SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2025
ORGANIZATION AND MANAGEMENT SERIES NO. 239

TRUST AND RESPONSIBILITY IN AI-DRIVEN LEADERSHIP: ACCOUNTABILITY FOR AI DECISIONS

Violetta ARUTIUNIAN

Wroclaw University of Economics and Business; violetta.arutiunian@ue.wroc.pl,
ORCID: 0009-0001-9950-3331

Purpose: The purpose of this paper is to explore the multifaceted relationship between Artificial Intelligence and organizational leadership, specifically addressing the "responsibility gap" that arises when decision-making is delegated to algorithms. It aims to redefine leadership roles in the "algocratic" era by distinguishing between moral trust and technical reliance. Furthermore, the research seeks to provide a comprehensive governance framework to help leaders navigate the ethical challenges of shared responsibility, ensuring that human judgment remains central to strategic decisions.
Methodology: This study adopts an integrative tertiary review approach to synthesize insights across high-stakes domains including healthcare, the judiciary, and management. By analysing the intersection of technical explainability (XAI) and ethical governance, the paper adapts the TRUST framework (Transparency, Responsibility, Understanding, Social Justice, Trustworthy Oversight) originally proposed in legal scholarship to the context of organizational leadership. The approach moves beyond descriptive analysis to offer a critical synthesis of how leaders can maintain accountability in Human-Machine Teaming.
Findings: The research identifies a critical distinction between "trust" (a moral relationship requiring goodwill, applicable to humans) and "reliance" (based on technical predictability, applicable to AI). The study finds that while AI enhances decision-making efficiency and analytics, it introduces a "responsibility gap" where assigning blame for errors becomes difficult. Furthermore, a paradox is identified where users report higher trust in humans but are behaviorally more willing to disclose sensitive information to AI systems due to lower social evaluation anxiety. The paper concludes that effective leadership requires "Meaningful Human Oversight" where humans retain the authority to override AI, rather than acting as passive validators.
Research limitations/implications: As a conceptual paper based on an integrative review, this study lacks primary longitudinal data regarding the long-term implementation of the TRUST framework in corporate settings. Future research should empirically test the impact of distinguishing "trust" from "reliance" on employee satisfaction and error rates in hybrid teams. Additionally, further investigation is needed into the long-term cognitive effects of "automation bias" on the degradation of human leadership skills.
Practical implications: The paper offers a strategic implementation framework for leaders. Organizations must prioritize AI literacy to ensure leaders can critically evaluate algorithmic outputs rather than blindly accepting them. Leaders are advised to implement specific protocols for Meaningful Human Control, ensuring that AI serves as a decision-support tool rather than

a decision-maker in high-stakes scenarios. Furthermore, the adoption of the TRUST framework provides a roadmap for operationalizing ethics, moving from abstract principles to concrete actions such as algorithmic auditing and equity impact assessments.
Social implications: This research addresses the broader societal risk of algocracy, where opaque systems erode the legitimacy of decision-making in public and private institutions. By emphasizing Social Justice, the paper highlights the necessity of preventing AI from perpetuating historical biases found in training data against marginalized communities. Promoting a culture of "trustworthy oversight" ensures that technological advancement does not come at the cost of human dignity, employee well-being, or democratic values.
Value: This paper provides a novel theoretical contribution by bridging the gap between technical definitions of reliability and philosophical concepts of trust within a leadership context. It adapts the judicial TRUST framework for broader management application, offering a structured solution to the "responsibility gap". The value lies in its actionable guidance for leaders to navigate the paradox of "algorithmic appreciation" versus "automation bias" ensuring AI is used to augment rather than hollow out human agency.
Keywords: AI leadership, accountability, organizational trust, ethical decision-making, transparency.
Category of the paper: Conceptual paper / Literature review.

1. Introduction

Today's organizations and public institutions are facing a transformation that is referred to as the fourth industrial revolution. A central element of this change is the ubiquity of algorithmic decision systems, which have ceased to be just analytical tools and have become active participants in decision-making processes (European Parliament, 2019). This fact is no longer controversial: the universality and increasing autonomy of algorithms are well documented.

