JIRRS
Journal International Review of Research Studies

Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

Artificial Intelligence, Ethics, and Organizational Trust: A Systematic Review and Future Research Agenda

Fajer Danish[1] and Muneera Al-Khalifa[2]
[1],[2]Royal University for Women, Bahrain
fajerdanish@gmail.com
mmalkhalifa@ruw.edu.bh

Abstract

Artificial Intelligence (AI) systems are becoming deeply ingrained in organizational decision-making processes due to their ability to improve efficiencies, scalability and analytics-related accuracy. Consequently, organizations have faced more severe ethical issues around transparency and bias, as well as sustainability challenges around trustworthiness. While there has been some research conducted on AI capabilities, ethics and human trust in AI systems, research has remained siloed across academic fields. This paper aims to conduct a review of recent literature exploring AI capabilities and ethics and their relationship with organizational trust.

The study will be completed by conducting a literature review on peer-reviewed articles published from 2020 to 2025 using a PRISMA-guided methodology. Articles will be reviewed based on keywords related to artificial intelligence-enabled decision-making, ethical issues in organizations, explainability and trust in organizational settings.

After applying inclusion and exclusion criteria, four central themes were found: 1) AI as a source of competitive advantage for organizational decision-making; 2) AI ethics, including bias, opacity, and lack of accountability; 3) Trust as a multicomponent process developed through performance, transparency, and accountability; and 4) Responsible AI governance as a moderating factor between efficiency and legitimacy. There is existing research to suggest that trust in technology is formed not only by performance but by transparency and explainability factors as well as ethical acceptability (Afroogh, 2024; Papagiannidis et al., 2025).

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

This paper offers several contributions. First, it combines both strategic and behavioral schools of thought into one framework for understanding how trust can be formed in AI-mediated decision-making processes. Second, it identifies gaps in current literature and suggests a future research agenda including multilevel trust, industry differences in AI and ethical considerations, and measurement of responsible AI. Overall, AI-enabled decision-making is not sustainable if the organization cannot effectively implement AI with ethical risk mitigation practices.

Keywords: Artificial Intelligence (AI); Organizational trust; Ethical AI; Explainable AI; (XAI), Trustworthiness (ability, benevolence, integrity).

1. Introduction

Artificial Intelligence (AI) decision-making has become an integral part of how firms process information, predict outcomes, and take action. Organizations are adopting AI to improve decision quality, speed, and scale, making AI a foundational capability of smart organizations. Yet AI systems bring with them performance benefits risks to organizational transparency, fairness, and accountability (Papagiannidis et al., 2025).

According to recent literature, there is a significant gap in how organizations balance efficiency vs legitimacy when implementing AI. Improved decision-making comes at the cost of less explainability or justifiability of these decisions to stakeholders, also known as a "black box" effect (Afroogh, 2024) . This concern is particularly acute in organizational settings where AI is used to make decisions that impact employees, customers, and other stakeholders.

Recent reviews have shown that trust in AI is not only determined by performance expectations, but also by social, contextual, and governance factors. Factors that affect trust include transparency, fairness, accountability, and other perceptions related to organizational intent (Reinhardt, 2023) . The prevalence of the term "trustworthy AI" in academic and policy making circles has also created a need to better understand how this concept is measured and utilized by organizations.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

Trustworthy AI, governance practices, and ethics in AI have each been studied extensively. However, research has remained somewhat siloed. Research around AI focuses on the performance capabilities of AI, AI ethics research has centered around governance and risks, and AI and trust research has focused on end-user perspectives. A recent systematic literature review found that while there is emerging research on AI ethics and trust, a comprehensive review of the current evidence is still lacking (Papagiannidis et al., 2025).

To address this gap, this literature review will explore the relationship between AI decision-making, AI ethics and governance, and trust within organizations. The study will draw from current evidence to map out mechanisms of how and why trust is formed in AI capabilities and identify avenues for future research.

---

2. Literature Review

2.1 AI as an Organizational Capability

Artificial intelligence (AI) capabilities have transitioned from an operational enabler to a strategic capability in modern organizations. With the emergence of smart organizations powered by data analytics and automation, AI has allowed firms to leverage large-scale data processing, predictive decision-making, and real-time responsiveness (Afroogh et al., 2023). Recent literature conceptualizes AI as a source of competitive advantage when applied to value-producing processes and architectures (Dwivedi et al., 2021; Papagiannidis et al., 2025).

