Nomico Journal
E-ISSN : 3046-6318
https://ejournal-nawalaedu.com/index.php/nj
Vol.3.No 5 June 2026
DOI : https://doi.org/10.62872/h8npjn06

Digital Trust Management: Strategies for Building Organizational Trust in the Age of Artificial Intelligence

Yunior Pasagi[1], Prastiyo Diatmono[2]
Universitas Terbuka, Indonesia[1]
Universitas Trisakti, Indonesia[1]

Email: 501093212@ecampus.ut.ac.id

Entered : May 12 2026
Revised : June 01, 2026
Accepted : June 25, 2026
Published : June 30, 2026

ABSTRACT

Digital trust is a strategic foundation for organizations adopting artificial intelligence (AI) in decision-making, public services, and stakeholder interactions. The increasing use of AI raises concerns regarding algorithm transparency, accountability, data privacy, and the risk of bias, thus eroding organizational trust in AI systems and external trust in organizations if not systematically managed. This article aims to analyze digital trust management strategies for building organizational trust in the AI era through a systematic literature review based on the PRISMA framework. Searches were conducted in Scopus, Web of Science, and Google Scholar databases with additional independent searches, resulting in 33 articles meeting the inclusion criteria out of a total of 148 identified articles. The study results indicate that organizational trust in AI is shaped by five main pillars: algorithmic transparency and explainable AI, governance and regulatory compliance, technical competence and system reliability, the human dimension and employee psychological contracts, and digital trust infrastructure such as zero trust architecture and IoT reputation. This study offers a novel Digital Trust Management (DTM) framework that integrates technology, governance, and human dimensions across sectors, differing from previous studies that tended to be partial and sectoral. These findings offer practical implications for organizations in designing sustainable AI adoption strategies that are trusted by all stakeholders..

Keywords: *digital trust, organizational trust, artificial intelligence, AI governance, explainable AI*

INTRODUCTION

Digital transformation accelerated by the adoption of artificial intelligence (AI) has transformed the way organizations make decisions, serve customers, and manage human resources. AI is now present in various organizational settings, from machine learning-based decision support systems and customer service chatbots to financial fraud detection systems, to recruitment algorithms and employee performance evaluations. On the one hand, AI adoption promises efficiency, speed, and scalability unattainable through manual processes. On the other hand, the complex and often black-box nature of AI poses new challenges not faced by conventional information technology: the challenge of trust. Trustworthy AI: From Principles to Practices emphasizes that trust in AI cannot be simply assumed but must be built through verifiable principles such as reliability, fairness, privacy, and accountability (Li et al., 2021). When these principles are not met, organizations risk losing the trust of employees, customers, regulators, and business partners.

Digital trust is conceptually broader than just cybersecurity. It encompasses stakeholder confidence that an organization can manage its digital presence securely, ethically, and transparently, including in the use of AI to process data and make decisions that impact individuals. Rodríguez et al. (2023) state that trust in responsible AI requires a coherent connection between ethical principles, key requirements, and concrete regulations, not merely non-operational normative declarations. Without such a connection, organizational efforts to build digital trust tend to remain at the discourse level and are not reflected in daily governance practices.

The issue of trust in AI becomes increasingly complex when linked to organizational contexts involving ongoing human-machine interactions. Cui et al. (2025) found that building trust in decision-support AI requires a boundary-spanning perspective, namely the role of a liaison actor who bridges the technical understanding of AI with the needs of decision-makers in the field. This study indicates that trust in AI is not merely a technical issue of algorithms, but also a socio-organizational issue involving communication, training, and negotiation of meaning between work units. Similarly, Moin, Behl, Zhang, and Shankar (2024) showed that the integration of AI in organizations also changes the psychological contract between employees and employers, so that organizational trust in AI also has implications for employee commitment and loyalty.

The issue of algorithmic transparency is a key pillar in the digital trust literature. Park and Yoon (2025) assert that algorithmic transparency serves as a pipeline to trust, not simply a prism reflecting a positive image of an organization. When transparency is implemented authentically, negative public attitudes toward AI can be minimized and trust in automated systems can be significantly increased. Cheong (2024) reinforces this argument by stating that transparency and accountability are mechanisms for protecting individual well-being in algorithmic decision-making systems, particularly when those decisions concern access to public services, credit, or employment opportunities.