However, these facts lead to a less obvious research question: *what forms of coexistence between humans and AI are possible, and how should responsibility for decisions made with the participation of intelligent systems be understood?* Trust in algorithmic systems and the question of accountability are becoming central concerns: Who is responsible for decisions supported or made by AI? Since 2020, there has been a growing interest in the management and quality literature on the issue of shared responsibility and the redefinition of leadership roles in an environment in which some analytical or decision-making functions are delegated to algorithmic systems. From automating routine tasks to augmenting complex decision-making processes, AI's influence is reshaping how organizations operate, strategize, and interact with their stakeholders (Joshi, 2025, p. 49).

In traditional models, the responsibility rested with the man-leader, who made decisions based on available information and his own competences. In the algocratic era, where the decision-making process is shared with autonomous or semi-autonomous systems, the boundaries of responsibility are blurred. As Madanchian et al. (2024) note, the integration

of AI in management is not just a technological issue, but a fundamental cultural change, requiring new adaptive competencies from leaders. However, with the increasing autonomy of AI systems, there is a pressing problem of the responsibility gap, in which it is difficult to assign blame to a specific entity (designer, user, or algorithm itself).

The article reviews the conceptual literature to explore the multifaceted relationship between AI and organizational dynamics. We delve into the impact of AI on key areas such as management and leadership, decision-making methodologies, stakeholder relations, organizational change management, cross-functional collaboration, and conflict resolution.

The purpose of this article was to provide a comprehensive analysis. Based on a literature review, the article:

- It identifies the key factors influencing trust in AI.
- It examines the ethical and societal challenges of dispersion of responsibility in AI-based management.

2. Methodology

This study utilizes an integrative review approach to synthesize theoretical and empirical literature, facilitating a comprehensive understanding of AI-driven leadership. Given the rapid pace of technological change, a transdisciplinary lens was adopted, integrating insights from management, law, and bioethics.

The aim of this approach is to identify universal mechanisms for building trust and accountability that go beyond the specifics of individual sectors.

2.1. Selection criteria

The literature selection process was not limited to a single database or a narrow definition of management, but deliberately included "high-stakes domains" where AI decisions have a direct impact on human well-being or social justice. To be included in the analysis, an article must meet the following inclusion criteria:

1. *Intersection of technology and ethics:* the works analyzed had to address the point of contact between the technical aspects of AI systems and social or ethical oversight. In particular, studies combining technical explainability with ethical governance were sought.
    - As an example of literature focusing on the technical and business dimensions, works such as Behera et al. (2023) are included, which analyze the impact of XAI on sustainable growth and competitive advantage.

- Critical analyses such as the work of Lima et al. (2022), which examines the conflict between the explainability of algorithms and their legal and moral accountability, have been included as a counterpoint.

2. *Cross-domain synthesis:* because leadership challenges in the algorithmic age are often most evident in highly regulated sectors, the review included literature on healthcare, the judiciary, and operational management.
   - Analysis of the judicial sector (e.g. Flórez Rojas, 2025; Gravett, 2024) allowed to identify risks related to algocracy and automation bias.
   - The medical literature (e.g. Mendel et al., 2024) has provided evidence of paradoxical phenomena of patients' trust in AI compared to doctors.
   - This approach justifies the thesis that the dilemmas faced by judges and doctors are an early warning and a model for business leaders in other sectors.
3. *Timeliness and scientific rigor:* peer-reviewed articles published between 2019 and 2025 were prioritized to capture the *dynamics* of change related to the implementation of generative AI and the latest legal regulations.

2.2. Analytical procedure

The collected material was subjected to thematic analysis to identify the key dimensions of trust and reliance. The synthesis process consisted in mapping the control mechanisms and risks identified in the literature to the conceptual framework of leadership. In particular, the TRUST framework identified in the legal literature has been adapted and operationalized for the needs of management practice.

This methodological approach allowed us to go beyond a simple description of AI applications, allowing for a critical assessment of the responsibility gap and formulation of recommendations for leaders who need to manage not only technology, but above all relationships in hybrid human-machine teams.