Still, researchers caution against presuming organizations achieve a competitive advantage through AI adoption. Several authors propose AI's effectiveness may be determined by complementary capabilities related to learning agility, governance mechanisms, and human-AI symbiosis (Mikalef et al., 2021; Ransbotham et al., 2020). Additionally, others find AI-assisted decision-making can improve decision efficiency and quality by mitigating human bias and partially introducing machine bias (Mehrabi et al., 2021).

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

2.2 Ethical Challenges Associated with AI Decision-Making

Ethical issues emerge as one of the most prolific topics associated with AI adoption and diffusion. Themes related to bias, transparency, explainability, accountability, and fairness dominate the ethical AI conversation among scholars and practitioners. Algorithmic bias, which can arise when an AI system is trained on a set of data that includes historical inequities, has been found to be prevalent in AI systems (Mehrabi et al., 2021). Unfortunately, algorithmic bias is difficult to detect in machine learning systems as many classification algorithms are not easily interpretable by users.

Research in the space has emphasized issues related to transparency and explainability. Organizations have adopted explainable AI (XAI) initiatives to make algorithmic decisions more interpretable to stakeholders. But critics argue that AI decisions can be technically explainable while still being perceived as unethical by stakeholders (Rai, 2020; Afroogh, 2024).

Existing research further indicates that accountability in AI decision-making may be unclear if decision-making authority is blurred between humans and machines. According to Floridi et al. (2021), organizations with poor AI accountability mechanisms put their legitimacy at risk. To help standardize ethical AI decision-making, several authors have proposed frameworks and guidelines. For instance, Papagiannidis et al. (2025) offered a theoretical ethical AI governance framework based on four principles: fairness, transparency, responsibility, and human agency.

2.3 Trust within an Organization

Trust within organizations has been defined as an attitude that creates a willingness to rely on another party. Parties within an organization develop this willingness when they have positive expectations of another party's trustworthiness (Kersten et al., 2022). Several studies have focused on trust within the context of AI decision-making. These studies help to determine if stakeholders will trust the decisions generated by AI systems.

Afroogh (2024) identified three factors that influence trust within an organization: technical, social, and contextual. Technical factors include the ability of AI to make accurate

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

predictions. From a social standpoint, trust is influenced by one's beliefs about the fairness of AI decisions. Lastly, context can also impact trust if the stakes are high or decision-making is new to the organization. Overall, Afroogh (2024) found that AI can be trustworthy but is not overridden by performance alone.

Research by Reinhardt (2023) offers additional insight into trust within organizations related to AI. The author stresses that when humans are presented with AI decisions, they do not only evaluate the decision but rather what the organization values. If the decision created by AI is perceived as unfair, unethical, or biased by the stakeholder, then the trust of the organization as a whole may be compromised.

2.4 Trustworthiness: Ability, Benevolence, and Integrity

Earlier, organizational trust was described as a function of one party's trustworthiness. But what defines trustworthiness? According to organizational behavior literature, trustworthiness has three fundamental dimensions: ability, benevolence, and integrity (Cribbin, 2022). The ability to assess these constructs in an AI framework can provide meaningful insight into AI and trust.

Ability is defined by a party's competency regarding a specific subject. Relating this to AI, some may question an AI system's ability if it does not perform well on given tasks. However, Glikson and Woolley (2020) found that AI is likely to be trusted if the machines demonstrate a high level of performance.

Benevolence within an organization can be exemplified by the intentions of decision-makers. As AI is presented with more decisions that can affect people, stakeholders want to ensure the organizations are not using AI to take advantage of their employees (Afroogh, 2024). Lastly, integrity can be determined by AI's adherence to ethical guidelines. Floridi et al. (2021) discuss how algorithmic decisions can be perceived as unbiased but unethical. If that is the case, stakeholders will not trust the organization's AI decisions.

2.5 Responsible AI Governance

Taken together, recent literature suggests responsible AI governance serves as the mechanism through which AI capabilities are translated into ethical and trustworthy AI

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

decisions. Responsible AI governance includes principles, mechanisms, and processes that ensure AI systems operate according to ethical and moral standards.