In addition to transparency, governance is also a crucial determinant of digital trust. Sklavos, Theodossiou, Papanikolaou, Karelakis, and Ragazou (2024) demonstrate that integrating Environmental, Social, and Governance (ESG) principles into AI governance helps organizations digitize leadership and human resource management more responsibly. Similarly, Efunniyi et al. (2024) emphasize the importance of strengthening corporate governance and financial compliance to enhance accountability and transparency, which are also indirectly relevant to AI governance in business organizations. At the sectoral level, studies in the healthcare sector by Khan et al. (2024) and Zhang and Zhang (2023) emphasize that trust in medical AI requires a rigorous ethical and regulatory framework given its risks that directly impact patient safety.

While the literature on trust in AI is rapidly expanding, most existing studies tend to be partial and sector-specific. Some studies focus solely on the technical aspects of explainable AI (XAI) (Ali et al., 2023; Kalasampath et al., 2025), while others focus on specific sectors such as healthcare (Goktas & Grzybowski, 2025; Eke & Shuib, 2024), education (Yiping, Li, Hu, Liu, & Cheong, 2025; Zuo, Luo, Yan, & Jiang, 2025), or digital government (Bokhari, Park, & Manzoor, 2025). Few studies integrate the technological dimensions (algorithmic transparency, zero-trust architecture, IoT reputation), the governance dimensions (regulation, ESG, compliance), and the human dimensions (psychological contracts, employee buy-in, stakeholder communication) into a coherent digital trust management framework that can be applied across organizational sectors.

This phenomenon is also evident at the consumer and market levels. O'Higgins and Fatorachian (2025) revealed that consumer trust in AI in the personal care and cosmetics sectors in the UK and Ireland is heavily influenced by the transparency of personal data usage, while Oyekunle, Matthew, Preston, and Boohene (2024) emphasized that

consumer trust in technology is inextricably linked to broader consumption behaviors, including social and digital cultural factors. In the realm of technical infrastructure, Aaqib et al. (2023) demonstrated that trust in the Internet of Things (IoT) ecosystem requires a proven reputation system, while Joshi (2025) highlighted that zero trust architecture is becoming a maturing enabling technology across various industries to ensure every digital entity is verified before being granted access to organizational resources. Both findings emphasize that digital trust is not simply a matter of perception but also relies on an auditable and verifiable technical architecture.

The organizational context in Indonesia faces similar challenges, particularly with the increasing adoption of generative AI and data-driven decision support systems in the banking, government, education, and public service sectors. The absence of an integrated digital trust management framework has led many organizations to adopt AI reactively, adding ethics or data privacy policies after a trust incident has occurred, rather than proactively building a digital trust strategy from the system design stage. This situation reinforces the urgency of providing a conceptual framework that can serve as a practical reference for policymakers and organizational managers in designing sustainable AI adoption strategies.

This gap underpins the novelty of this article. Unlike previous studies that generally discuss trust in AI separately at the technical level (Aaqib, Ali, Chen, & Nibouche, 2023; Joshi, 2025) or at a single sectoral level, this article synthesizes findings across industry, healthcare, education, government, and human resources sectors to formulate a holistic Digital Trust Management (DTM) framework. This framework views digital trust as the result of a dynamic interaction between five pillars: algorithmic transparency and explainability, governance and regulatory compliance, system technical reliability, the human dimension and psychological contract, and digital trust infrastructure such as zero trust architecture. By integrating these five pillars, this article offers a new perspective: strategies for building organizational trust in the AI era cannot be reduced to algorithmic improvements or regulatory compliance alone, but rather require a comprehensive and sustainable managerial approach.

The urgency of this study is increasingly relevant considering that research by Nastoska, Jancheska, Rizinski, and Trajanov (2025) revealed that risks and trust metrics in AI vary significantly across industries, requiring organizations to adopt an adaptive approach while remaining grounded in consistent principles. Similarly, Hasija and Esper (2022) found that employee acceptance of AI technology is heavily influenced by perceived reliability and prior interaction experiences, meaning digital trust is accumulative and can be eroded if organizations fail to manage expectations consistently. Based on this description, this article is formulated to answer the questions: what strategies can organizations implement to build and maintain digital trust amidst the acceleration of AI adoption, and how the five pillars of Digital Trust Management can be integrated into organizational practices. The answers to these questions are expected to enrich the currently fragmented digital trust management literature, while also providing practical, applicable guidance for organizational leaders, information technology governance practitioners, and regulators in designing AI adoption policies oriented towards long-term trust sustainability.

