaceta Sanitaria, 35, Suppl 1,* S67-S70.

43. Taddeo, M., McCutcheon, T., Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence, 1,* 557-560.
44. Varriale, V., Cammarano, A., Michelino, F., Caputo, M. (2023). Critical analysis of the impact of artificial intelligence integration with cutting-edge technologies for production systems. *J. Intell. Manuf.*
45. Wang, S., Chen, Z., Xiao, Y., Lin, C. (2021). Consumer Privacy Protection With the Growth of AI-Empowered Online Shopping Based on the Evolutionary Game Model. *Frontiers in Public Health, 9*, 705777.
46. Wu, D., Ren, A., Zhang, W., Fan, F., Liu, P., Fu, X., Terpenny, J. (2018). Cybersecurity for digital manufacturing. *Journal of Manufacturing Systems, 48,* 3-12.
47. Wu, J., Yang, R., Zhao, P., Yang, L. (2024). Computer-aided mobility solutions: Machine learning innovations to secure smart urban transportation. *Sustainable Cities and Society, 107*, 105422.
48. Yang, Z., Silcox, C., Sendak, M., Rose, S., Rehkopf, D., Phillips, R., Peterson, L., Marino, M., Maier, J., Lin, S. et al. (2022). Advancing primary care with Artificial Intelligence and Machine Learning. *Healthc (Amst), 10,* 100594.
49. Zywiolek, J., Sarkar, A., Sial, M.S. (2022). *Biometrics as a method of employee control.* 16th International Conference on Ubiquitous Information Management and Communication (IMCOM). IEEE, pp. 1-5.
50. Żywiołek, J., Rosak-Szyrocka, J., Jereb, B. (2021). Barriers to Knowledge Sharing in the Field of Information Security. *Management Systems in Production Engineering*, *29*, 114-119.
51. Żywiołek, J., Rosak-Szyrocka, J., Nayyar, A., Naved, M. (2024) *Modern Technologies and Tools Supporting the Development of Industry 5.0.* New York: CRC Press.
52. Żywiołek, J., Schiavone, F. (2021) Perception of the Quality of Smart City Solutions as a Sense of Residents' Safety. *Energies, 14,* 5511.
53. Żywiołek, J., Tucmeanu, E.R., Tucmeanu, A.I., Isac, N., Yousaf, Z. (2022). Nexus of Transformational Leadership, Employee Adaptiveness, Knowledge Sharing, and Employee Creativity. *Sustainability, 14*, 11607.