Similar to Afroogh et al. (2023), responsible AI governance mechanisms include explainability, auditability, establish human monitoring and oversight, and other tasks that enhance organizational trust. Ultimately, governance can be considered a capability that organizations can possess.

If an organization emphasizes building AI governance, they will be more likely to create efficient and trustworthy AI systems. On the contrary, if an organization neglects this critical capability, they may see high performance but low trust in their AI decisions.

2.6 Summary and Gap

In summary, while each body of literature provides meaningful insight into AI, there are still gaps that have yet to be addressed. Specifically, AI as a strategic capability, ethical issues with AI, and trust have been studied heavily but rarely together. Additionally, while scholars such as Papagiannidis et al. (2025) provide frameworks for ethical AI, there are no clear implications on how these can be practiced in organizations. Lastly, while recent studies such as Afroogh et al. (2023) mention responsible AI governance, there are limited studies that explore AI capabilities leading to trust with governance as a mediator.

As a result, there is a gap in the literature identifying how organizations can use AI to make better decisions while maintaining ethical standards and earning stakeholders' trust.

Table 1. Literature Base for the Review: AI, Ethics, and Organizational Trust (2020–2025)

No.	Study	Year	Theory / Lens	Key Variables	Method	Core Focus	Contribution to This Study
1	Glikson & Woolley	2020	Trust Theory	Trust, reliability, transparency	Review	Trust in AI	Foundational trust dimensions
2	Rai	2020	Information Processing	Explainability, transparency	Conceptual	XAI	Links explainability to trust
3	Siau & Wang	2020	Ethics Theory	Fairness, accountability, transparency	Conceptual	AI ethics	Supports ethical AI construct

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

No.	Study	Year	Theory / Lens	Key Variables	Method	Core Focus	Contribution to This Study
4	Dwivedi et al.	2021	Technology Adoption	AI adoption, performance	Review	AI impact	Supports AI capability
5	Mehrabi et al.	2021	Fairness Theory	Bias, discrimination	Review	AI bias	Supports ethical risks
6	Shin	2021	Trust Theory	Explainability, trust, causability	Empirical	Trust in AI	Strong trust mechanism evidence
7	Mikalef et al.	2021	RBV	AI capability, firm performance	Empirical	Performance	Supports strategic capability
8	Chen & Wen	2021	Institutional Theory	Trust, attitudes	Empirical	Public trust	Contextual trust formation
9	Yam & Skorburg	2021	Ethics / Justice Theory	Fairness, rights	Conceptual	Hiring AI	Ethics in HR decisions
10	Vassilakopoulou et al.	2022	Responsible AI	Governance, accountability	Conceptual	AI responsibility	Governance framing
11	Figueroa-Armijos et al.	2023	Trust Theory	Ethics, trust	Empirical	Hiring AI	Ethics → trust link
12	Balasubramaniam et al.	2023	Governance Theory	Transparency, explainability	Conceptual	AI requirements	Operationalizing ethics
13	Jackson	2023	Behavioral Theory	Trust, fairness perception	Empirical	Algorithmic grading	Trust breakdown cases
14	Afroogh et al.	2024	Trust Theory	Trust, transparency	Review	AI trust	Updated trust synthesis
15	Henrique et al.	2024	Trust Framework	Trust, adoption	Review	Trust evolution	Field fragmentation
16	Coussement et al.	2024	Decision Theory	Explainability, decision quality	Conceptual	XAI	Decision legitimacy
17	Sadeghi et al.	2024	Dynamic Capability	XAI, agility	Conceptual	Decision agility	AI responsiveness
18	Vanneste et al.	2024	Agency Theory	Perception, trust	Conceptual	AI agency	Human vs AI perception