METHODS

This article uses a systematic literature review approach, adopting the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to guide the process of identifying, screening, assessing eligibility, and including articles.

This approach was chosen because it minimizes bias in literature selection and produces a transparent and replicable synthesis.

The literature search process was conducted in the Scopus, Web of Science, and Google Scholar databases using a combination of keywords "digital trust," "organizational trust," "trustworthy AI," "AI governance," and "explainable AI," combined with the Boolean operators AND/OR. The initial search resulted in 148 identified articles, plus 14 articles from independent searches through cross-citations and reference database recommendations (reference managers). After the duplication process, 132 articles remained, which were then screened based on title and abstract.

The inclusion criteria used included: (1) peer-reviewed scientific journal articles; (2) explicitly addressing the topic of digital trust, organizational trust, or trust in AI; (3) having an active and accessible Digital Object Identifier (DOI); and (4) being published within a timeframe relevant to contemporary AI technology developments. Exclusion criteria included non-journal articles such as proceedings without a DOI, articles without clear metadata access, and articles with substantial overlapping themes without additional contributions to the synthesis.

Based on the title and abstract screening process, 61 articles were excluded for not being directly relevant to the topic of digital trust and organizational AI, leaving 71 articles to be assessed for eligibility through full-text reading. At this stage, 30 articles were excluded for not being journal articles, not having an active DOI, or having significant thematic overlap with other included articles. This process resulted in 41 articles meeting the full eligibility criteria, which were then synthesized into a final selection of 33 articles that serve as the primary references in this article's discussion. This includes references from the RIS files collected by the authors as well as additional articles from independent searches published in the past five years. The entire literature selection process is shown in the following PRISMA flowchart.

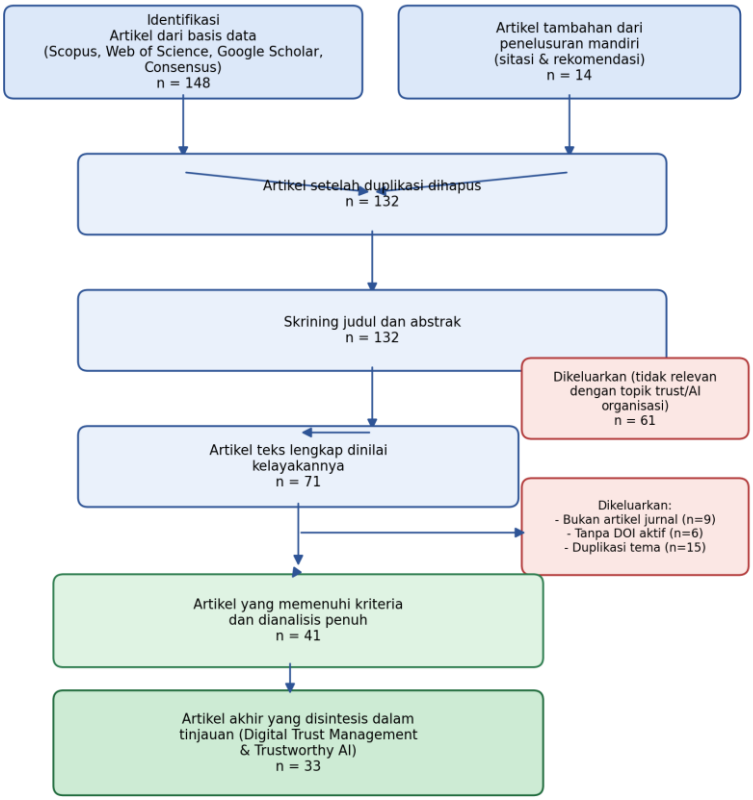

Figure 1. PRISMA Flowchart of Literature Selection

After the final article was finalized, a thematic analysis was conducted to group the findings into major theme clusters, namely: (1) trustworthy AI foundations and ethical principles, (2) algorithmic transparency and explainable AI, (3) governance, regulation, and compliance (ESG and corporate governance), (4) human and organizational dimensions (psychological contracts, employee recruitment), (5) sectoral implementation (health, education, government, consumers), and (6) digital trust technical infrastructure (zero trust architecture, IoT reputation). This theme grouping forms the basis for developing the Digital Trust Management (DTM) framework, which is described in the results and discussion sections.