3. AI and decision process transformation

3.1. AI as a decision catalyst

Algorithmic decision-making based on big data has become an essential tool and is pervasive in all aspects of our daily lives: the news articles we read, the movies we watch, the people we spend time with, whether we get searched in an airport security line, whether more police officers are deployed in our neighborhoods, and whether we are eligible for credit, healthcare, housing, education and employment opportunities, among a litany of other commercial and government decisions (Gravett, 2024, p. 282).

AI, especially when combined with big data analytics, is seen as a technology capable of revolutionizing decision-making, providing real-time recommendations and enabling faster response to crises. Scientific research proves that AI plays a significant role in the decision-making process. For example, in a study of 133 employees of Saudi organizations, the results clearly showed that AI has a direct positive impact on improved decision-making. Moreover, improved decision-making has a significant positive impact on individual productivity, organizational performance and organizational culture (Alasmri and Basahel, 2022, p. 1).

AI supports leaders in:

- Analytics and prediction: analyze historical data, forecast future trends (e.g., sales, customer behavior), and detect anomalies that signal upcoming issues early.
- Optimization and automation: it helps automate routine activities and processes, which frees up leaders and teams to focus on strategic tasks. In the IT sector, almost 40% of automation processes concern this department (Trojańczyk, 2024, p. 100).
- Reducing decision-making stress: faced with an increasing number of decisions (a tenfold increase in the last three years) and decision-making stress (85% of leaders), AI can provide recommendations, supporting the selection process under conditions of uncertainty (Trojańczyk, 2024, p. 100).

However, it should be remembered that AI is only a tool, and its analyses, although advanced, may be subject to error. Final, strategic decisions must be made and accepted by humans.

3.2. Epistemic trust and explainability

Trust is a key element in interpersonal relationships and in the functioning of the organization, and in the case of AI implementation, it becomes the foundation for its acceptance. The problem of trust has become especially relevant in connection with the globalization of threats caused by accelerated technological development (Sychev, 2024, p. 39). In the technological context, trust is closely linked to the predictability and reliability of systems.

A key problem in delegating decisions to AI is the so-called black box problem, when even the creators of the algorithm are not able to fully explain how the system arrived at a given result. Behera and co-authors (2023) note that while AI models learning from massive data sets have powerful predictive power, their internal processes are often incomprehensible to humans. Without understanding how the algorithm arrived at a specific recommendation – such as refusing a loan or rejecting a job candidate – leaders cannot take full responsibility for those decisions. As González Arencibia et al. (2024) note, the lack of explainability undermines transparency and trust, which are essential for ethical leadership. This creates a tension between effectiveness and explainability. This opacity undermines trust, especially in autonomous and high-stakes systems (Busuioc, 2021, p. 825).

The drive to build trust is one of the main motivators for the development of Explainable Artificial Intelligence. XAI aims to develop methods and tools that enable humans to understand, rationally explain, and trust the results generated by AI (Behera et al., 2023, p. 1).

Explanations are often not reliable, and can be misleading. If we instead use models that are inherently interpretable, they provide their own explanations, which are faithful to what the model actually computes. The lack of explainability limits the understanding of how AI comes to conclusions, leading to user frustration and hindering the adoption of technology in critical sectors such as medicine and the judiciary.

Empirical research shows that transparency is key to building trust: employees in companies that focus on a high level of communication transparency feel trustworthy 1.6 times more often, are 1.8 times more productive, and 2.3 times more focused on tasks compared to companies with low transparency. Transparency of decisions and processes is therefore not only a matter of ethics, but a factor that directly affects the performance of an organization (Trojańczyk, 2024, p. 38).

In scientific discourse, the concepts of "trust" and "reliance" are increasingly distinguished in the context of interaction with technology. Bruynseels, Asveld, and van den Hoven (2025) argue that assigning "trust" to algorithms is a categorical fallacy because machines lack will or intentionality. Instead, they propose the concept of "epistemic reliance". Allocation of trust should be reserved for human agents and the organizations they operate in. Reliance applies to foundation models and artificial intelligence agents (Bruynseels et al., 2025).