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

No.	Study	Year	Theory / Lens	Key Variables	Method	Core Focus	Contribution to This Study
19	Lahusen et al.	2024	Governance Theory	Trustworthiness, governance	Conceptual	AI governance	Trust-building structures
20	Batool et al.	2025	Governance Theory	AI governance mechanisms	SLR	AI governance	Multi-level governance
21	Papagiannidis et al.	2025	Responsible AI	Governance, ethics	Review	AI governance	Core integrator
22	Bach et al.	2025	Practice Theory	Governance practices	Review	RAI implementation	Practical governance
23	Vining et al.	2025	Socio-technical Systems	Trust, systems	Review	Trustworthy AI	Organizational trust
24	Blanco	2025	Trust Theory	Trust vs reliance	Conceptual	Trust distinction	Conceptual clarity
25	Chang	2025	XAI Theory	Explainability, adoption	Review	XAI in business	Expands XAI literature
26	Chaudhary et al.	2025	Innovation Theory	XAI adoption	Conceptual	Organizational AI	Adoption mechanisms
27	Cetinkaya & Krämer	2025	Perception Theory	Transparency, trust	Empirical	AI perception	Trust drivers
28	Schlicker et al.	2025	Trustworthiness Model	Trustworthiness, evaluation	Conceptual	TrAM model	Trust measurement
29	Schilke & Reimann	2025	Behavioral Theory	Transparency, trust	Empirical	Disclosure effects	Transparency paradox
30	Do Khac et al.	2025	Trust Model	Human-AI trust	Conceptual	Org trust	Multi-level trust
31	Staley	2025	Finance / Decision Theory	XAI, trust	Mixed	Financial AI	Sector application
32	Daly et al.	2025	Organizational Behavior	Trust, adoption	Qualitative	AI adoption	Social trust dynamics
33	Phan et al.	2025	Risk Theory	Risk, trust, transparency	Empirical	AI platforms	Risk–trust link
34	Dahlin	2025	Philosophy of Trust	Trust, ethics	Conceptual	AI trust	Deep trust theory
35	Bar-Gil et al.	2025	Institutional Theory	Ethics, governance	Comparative	AI ethics	Contextual variation

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

No.	Study	Year	Theory / Lens	Key Variables	Method	Core Focus	Contribution to This Study
36	Lee et al.	2025	ESG / Governance	AI governance, ESG	Conceptual	Responsible AI	Strategic governance

Table 1 comes from an extensive list of key peer-reviewed and curated high-quality review literature published in recent years, from 2020 to 2025. This time frame was chosen to ensure inclusion of cutting-edge research and foundational studies (especially in trust literature, Glikson and Woolley, 2020), nascent literature that is beginning to emerge on explainability/transparency (Rai, 2020; Shin, 2021), and the prolific responsible AI literature published recently (Papagiannidis et al., 2025; Batool et al., 2025). Explicit theoretical lenses and variables were added to help summarize the rich complexity within this body of literature, both conceptually and empirically.

Based on Table 1, the dominant themes in the literature between 2020 and 2025 center around four key streams of thought: (1) AI and trust/trustworthiness, (2) Explainability/ transparency as interpretability, (3) Bias/Fairness/ Accountability as ethical risks, and (4) Responsible AI governance as an organizational capability. Table 1 not only shows themes clustering around these four streams, but also indicates a chronology to the literature. Earlier publications tend to focus on antecedents of trust and explainability at the individual or system level. Later publications tend to focus on responsible AI governance, operationalization of responsible AI, and organization-level and multilevel trust.

Finally, the table shows the inclusion of theory is on the rise within this domain. Theory ranged from well-established theory (Trust Theory, Resource-Based View (RBV), Governance Theory) to more hybridized theory (Socio-technical Systems Theory). While integration is forthcoming within this body of literature, Table 1 highlights that few studies make the connection between performance (efficiency, quality of decisions) afforded from AI at the organizational level and how ethics/governance mechanisms and/or organizational trust form within that same model.

This need serves as motivation for this systematic review because it places efficiency, ethics, and organizational trust within one comprehensive framework. Recently, reviewers have

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

advocated for expanding research beyond siloed constructs and focusing on a comprehensive understanding of AI adoption (Siau et al., 2021).

3. Methodology

3.1 Research Design

The methodology utilized in this paper is Systematic Literature Review (SLR). This type of literature review will allow the authors to summarize and critically analyze previous research done in the areas of AI, ethics, and organizational trustworthiness. Artificial intelligence research requires SLR since this phenomenon is complex and spread across disciplines, and SLR will allow for both conceptual papers, empirical papers, and review papers to be analyzed together. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA 2020 Statement will be utilized when completing this literature review (Page et al., 2021). By using this strategy, complete transparency of locating, screening, and analyzing literature will be reported. This method was selected to increase the methodological rigors, transparency, and reproducibility of this review. Readers can expect this methodology to be used in papers within Scopus Q1 journals.