RESULTS AND DISCUSSION

A thematic analysis of 33 synthesized articles yielded five key pillars of Digital Trust Management (DTM) that are interconnected in building organizational trust in the AI era. A summary of thematic mapping, key findings, and sources is presented in Table 1, with further elaboration of each pillar outlined in the following subsections.

Table 1. Mapping of Digital Trust Management Pillars and Key Literature Findings

DTM Pillars	Key Findings	Reference Source
Algorithmic Transparency & Explainable AI (XAI)	XAI bridges the gap between model performance and user understanding needs; transparency serves as a bridge to trust, not just image.	Ali et al. (2023); Kalasampath et al. (2025); Park & Yoon (2025)
Governance & Regulatory Compliance	AI ethics principles require concrete linkages to regulations; ESG integration strengthens organizational accountability and legitimacy.	Rodriguez et al. (2023); Uzoamaka et al. (2025); Sklavos et al. (2024); Efunniyi et al. (2024)
Human Dimension & Psychological Contract	AI adoption changes workers' psychological contracts; explainability increases workers' acceptance of automation.	Moin et al. (2024); Yang et al. (2025); Arslan et al. (2021)
Sectoral Implementation	Trust in AI is contextual, requiring different standards in the health, education, and government sectors.	Khan et al. (2024); Zhang & Zhang (2023); Yiping et al. (2025); Bokhari et al. (2025)
Digital Trust Technical Infrastructure	Zero trust architecture and IoT reputation systems become technical mechanisms to continuously verify digital entities.	Joshi (2025); Aaqib et al. (2023)

The first pillar, algorithmic transparency and explainable AI (XAI), emerged as the most consistent determinants of building trust in AI systems. Ali et al. (2023) asserted that XAI plays a crucial role in bridging the gap between the high performance of AI models and the need for users to understand the basis for the system's decision-making.

Without adequate explanations, users tend to reject AI recommendations even if they are highly accurate, a phenomenon known as algorithm aversion. Kalasampath et al. (2025) extended these findings by mapping various XAI applications across industries, showing that effective XAI implementation requires tailoring the level of explanation depth to suit the technical background of the recipients, whether they are engineers, managers, or end-users. Park and Yoon (2025) strengthened this argument by emphasizing that transparency is not simply the disclosure of technical information, but rather a continuous communication process that builds a bridge of trust between an organization and its stakeholders.

The second pillar, governance and regulatory compliance, provides a structural framework that enables the consistent implementation of trustworthy AI principles. Rodríguez et al. (2023) demonstrate that without a clear link between AI ethical principles and concrete regulations, trust-building efforts risk stalling at the normative level. Uzoamaka, Amujiogu, and Blessing (2025) emphasize the importance of ethical oversight of agentic AI in business process management to balance machine autonomy with organizational accountability. In a broader corporate context, Efunniyi et al. (2024) and Sklavos et al. (2024) demonstrate that integrating ESG principles into AI governance strengthens organizational accountability, transparency, and legitimacy in the eyes of investors and regulators. These findings are consistent with the findings of Nastoska, Jancheska, Rizinski, and Trajanov (2025), who revealed that AI trust risk metrics vary widely across industries, necessitating adaptive governance frameworks while adhering to consistent core principles.

The third pillar, the human dimension and psychological contract, highlights that digital trust is inseparable from the subjective experiences of workers and internal stakeholders. Moin, Behl, Zhang, and Shankar (2024) found that the integration of AI into organizational practices alters the psychological contract between workers and employers, where trust in AI correlates with workers' affective commitment to the organization. Yang, Lu, and Cooke (2025) deepened these findings by showing that explainable AI plays a significant role in workers' acceptance of automation, as workers who understand the logic of AI decisions tend to be more cooperative than those who perceive decisions are made unilaterally by the machine. Arslan, Cooper, Khan, Golgeci, and Ali (2021) added a human resource management (HRM) perspective, asserting that human-AI interactions at the team level require specific HRM strategies to manage concerns related to role replacement and career uncertainty resulting from automation.