This distinction is crucial for accountability. If we treat AI as a partner that we "trust" in a moral sense, we risk abdicating our own responsibility to verify its actions. Research by Hallowell et al. (2022) on the use of AI in the diagnosis of rare diseases confirms that relational trust (to the human using the tool) is crucial for clinicians and patients, and evidence-based reliability (epistemic trust) is expected in relation to the tool itself.

Research by Mendel, Nova, and Wiesenfeld (2024) reveals an interesting paradox in the context of remote patient monitoring. Although patients declare a higher level of trust in doctors than in AI, in practice they are more likely to disclose sensitive health information to AI systems when the level of trust is already established. This is due to the fact that users disclose health information more readily to AI than to doctors (Mendel et al., 2024), which may suggest less fear of social evaluation by the machine.

However, as Maris et al. (2024) point out, patients still expect a right to a human doctor. They fear that the introduction of AI into clinical practice will lead to the loss of the "human touch" that is essential for empathetic care and the inclusion of individual patient values in the decision-making process.

3.3. Traditional trust models and autonomous systems

Traditionally, leadership has been based on human judgment, intuition, and experience. The development of AI technologies, in particular machine learning and large language models, has changed the landscape of decision-making. In 2025, AI systems are increasingly being used not only to support managerial decisions, but also to automate routine and strategic choices in sectors ranging from finance and healthcare to public administration. Traditional leadership models are evolving towards hybrid systems, in which leaders collaborate with AI as part of *Human-Machine Teaming (HMT*). Assaad (2024) defines this dynamic as a transition from a hierarchical to a partnership relationship, where in a teaming operation, the human relies on the roles and responsibilities of the machine to fulfil their own roles and responsibilities, and vice versa. However, this interdependence increases the complexity of the system and introduces new risks, in particular related to the unpredictability of autonomous systems.

In the context of crisis management, as Trojańczyk (2024) points out, the role of the leader changes from an authoritarian decision-maker to an interpreter of data provided by AI. Leaders must have the ability to critically evaluate algorithmic recommendations so as not to fall into the trap of overtrusting technology. Joshi (2025) highlights that in matrix organizations, AI can increase decision-making efficiency by 23% and conflict resolution efficiency by 37%, however, this requires strategic leadership capable of managing change and mitigating employee resistance.

3.4. AI as a co-decision-maker

In more recent years, the information system field is beginning to place its attention on complex and often alarming ways in which the use of IT affects organisational and social life – the dark side of technology and its use. (Mikalef et al., 2022, p. 257). For instance, in gig-economy companies, such as Task Rabbit or Uber, algorithms can independently "deactivate", thus sanction a worker's account based on an opaque algorithmic decision. Other companies apply integrated and AI-driven performance management platforms that optimize working times, track work procedures, and rate employee productivity in a short-paced manner. (Schafheitle et al., 2021, p. 5473).

The use of AI in the judiciary raises the spectre of "algocracy" – a system in which algorithms take over the decision-making functions of public authority (Flórez Rojas, 2025). An example is the COMPAS system used in the USA to assess the risk of recidivism, which became the subject of the famous State versus Loomis case. Gravett (2024) analyses this phenomenon through the lens of the "technology effect" and "automation bias".

Judges, under time and cognitive pressure, may be tempted to uncritically accept algorithms' recommendations. Most people will take the road of least cognitive effort, and rather than systematically analyze each decision, will use decision rules of thumb or heuristics. Automated decision aids may act as one of these decision-making heuristics (Gravett, 2024).

This leads to a situation in which the algorithm becomes the de facto final decision-maker, even though it formally only has an advisory function. Flórez Rojas (2025) proposes the TRUST framework – Transparency, Responsibility, Understanding, Social justice, Trustworthy oversight, as an essential safeguard against the erosion of the legitimacy of the justice system in the face of automation. They do not replace traditional legal standards, but complement them by emphasizing the relational and procedural dimensions of trust. Below is a table presenting the TRUST model.