3.2 Review Protocol and Research Questions

The review is structured around four key research questions:

- RQ1: How is AI conceptualized as a decision-making capability in organizational contexts?
- RQ2: What ethical challenges are associated with AI-driven decision-making?
- RQ3: What factors influence organizational trust in AI systems?
- RQ4: How do governance mechanisms mediate the relationship between AI adoption and trust?

These questions guide the selection, coding, and synthesis of the literature.

3.3 Search Strategy

A systematic search was conducted across major academic databases, including:

- Scopus
- Web of Science

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

- ScienceDirect
- SpringerLink

The search focused on peer-reviewed journal articles published between 2020 and 2025, ensuring relevance to contemporary developments in AI and organizational research. Search String includes:

- ("artificial intelligence" OR "AI" OR "algorithmic decision-making")
  AND
- ("organizational trust" OR "trust in AI" OR "trustworthiness")
  AND
- ("ethics" OR "ethical AI" OR "responsible AI" OR "fairness" OR "accountability" OR "transparency" OR "explainability")
  AND
- ("organization*" OR "firm*" OR "enterprise*" OR "business")

3.4 Inclusion and Exclusion Criteria

Inclusion Criteria

- Peer-reviewed journal articles
- Published between 2020–2025
- Focus on AI in organizational or decision-making contexts
- Address at least one of the following: trust, ethics, governance, fairness, transparency, or explainability
- Written in English

Exclusion Criteria

- Purely technical or algorithmic studies without organizational implications
- Conference papers, editorials, dissertations, and book chapters
- Studies lacking relevance to ethical or trust-related dimensions

3.5 Study Selection Process (PRISMA Framework)

The study selection followed the four-stage PRISMA process: identification, screening, eligibility, and inclusion. Figure 1 illustrates the study selection process conducted.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

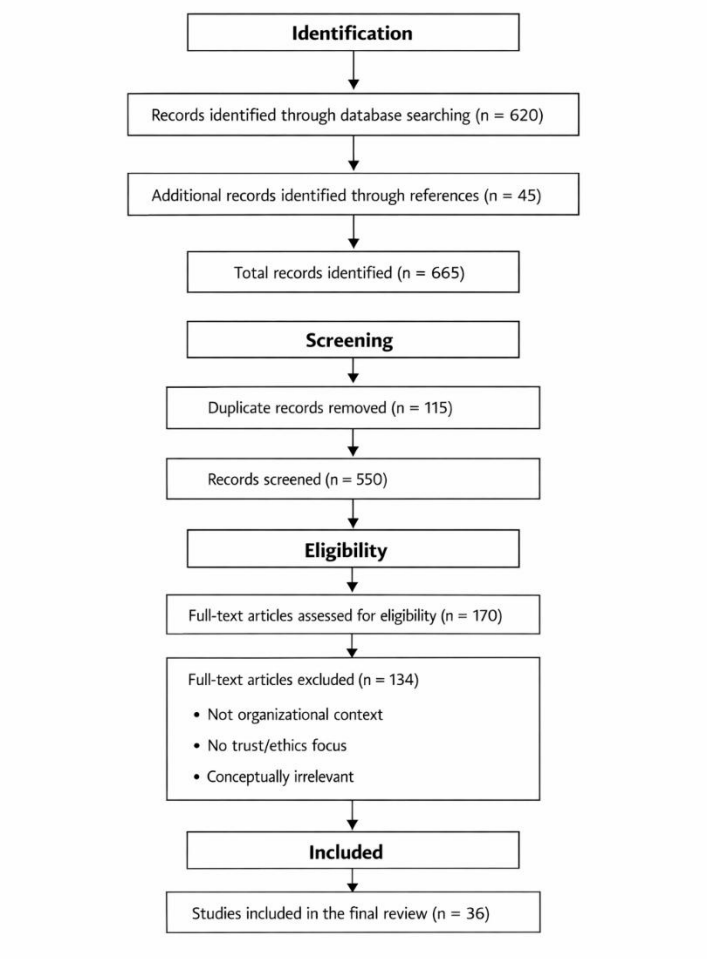

Figure 1. PRISMA flow diagram illustrating the study selection process.