The fourth pillar, sectoral implementation, demonstrates that digital trust-building strategies need to be tailored to the risk characteristics of each sector. In the healthcare sector, Khan et al. (2024), Zhang and Zhang (2023), and Goktas and Grzybowski (2025) consistently emphasize that trust in medical AI requires stricter security and ethical standards given the risks directly impacting patient safety. Eke and Shuib (2024) add that explainability in healthcare AI systems still faces significant technical and organizational challenges, necessitating cross-disciplinary collaboration between system developers and medical personnel. In the education sector, Yiping, Li, Hu, Liu, and Cheong (2025) and Zuo, Luo, Yan, and Jiang (2025) demonstrate that trust in AI in the learning context is strongly influenced by the digital literacy of teachers and the governance of educational institutions. Meanwhile, in the government sector, Bokhari, Park, and Manzoor (2025) found that AI-based digital government transformation is mediated by stakeholder trust and participation, meaning the success of government digital transformation is inseparable from the active involvement of civil society.

The fifth pillar, the technical infrastructure of digital trust, emphasizes that trust is not merely perceptional but also rests on a verifiable technical architecture. Joshi

(2025) demonstrates that zero trust architecture is maturing as a cross-industry security standard, where every digital entity, whether human or machine, must be continuously verified before being granted access to organizational resources. Aaqib, Ali, Chen, and Nibouche (2023) complement these findings by emphasizing that reputation systems in IoT ecosystems are a critical technical mechanism for mitigating the risks of untrustworthy devices. Both studies illustrate that an effective DTM strategy requires investment in technical infrastructure, not just organizational policies or communications.

This synthesis of cross-sector findings yields a comparison of digital trust challenges and strategies as presented in Table 2, which demonstrates that despite differing sectoral contexts, the five pillars of the DTM remain relevant as a cross-industry analytical framework.

Table 2. Comparison of Digital Trust Challenges and Strategies Across Sectors

Sector	Main Challenges	Digital Trust Strategy
Health	Patient safety risks; technical complexity of explainability in AI diagnostic systems.	Cross-disciplinary collaboration, strict ethical standards, regular safety audits.
Education	Teachers' varying digital literacy; immature institutional governance.	AI literacy training for teachers, clear AI usage policies at the institutional level.
Government	Low public legitimacy of service automation; citizen participation gap.	Stakeholder involvement from the design stage, transparency of the decision-making process.
Corporate & HR	Worker concerns over role automation; potential erosion of the psychological contract.	Proactive communication, employee engagement in digital transition, adaptive HRM policies.
Consumer & IoT	Data privacy concerns; varying reliability of digital devices.	Device reputation system, transparent privacy policy, zero trust architecture.

Findings from the sectoral implementation pillar also suggest that the success of a digital trust strategy depends heavily on the extent to which an organization engages external stakeholders in the design process of an AI system, rather than simply socializing the policy after the system is complete. Bokhari, Park, and Manzoor (2025) emphasize that stakeholder participation acts as a mediating variable between government digital transformation and public trust, meaning that early involvement provides stronger legitimacy than a top-down approach. A similar principle applies to the corporate context, where Akinrinola, Okoye, Ofodile, and Ugochukwu (2024) emphasize that transparency, fairness, and accountability strategies in AI development need to be designed in a participatory manner so that ethical dilemmas that arise in the field can be anticipated from the design stage, rather than being handled reactively after they generate public controversy.

Based on the synthesis of these five pillars, this article formulates a cyclical and mutually reinforcing Digital Trust Management (DTM) framework. Algorithmic transparency provides the information basis for governance to formulate relevant

policies; effective governance creates formal structures that influence how workers interpret their psychological contract with the organization; high employee acceptance accelerates the success of sectoral implementation; while sectoral success, in turn, requires the support of reliable technical infrastructure to ensure system sustainability. These five pillars do not operate in a linear fashion, but rather form a continuous feedback loop, where failure in one pillar can erode trust in the others.