Table 1.
The TRUST Framework for AI Governance

Principle	Description & Objective	Managerial Application	Key Mechanisms & Actions
T – Transparency	Requires openness regarding how AI systems operate and how decisions are reached. It ensures that algorithmic processes are not opaque "black boxes" but are accessible for scrutiny.	• Implementing mandatory "explainability reports" for AI-driven strategic decisions. • Ensuring employees understand why an AI recommended a specific resource allocation.	• Publishing algorithmic transparency reports. • Providing accessible, comprehensible explanations of algorithmic decisions. • Ensuring decisions are open to public and professional feedback.
R – Responsibility	Establishes clear lines of accountability for both the design and deployment of AI systems. It ensures that human judgment is not displaced and that judicial actors remain accountable for decisions mediated by algorithms.	• Defining specific protocols for "Meaningful Human Oversight". • Leaders must sign off on AI decisions, retaining liability for errors.	• Maintaining meaningful human oversight (human-in-the-loop). • Defining liability for system malfunctions or errors. • Ensuring algorithms support rather than supplant human deliberation.
U – Understanding	Focuses on education and AI literacy for legal professionals and the public. Judges and court administrators must comprehend how tools operate to critically evaluate outputs and recognise potential biases.	• Mandating AI literacy training for managers to recognize "hallucinations" or bias. • Counteracting automation bias through critical evaluation workshop.	• Training programs for judges on AI capabilities and limitations. • Bridging the gap between technological advancements and judicial expertise. • Promoting critical evaluation of generative AI outputs.
S – Social Justice	Aims to prevent AI from perpetuating systemic inequalities or discrimination. It ensures that the integration of AI aligns with fundamental rights and principles of fairness.	• Conducting regular algorithmic equity audits to ensure AI tools in HR (hiring/firing) do not discriminate against marginalized groups.	• Auditing systems for bias against marginalised communities. • Ensuring equal access to justice. • Aligning technological functionality with procedural safeguards to protect human rights.
T – Trustworthy Oversight	Involves institutional mechanisms to ensure algorithmic governance remains accountable. It requires independent bodies to monitor compliance with legal standards.	• Establish an internal "AI Ethics Board" or independent audit committee to review high-stakes algorithmic deployments.	• Establishing independent bodies empowered to audit AI systems. • Conducting post-implementation reviews. • Intervening in cases of system abuse or malfunction.

Source: own study based on Flórez Rojas, 2025.

The above model is proposed as a normative guide to mitigate the risks associated with algocracy, a situation in which algorithms shape decision-making processes in a way that limits human participation and the ability to challenge decisions.

4. The paradox of accountability for AI-supported decisions

The study, titled *Who is Responsible, the Advisor or the AI? Understanding the Effects of Advisors Disclosing Their AI Use on Their Perceived Responsibility and AI Reliance* by Tamir Mednel, Soumik Mandal, Oded Nov, and Batia Mishan Wiesenfeld of New York University looked at the impact of disclosing AI use. The aim of the study was to examine how advisors' disclosure of the use of AI systems affects their perceived accountability for decisions and the degree of reliance on AI advice.

Financial advisors took part in the experiment and participated in a simulation of financial advice. At the same time, the condition of disclosing to clients whether the advisor used AI disclosed vs. undisclosed was manipulated. The results showed that advisors felt less accountable for their recommendations when their use of AI was disclosed to clients, compared to when it remained undisclosed. In addition, a higher sense of personal responsibility correlated with less reliance on AI. The perceived self-competence of advisors increased their sense of responsibility towards investment decisions compared to AI, while greater trust in AI decreased their sense of responsibility (Mendel et al., 2025, p. 8).

One hypothesis is that advisors' perceived self-responsibility is negatively related to relying on AI advice – in other words, advisors who consider themselves more responsible will rely less on AI, while those who feel less responsible will trust AI recommendations more.

"In combination, the above predictions suggest that disclosure of AI use will lower advisor's perception of their responsibility, which will in turn increase their reliance on AI. Thus, disclosure of AI use increases reliance on AI by reducing advisor's perceived self-responsibility" (Mendel et al., 2025, p. 8). The authors also examine whether a sense of personal responsibility mediates the positive relationship between disclosing the use of AI and the degree of reliance on its recommendations.