3.6 Data Extraction and Coding

A structured data extraction protocol was employed to ensure consistency and analytical rigor. Each selected study was coded across multiple dimensions, including:

- author(s) and publication year
- theoretical framework
- key variables and constructs
- research methodology
- sector or organizational context
- principal findings
- relevance to AI, ethics, and trust

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

This structured approach enabled the development of a comprehensive synthesis (Table 1), facilitating cross-study comparison and thematic integration.

3.7 Data Analysis Approach

The study employs a thematic synthesis approach, which is particularly suitable for integrating heterogeneous literature across conceptual, empirical, and review-based studies. Unlike meta-analysis, which requires homogeneous quantitative datasets, thematic synthesis allows for the identification of recurring patterns, relationships, and conceptual linkages across diverse research designs.

Through iterative coding and analysis, four dominant themes were identified:

1. AI as a strategic organizational capability
2. Ethical risks and governance challenges
3. Trust formation in AI-mediated environments
4. Responsible AI governance as an integrative mechanism

These themes form the analytical foundation of the literature review presented in Section 2.

3.8 Quality Assessment of Included Studies

To enhance methodological rigor, a structured quality assessment protocol was applied to all included studies. Each article was evaluated based on:

- methodological robustness (e.g., empirical rigor, sample adequacy)
- clarity and relevance of theoretical grounding
- alignment with the study's research objectives
- transparency of findings and acknowledgment of limitations

Rather than excluding studies solely based on quality, the review adopts a weighted analytical approach, giving greater emphasis to high-quality empirical and systematic review studies. This ensures a balanced synthesis that captures both theoretical development and empirical evidence.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

3.9 Data Coding and Reliability

To ensure consistency, the coding process followed an iterative refinement approach, whereby initial coding categories were continuously reviewed and adjusted to improve conceptual clarity and alignment with the research questions.

Although the review was conducted by a single researcher, reliability was enhanced through:

- repeated cross-checking of extracted data
- alignment with established constructs (e.g., trustworthiness dimensions, AI governance principles)
- systematic comparison across studies

This approach minimizes subjectivity and enhances the credibility of the thematic synthesis.

3.10 Bias and Limitations in the Review Process

There were several potential areas for bias that were addressed. Firstly, database bias was addressed by searching a variety of databases known to be large, academic databases covering most disciplines. Secondly, publication bias is present in this review as only peer reviewed journal articles are being sourced. Journal articles may overrepresent research that have significant findings or positive findings. Thirdly, there will be language bias as we are only searching for English language studies. Finally, only including studies within the time range of 2020–2025 will limit the studies that may have occurred prior. However, this was kept in mind while searching for applicable literature as AI technology and governance are quickly changing and evolving. We wanted literature that could best apply to the organizations today

3.11 Methodological Rigor and Transparency

Following PRISMA 2020 statement (Page et al., 2021), this study provides transparency regarding its search, screening, and selection procedures. Explicit inclusion criteria, standardized data extraction form, and methodical thematic analysis strengthen the dependability and reproducibility of this review.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

Taken together, these rigorous processes ensure the quality of this review commensurate with top-tier academic journals and supply valuable knowledge to further support inquiry into AI, ethics, and organizational trust.

---

4. Discussion and Theoretical Contributions

4.1 Integrating Efficiency, Ethics, and Trust in AI-Driven Organizations

This review not only complements but also advances the body of literature reviewed in Section 2. It echoes prior findings that AI serves the dual roles of efficiency booster and organizational source of tension (Dwivedi et al., 2021; Mikalef et al., 2021). In line with previous literature, findings from this review reaffirm the robust relationship between AI adoption and decision-making speed, scalability, and analytic precision. However, consistent with recent streams of literature focusing on governance (Papagiannidis et al., 2025; Batool et al., 2025), they also illustrate how these efficiency dividends are often coupled with trust and ethical tensions.

Crucially, this review highlights that efficiency and trust dynamics do not evolve conjointly; rather, they unfold via dissociated mechanisms. If initial publications were primarily focused on performance outcomes (Ransbotham et al., 2020), more recent works document AI adoption dysfunctions such as opacity or lack of interpretability (Rai, 2020; Afroogh et al., 2024). Building on these findings, this review reveals that tensions between efficiency and legitimacy levers must be managed via conscious interventions.