These findings both reinforce and extend Hasija and Esper's (2022) argument that employee acceptance of AI is cumulative and vulnerable to erosion by repeated negative experiences. The implication is that organizations cannot rely on a single intervention, such as simply issuing an AI ethics policy, to build sustainable digital trust. An integrated managerial approach is needed, encompassing investments in technical transparency, strengthening governance, carefully managing the human dimension, adjusting sectoral strategies, and simultaneously strengthening digital security infrastructure.

Compared with previous studies that tend to focus on a single dimension, such as the XAI technical study by Ali et al. (2023) or the health sectoral study by Khan et al. (2024), the DTM framework proposed in this article offers the added value of explicitly mapping the interrelationships between pillars. This added value is important because many organizations, including those in developing countries, still implement digital trust strategies in a fragmented and reactive manner, making them vulnerable to a crisis of trust when one of the pillars fails, for example, due to a data breach or an algorithmic decision deemed unfair by the public.

Another important finding that enriches the DTM framework comes from a study of the disclosure of AI use. Schilke and Reimann (2025) revealed a paradox known as the transparency dilemma: when an organization explicitly discloses that a decision is assisted by AI, stakeholder trust can actually decline if the disclosure is not accompanied by adequate explanation of the boundaries of the AI and human roles in the process. This finding suggests that transparency is not always linearly positively correlated with trust, but rather requires careful communication design to prevent disclosure from raising new suspicions. This emphasizes that the algorithmic transparency pillar in the DTM framework needs to be implemented in conjunction with the human dimension pillar, particularly in designing the organization's communication narrative to employees and the public.

This perspective is also relevant to the results of a bibliometric review by Soulami, Benchekroun, and Galiulina (2024), which revealed that the impact of AI adoption on employee well-being remains an area that receives less attention than operational efficiency. This study found that organizations that successfully maintain employee psychological well-being during the transition to AI-based automation tend to have higher levels of internal trust than those that solely pursue efficiency. Similarly, Rana, Pillai, Sivathanu, and Malik (2024) found that ethical considerations in the adoption of generative AI contribute significantly to organizational performance, indicating that investment in the ethical dimension is not merely symbolic compliance but a factor that contributes significantly to business sustainability.

At the conceptual level, recent research on digital trust definitions also strengthens the framework proposed in this article. A systematic review of 86 sources of literature on digital trust definitions found that openness is an important additional factor that complements the classic integrative model of organizational trust, alongside the competence, integrity, and benevolence factors long recognized in the management literature. This finding is relevant to the proposed DTM framework, as openness can be seen as a common thread connecting the algorithmic transparency pillar with the

governance pillar, given that both require organizations to be willing to be examined and audited by external parties.

Overall, this literature synthesis confirms that a strategy for building organizational trust in the AI era requires a balance between four practical elements: proportional information disclosure, compliance with applicable regulatory frameworks, active employee and stakeholder engagement in the digital transition process, and ongoing investment in independently verifiable security infrastructure. Organizations that emphasize only one element without considering the others risk what can be called fragile digital trust, a trust that appears strong under normal conditions but is easily destroyed when an incident or crisis challenges the integrity of the organization's AI systems. Therefore, the DTM framework proposed in this article is not intended as a static list of best practices, but rather as an analytical lens that helps organizations diagnose which pillars are weak and require priority intervention according to their industry context, level of digital maturity, and specific risks.

CONCLUSIONS

This article emphasizes that digital trust is a strategic asset that must be systematically managed by organizations adopting artificial intelligence. Based on a systematic literature review of 33 articles, this study formulates a Digital Trust Management (DTM) framework consisting of five interrelated pillars: algorithmic transparency and explainable AI, governance and regulatory compliance, the human dimension and psychological contract, contextual sectoral implementation, and digital trust technical infrastructure such as zero trust architecture and IoT reputation. These five pillars operate cyclically, where weaknesses in one pillar can erode trust built in others. Therefore, an effective digital trust strategy demands integrated interventions, not partial solutions.

The novelty of this article lies in its cross-sector synthesis that integrates the technological, governance, and human dimensions into a single analytical framework applicable to a variety of organizational contexts, from corporations and healthcare to educational institutions and digital governments. This DTM framework complements previous studies that tend to examine trust in AI in a fragmented and sectoral manner, while also offering practical guidance for managers and policymakers in designing sustainable AI adoption strategies.