In their study, the authors also referred to Agency Theory, which is a theoretical framework for analyzing the conflicts and challenges that arise when an "agent" (in the context of this research, an advisor) acts on behalf of another person (e.g., a client). Agency theory emphasizes that the interests of the advisor and the client may not be fully aligned, and the effects of the discrepancy are referred to as "agency problems". Disclosure of an advisor's use of AI can reduce the asymmetry of information between advisor and client, increasing process transparency and making it easier for clients to evaluate the advisor. In agency theory, disclosure should motivate the advisor to act in the interest of the client, however, in practice,

it can also lower the advisor's sense of personal responsibility, leading to greater reliance on AI and potentially more biased recommendations.

Such a perception of responsibility is not limited to financial advice. It can translate into other areas of human activity supported by AI, where delegating decisions or knowledge to automated systems creates similar ethical and practical dilemmas. Artificial intelligence (AI)-based decision support systems are used extensively in various domains, such as accounting, law, health, and finance (Mendel et al., 2025, p. 2).

One of the most striking innovations in the United States criminal justice system in the course of the last three decades has been the introduction of risk-assessment software, powered by sophisticated and often proprietary algorithms, to predict whether individual offenders are likely to reoffend (the so-called "risk of recidivism") (Gravett, 2024, p. 282).

In medicine, for example, in radiology an AI system may be used to identify anomalous patterns in image data and In medicine suggest a diagnosis, whose correctness a radiologist verifies before sharing it with the patient (Mendel et al., 2025, p. 2).

In the military sector, AI is used in almost all areas of activity, including military intelligence. "The primary role of military intelligence is to gather and analyze information to help military leaders make informed decisions." (Nitzl et al., 2025, p. 1). "Military intelligence is thus concerned with the collection and analysis of information to provide a comprehensive picture of the situation. This may entail the collection of data on the armed forces and the examination of the plans and operations of other nations, as well as the gathering of information on developments affecting a nation's security (Sadiku, Musa, 2021; Nitzl et al., 2025, p. 1). New developments in AI and its integration as an analytical tool promise a wide range of support capabilities for the analyst (Nitzl et al., 2025, p. 2).

"Researchers have identified a tendency towards excessive optimism when making decisions involving technology. This bias towards optimism in technology has been labeled the technology effect" (Gravett, 2024, p. 288).

5. Discussion and conclusions

An analysis of the literature leads to the conclusion that in the age of AI, responsibility cannot be delegated to algorithms. Technology, regardless of its sophistication and autonomy, remains a tool in the hands of people. Trust in AI systems cannot be blind; it must be a conditional trust, based on evidence of reliability and transparency of processes (Hallowell et al., 2022).

An analysis of the literature leads to the conclusion that in the age of AI, responsibility cannot be delegated to algorithms. The growing autonomy of decision-making systems is one of the greatest ethical and legal challenges, leading to the risk of creating the so-called

"responsibility gap". In a moral and legal context, the literature consistently indicates that artificial intelligence, without intentionality, cannot be considered a moral subject. Technology, regardless of its sophistication and autonomy, remains a tool in the hands of people. Trust in AI systems cannot be blind; it must be a conditional fund, based on evidence of credibility and transparency of processes (Hallowell et al., 2022).

Still, there is a tendency for human decision-makers to use AI as a "scapegoat" to avoid blame for mistakes or adverse outcomes.

This phenomenon is particularly dangerous in the light of modern research. Research proves that just disclosing the fact of using AI in the decision-making process can psychologically relieve human advisors or decision-makers of a strong sense of responsibility for the final outcome. This psychological release can lead to uncritical acceptance of algorithmic recommendations, known as "automation bias", where leaders, managers, or judges thoughtlessly accept the results generated by the systems.