4.2 Reframing Organizational Trust in AI Contexts

The findings directly support the theory discussed in Section 2. Most notably, previous research conducted by Glikson and Woolley (2020) suggested that trust with AI was multi-dimensional and driven by both technical and social components. This literature review further supports this idea by showing that trust is less about performance metrics and more about perceptions of fairness, transparency and organizational motivations (Afroogh et al., 2024; Reinhardt, 2023).

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

The review also expands upon past literature concerning trust by providing insight on how each dimension of trustworthiness is remodeled within AI contexts. Aligning with past data (Shin, 2021), ability continues to have a strong positive correlation with trust as technical performance still matters. However, this review found that both benevolence and integrity are less attributed to human-related factors and are more associated with governance measures. This is in line with current discussions regarding trust with AI being institutional rather than interpersonal (Vining et al., 2025).

4.3 Responsible AI Governance as the Integrative Mechanism

Responsible AI governance emerged as the main moderator between AI capability and trust. This finding contributes to the governance stream of literature discussed in Section 2 (Papagiannidis et al., 2025; Batool et al., 2025). Previous literature examined governance in terms of ethical principles like fairness and transparency(Floridi et al., 2021; Siau & Wang, 2020). However, recent studies have placed greater emphasis on the concept of responsible AI governance.

Empirical evidence suggests that various governance mechanisms (e.g., explainability, auditability, and human oversight) influence trust outcomes. However, this relationship appears contingent on contextual factors as well. Similar to recent studies (Schilke & Reimann, 2025), our review unearthed evidence for transparency paradoxes in certain instances as well. This means governance mechanisms must not only be present but must also be recognized by stakeholders as well.

Therefore, this research further develops the literature by conceptualizing governance as a process that translates organizational capabilities into legitimate outcomes.

4.4 Theoretical Interpretation of the Findings

First, our study helps consolidate prior streams of research related AI capability (Mikalef et al., 2021), AI ethical risk (Mehrabi et al., 2021), and trust formation (Glikson & Woolley, 2020). Second, we extend prior research on trust theory to AI ecosystems by demonstrating that dimensions of trustworthiness are mediated via constructs such as governance rather than trust formation from AI developers to data providers. Third, our study responds to

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

recent calls for reconceptualizing governance not just as a principle-driven or abstract notion (Papagiannidis et al., 2025) but rather as a capability. Finally, our study furthers recent efforts in AI research moving beyond exploring constructs of adoption and performance to legitimacy and trusting relationships longitudinally (Afroogh et al., 2024).

4.5 Practical Implications of the Findings

The findings provide several actionable implications for practitioners.

Consistent with governance literature (Batool et al., 2025), organizations should prioritize the development of structured AI governance frameworks, including explainability tools and accountability mechanisms. However, aligning with recent empirical evidence (Cetinkaya & Krämer, 2025), managers must also recognize that stakeholder perceptions play a critical role in shaping trust outcomes.

Furthermore, the review suggests that organizations should move toward human–AI collaboration models, as highlighted in recent studies (Do Khac et al., 2025), rather than relying solely on automation. This approach enhances both decision quality and stakeholder acceptance.

4.6 Research Gaps Emerging from the Findings

Building on the gaps identified in Section 2 and the synthesis presented in Table 1, several important avenues for future research have been identified throughout this Research Topic. For example, trust should be examined through a multilevel lens (Vining et al.). Although many works in this collection primarily focus on individual-level constructs of trust, it is important to understand how trust operates at organizational and institutional levels. In addition, empirical testing of the theoretical frameworks presented by Papagiannidis et al. and Batool et al. is needed to understand how proposed AI governance mechanisms affect trust within organizational contexts over time. Further, future research should account for potential differences between sectors. As AI is being adopted to fulfill ethical expectations across industries, understanding how ethical decision-making may differ between sectors such as health care, accounting, and human resource management is critical (Figueroa-Armijos et al.). Related, more research should be conducted to understand the relationship

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

between explainability and trust. While several articles touch on this concept, previous research has found that algorithmic explainability does not always lead to increased levels of trust and can even cause distrust (Rai, 2020; Schilke & Reimann, 2025). Exploring how humans and AI models can make decisions together will also provide researchers and practitioners with solutions that not only create quality decisions but are also accepted by key stakeholders (Do Khac et al.).