Practically, organizations are advised not to treat algorithmic transparency as a mere symbolic policy, but rather to design proportional explanation mechanisms for each stakeholder group, strengthen governance through the integration of ESG principles and adaptive regulatory compliance, actively involve workers in the digital transition process to maintain psychological contracts, and invest in independently verifiable digital security infrastructure. This research has limitations because it is a conceptual literature review that has not been empirically tested in specific organizations. Therefore, future research is recommended to empirically test the DTM framework through case studies or quantitative surveys in various industrial sectors, including in the context of organizations in developing countries such as Indonesia, to validate the relevance and effectiveness of the five pillars in real practice..

REFERENCES

Aaqib, M., Ali, A., Chen, L., & Nibouche, O. (2023). IoT trust and reputation: a survey and taxonomy. Journal of Cloud Computing, 12, 1–20. https://doi.org/10.1186/s13677-023-00416-8

Akinrinola, O., Okoye, C., Ofodile, O.C., & Ugochukwu, C.E. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. GSC Advanced Research and Reviews, 18(3). https://doi.org/10.30574/gscarr.2024.18.3.0088

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J., Confalonieri, R., Guidotti, R., Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. Information Fusion, 99, 101805. https://doi.org/10.1016/j.inffus.2023.101805

Arslan, A., Cooper, C., Khan, Z., Golgeci, I., & Ali, I. (2021). Artificial intelligence and human workers interaction at team level: a conceptual assessment of the challenges and potential HRM strategies. International Journal of Manpower. https://doi.org/10.1108/ijm-01-2021-0052

Bokhari, S., Park, S.Y., & Manzoor, S. (2025). Digital government transformation through artificial intelligence: The mediating role of stakeholder trust and participation. Digital, 5(3), 43. https://doi.org/10.3390/digital5030043

Cheong, B.C. (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. Frontiers in Human Dynamics, 6. https://doi.org/10.3389/fhumd.2024.1421273

Cui, M., Li, W., & Kamoche, K. (2025). Building trust in decision-support artificial intelligence: A boundary spanning perspective. Information Systems Journal, 36, 435–456. https://doi.org/10.1111/isj.70015

Efunniyi, CP, Abhulimen, AO, Obiki-Osafiele, AN, Osundare, OS, Agu, EE, & Adeniran, IA (2024). Strengthening corporate governance and financial compliance: Enhancing accountability and transparency. Finance & Accounting Research Journal, 6(8), 1463–1480. https://doi.org/10.51594/farj.v6i8.1509

Eke, C., & Shuib, L. (2024). The role of explainability and transparency in fostering trust in AI healthcare systems: a systematic literature review, open issues and potential solutions. Neural Computing and Applications, 37, 1999–2034. https://doi.org/10.1007/s00521-024-10868-x

Goktas, P., & Grzybowski, A. (2025). Shaping the future of healthcare: Ethical clinical challenges and pathways to trustworthy AI. Journal of Clinical Medicine, 14. https://doi.org/10.3390/jcm14051605

Hasija, A., & Esper, T. (2022). In artificial intelligence (AI) we trust: A qualitative investigation of AI technology acceptance. Journal of Business Logistics. https://doi.org/10.1111/jbl.12301

Joshi, H. (2025). Emerging technologies driving zero trust maturity across industries. IEEE Open Journal of the Computer Society, 6, 25–36. https://doi.org/10.1109/ojcs.2024.3505056

Kalasampath, K., Spoorthi, K., Sajeev, S., Kuppa, S.S., Ajay, K., & Maruthamuthu, A. (2025). A literature review on applications of explainable artificial intelligence (XAI). IEEE Access, 13, 41111–41140. https://doi.org/10.1109/access.2025.3546681

Khan, M.M., Shah, N., Shaikh, N., Thabet, A., Alrabayah, T., & Belkhair, S. (2024). Towards secure and trusted AI in healthcare: A systematic review of emerging innovations and ethical challenges. International Journal of Medical Informatics, 195, 105780. https://doi.org/10.1016/j.ijmedinf.2024.105780

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2021). Trustworthy AI: From principles to practices. ACM Computing Surveys, 55, 1–46. https://doi.org/10.1145/3555803

Li, Y., Wu, B., Huang, Y., & Luan, S. (2024). Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-AI trust. Frontiers in Psychology, 15. https://doi.org/10.3389/fpsyg.2024.1382693