Key takeaways for leaders and policymakers include:

1. Rejection of the anthropomorphization of responsibility: AI cannot be a moral subject. Attempts to assign blame to algorithms are a form of avoidance on the part of creators and users.
2. Critical evaluation and education: To counter the bias of automation, users of AI systems (judges, doctors, managers) must be trained not only in the use of the tools, but also in understanding their limitations and potential errors (Gravett, 2024; Mendel et al., 2024).
3. Design with collaboration in mind: AI systems should be designed to support human will, not limit it. Shared responsibility models prove to be more effective at building satisfaction and trust than simple automation (Han, Ko, 2025).

To effectively counter these threats and ensure that AI remains a supporting tool rather than a substitute for human judgment, it is crucial to implement "Meaningful Human Oversight" procedures. This means that regardless of the level of sophistication of the system, the final decision and legal responsibility must rest with the person. Artificial intelligence should be treated only as an advanced support tool, while it is the human operator who must retain the ability to verify, challenge and reject its recommendations, thus guaranteeing an ethical and responsible decision-making process.

6. Directions for further research

Future analyses should go beyond the conceptual framework and focus on empirical verification of the TRUST model in various high-risk sectors. A key area of research is to determine whether training leaders to distinguish between "moral trust" in humans and

"technical reliance" on AI actually reduces excessive technological optimism. Longitudinal studies on the impact of AI on leadership competencies are also necessary to assess whether long-term collaboration with algorithms leads to the erosion of critical thinking and decision-making skills in crisis situations when the system fails.

References

1. Alasmri Naeem, Sarah Basahel (2022). Linking Artificial Intelligence Use to Improved Decision-Making, Individual and Organizational Outcomes. *International Business Research, 15, 10,* 1-15. https://doi.org/10.5539/ibr.v15n10p1
2. Assaad, Zena (2024). *A Risk-Based Trust Framework for Assuring the Humans in Human-Machine Teaming.* Second International Symposium on Trustworthy Autonomous Systems (TAS '24), Austin, TX, USA: ACM, https://doi.org/10.1145/3686038.3686045
3. Behera, Rajat Kumar, Pradip Kumar Bala, Nripendra P. Rana (2023). Creation of Sustainable Growth with Explainable Artificial Intelligence: An Empirical Insight from Consumer Packaged Goods Retailers. *Journal of Cleaner Production, 399,* 136605. https://doi.org/10.1016/j.jclepro.2023.136605
4. Bruynseels, Koen, Lotte Asveld, Jeroen van den Hoven (2025). Foundation Models for Research: A Matter of Trust? *Artificial Intelligence in the Life Sciences, 7,* 100126. https://doi.org/10.1016/j.ailsci.2025.100126
5. Bunnell, Arianna, Sharon Rowe (2023). The Effect of AI-Enhanced Breast Imaging on the Caring Radiologist-Patient Relationship. In: H. Sousa Antunes et al. (eds.), *Law, Multidisciplinary Perspectives on Artificial Intelligence and the Law*. *Governance and Technology Series, 58*. Springer.
6. Busuioc, Madalina (2021). Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review, 81, 5,* 825-36. https://doi.org/10.1111/puar.13293
7. European Parliament (2019). *Directorate General for Parliamentary Research Services. Understanding Algorithmic Decision-Making: Opportunities and Challenges*. Brussels: European Union. doi:10.2861/536131
8. Flórez Rojas, María Lorena (2025). Algocracy in the Judiciary: Challenging Trust in the System. *Revista de Estudios Sociales, 93,* 107-28. https://doi.org/10.7440/res93.2025.06
9. González Arencibia, Mario, Hugo Ordoñez-Erazo, Juan-Sebastián González-Sanabria (2024). Explainable Artificial Intelligence as an Ethical Principle. *Ingeniería, 29, 2,* e21583. https://doi.org/10.14483/23448393.21583
10. Gravett, Willem H. (2024). Judicial Decision-Making in the Age of Artificial Intelligence." In: H. Sousa Antunes et al. (eds.), *Multidisciplinary Perspectives on Artificial Intelligence*

and the Law (pp. 281-97). *Law, Governance and Technology Series, 58.* Springer, https://doi.org/10.1007/978-3-031-41264-6_15