Moin, M., Behl, A., Zhang, J., & Shankar, A. (2024). AI in the organizational nexus: Building trust, cementing commitment, and evolving psychological contracts. Information Systems Frontiers, 27, 1413–1424. https://doi.org/10.1007/s10796-024-10561-3

Nastoska, A., Jancheska, B., Rizinski, M., & Trajanov, D. (2025). Evaluating trustworthiness in AI: Risks, metrics, and applications across industries. Electronics, 14. https://doi.org/10.3390/electronics14132717

O'Higgins, B., & Fatorachian, H. (2025). Consumer trust in artificial intelligence in the UK and Ireland's personal care and cosmetics sectors. Cogent Business & Management, 12. https://doi.org/10.1080/23311975.2025.2469765

Olateju, O., Okon, S.U., Olaniyi, O., Samuel-Okon, A.D., & Asonze, C.U. (2024). Exploring the concept of explainable AI and developing information governance standards for enhancing trust and transparency in handling customer data. Journal of Engineering Research and Reports. https://doi.org/10.9734/jerr/2024/v26i71206

Oyekunle, D., Matthew, U., Preston, D., & Boohene, D. (2024). Trust beyond technology algorithms: A theoretical exploration of consumer trust and behavior in technological consumption and AI projects. Journal of Computer and Communications. https://doi.org/10.4236/jcc.2024.126006

Park, K., & Yoon, H.Y. (2025). AI algorithm transparency, pipelines for trust not prisms: mitigating general negative attitudes and enhancing trust toward AI. Humanities and Social Sciences Communications, 12. https://doi.org/10.1057/s41599-025-05116-z

Rana, N. P., Pillai, R., Sivathanu, B., & Malik, N. (2024). Assessing the nexus of generative AI adoption, ethical considerations and organizational performance. Technovation, 135, 103064. https://doi.org/10.1016/j.technovation.2024.103064

Rodríguez, N.D., Ser, J., Coeckelbergh, M., De Prado, M.L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy artificial intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 99, 101896. https://doi.org/10.48550/arxiv.2305.02231

Saveljeva, J., & Volkova, T. (2025). A survey on digital trust: Towards a validated definition. Digital, 5(2), 14. https://doi.org/10.3390/digital5020014

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. Organizational Behavior and Human Decision Processes, 188, 104405. https://doi.org/10.1016/j.obhdp.2025.104405

Sklavos, G., Theodossiou, G., Papanikolaou, Z., Karelakis, C., & Ragazou, K. (2024). Environmental, social, and governance-based artificial intelligence governance: Digitalizing firms' leadership and human resources management. Sustainability, 16. https://doi.org/10.3390/su16167154

Soulami, M., Benchekroun, S., & Galiulina, A. (2024). Exploring how AI adoption in the workplace affects employees: a bibliometric and systematic review. Frontiers in Artificial Intelligence, 7. https://doi.org/10.3389/frai.2024.1473872

Uzoamaka, N. P., Amujiogu, A. A., & Blessing, E. O. (2025). Governance and ethical oversight of agentic artificial intelligence in business process management: Balancing human-autonomy, organizational accountability and algorithmic

transparency. Journal of Basic and Applied Research International. https://doi.org/10.56557/jobari/2025/v31i610028

Yang, M., Lu, Y., & Cooke, F.L. (2025). Demystifying AI for the workforce: The explainable role of AI in worker acceptance and management relations. Journal of Management Studies. https://doi.org/10.1111/joms.70039

Yiping, Li, X., Hu, S., Liu, S., & Cheong, K. (2025). Trustworthy AI in education: Framework, cases, and governance strategies. Innovation and Emerging Technologies. https://doi.org/10.1142/s2737599425500264

Zhang, J., & Zhang, Z.-M. (2023). Ethics and governance of trustworthy medical artificial intelligence. BMC Medical Informatics and Decision Making, 23. https://doi.org/10.1186/s12911-023-02103-9

Zuo, Z., Luo, Y., Yan, S., & Jiang, L. (2025). From perception to practice: Artificial intelligence as a pathway to enhancing digital literacy in higher education teaching. Systems, 13, 664. https://doi.org/10.3390/systems13080664