11. Hallowell, Nina, Shirlene Badger, Aurelia Sauerbrei, Christoffer Nellåker, Angeliki Kerasidou (2022). 'I Don't Think People Are Ready to Trust These Algorithms at Face Value': Trust and the Use of Machine Learning Algorithms in the Diagnosis of Rare Disease. *BMC Medical Ethics, 23, 112,* https://doi.org/10.1186/s12910-022-00842-4
12. Han, Jihyung, Daekyun Ko (2025). Consumer Autonomy in Generative AI Services: The Role of Task Difficulty and AI Design Elements in Enhancing Trust, Satisfaction, and Usage Intention. *Behavioral Sciences, 15, 4,* 534. https://doi.org/10.3390/bs15040534
13. Joshi, Satyadhar (2025). Comprehensive Review of Artificial Intelligence in Management, Leadership, Decision-Making and Collaboration. *International Journal of Science and Social Science Research, 3, 1,* p. 49.
14. Lima, Gabriel, Nina Grgić-Hlača, Jin Keun Jeong, Meeyoung Cha (2022). *The Conflict Between Explainable and Accountable Decision-Making Algorithms.* ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), 1-18. Seoul, Republic of Korea: ACM. https://doi.org/10.1145/3531146.3534628
15. Madanchian, Mitra, Hamed Taherdoost, Michele Vincenti, Nachaat Mohamed (2024). Transforming Leadership Practices through Artificial Intelligence. *Procedia Computer Science, 235,* 2101-11. https://doi.org/10.1016/j.procs.2024.04.199
16. Maris, Menno T., Ayca Koçar, Dick L. Willems, Jeannette Pols, Hanno L. Tan, Georg L. Lindinger, Marieke A.R. Bak (2024). Ethical Use of Artificial Intelligence to Prevent Sudden Cardiac Death: An Interview Study of Patient Perspectives. *BMC Medical Ethics, 25, 42.* https://doi.org/10.1186/s12910-024-01042-y
17. Mendel, Tamir, Oded Nov, Batia Wiesenfeld (2024). Advice from a Doctor or AI? Understanding Willingness to Disclose Information Through Remote Patient Monitoring to Receive Health Advice. *Proceedings of the ACM on Human-Computer Interaction, 8, CSCW2, Article 386*. https://doi.org/10.1145/3686925
18. Mikalef, Patrick, Kieran Conboy, Jenny Eriksson Lundström, Aleš Popovič (2022). Thinking Responsibly about Responsible AI and 'the Dark Side' of AI. *European Journal of Information Systems, 31, 3,* 257-68. https://doi.org/10.1080/0960085X.2022.2026621
19. Nitzl, Christian, Achim Cyran, Sascha Krstanovic, Uwe M. Borghoff (2025). The Use of Artificial Intelligence in Military Intelligence: An Experimental Investigation of Added Value in the Analysis Process. *Frontiers in Human Dynamics, 7,* 1540450. https://doi.org/10.3389/fhumd.2025.1540450
20. Schafheitle, S.D., A. Weibel, A. Rickert (2021). The Bermuda Triangle of Leadership in the AI Era? Emerging Trust Implications from 'Two-Leader-Situations' in the Eyes of Employees. In: T.X. Bui (ed.), *Proc. Annu. Hawaii Int. Conf. Syst. Sci., vols. 2020-January*. IEEE Computer Society. Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85108319863&partnerID=40&md5=741cc62ca6ca3cb5ac3991ee719c6143

21. Sychev, A.A. (2024). Looking for confidence: transformation of trust in the digital age. *Monitoring Obshchestvennogo Mneniya: Ekonomicheskie i Sotsial'nye Peremeny, 183, 5,* 38-59. Scopus. https://doi.org/10.14515/monitoring.2024.5.2595
22. Trojańczyk, Adam (2024). *Przywództwo w czasach AI: Minipodręcznik dla liderów o zarządzaniu kryzysem w epoce maszyn generatywnych*. Łódź: Adam Trojańczyk, ISBN: 978-83-966626-0-6