5. Conclusion

5.1 Conclusion

For this review, I systematically analyzed peer-reviewed literature on AI spanning 2020 to 2025 to gain additional insight into how AI and ethics influence organizational trust.

To that end, the review asked and answered four primary questions focused on understanding AI as a decision-making capability, understanding the ethical considerations that emerge with AI use, understanding what factors determine trust in AI, and understanding what governance mechanisms influence these relationships.

AI is defined as a decision-making capability that allows organizations to make faster, more scalable, and data-driven decisions. The ethical considerations of this capability include bias, opacity, and lack of accountability. While multiple factors determine trust in AI, perceptions of fairness, explainability, and organizational intention all play important roles. Finally, responsible AI governance mechanisms serve as the mechanism that links AI with trust.

Together, these findings show that the relationship between AI performance and organizational adoption is dependent on responsible AI governance, suggesting a shift in AI literature away from performance toward legitimacy.

5.2 Theoretical Contributions

First, this review bridges gaps in literature by bringing together three bodies of research: AI as a capability, responsible AI governance, and organizational trust. While each stream of research has explored these topics independently, this review found that they are inherently intertwined.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

Second, this review contributes to the organizational trust literature by providing additional insight into how traditional facets of trustworthiness (i.e., ability, benevolence, integrity) are manifested in AI. Unlike traditional determinants of trust, AI-based ability, benevolence, and integrity are not directly observable; rather, they are reflected by system performance, responsible AI governance, and organizational practices.

Finally, this review makes contributions to the strategic management literature by highlighting that responsible AI governance should be treated as a strategic capability as opposed to a regulatory afterthought. Treating governance as a capability will encourage leaders to align AI use with organizational practices and values.

5.3 Managerial Implications

The findings of this review have several managerial implications. Chief among them is that managers should begin to treat AI through a governance lens. While AI allows organizations to become more efficient, efficient decisions may not be trusted if stakeholders do not understand how they are made.

Instead, leaders should invest in establishing explainability protocols, responsibility frameworks, and human oversight of AI technologies. While many organizations have adopted AI to improve efficiency, careful attention should be paid to how these tools will be received by stakeholders. As such, managers should approach AI as a human–AI collaboration and ensure that AI technologies are not the sole decision-maker.

While it is tempting to use AI because you can, responsible AI use may mean using AI only when you should. Organizations that focus on gaining stakeholder trust through responsible AI practices will create long-term legitimacy. In contrast, others who pursue AI solutions purely for efficiency will lose that trust.

5.4 Policy and Societal Implications

Policy-wise, the findings suggest that governments and institutional bodies should work to create standards and regulations for AI use. Organizations will not police themselves, so it is up to these bodies to require organizations to be transparent, unbiased, and accountable.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

In addition, society as a whole should work to increase their knowledge of AI technology. By understanding how AI makes decisions, stakeholders will be better equipped to hold organizations accountable for AI decisions. As AI is used in high-stakes decision-making, these decisions can have societal consequences if left unchecked. Together, governments and society can work to ensure AI is being used responsibly.

5.5 Limitations and Future Research

This review has several limitations. For one, it only incorporates academic journal articles from 2020 to 2025. While this was done to ensure collection of recent data on the topic, it is possible that relevant books and older articles were excluded. In addition, this review was conducted using secondary data, which limits its originality. The scope of this review was also limited by the interdisciplinary nature of AI as various organizations and researchers may use the terminology differently.

For future research, it would be beneficial to test these relationships through empirical studies. Since this review found a link between responsible AI governance and trust, future research should aim to confirm this relationship. In addition, future research should look at how these factors influence trust in specific sectors. Since different sectors have different trust implications, future research should focus on a particular sector.

5.6 Final Insight

AI will not succeed or fail based on its capacity to make faster, data-driven decisions. Rather, AI will succeed or fail based on whether it is trustworthy. AI may make optimal decisions, but only you can make decisions that stakeholders will trust.

References

1. Afroogh, S., et al. (2024) 'Trust in artificial intelligence: Progress, challenges, and future directions', *Humanities and Social Sciences Communications*, 11(1), pp. 1–15.
2. Balasubramaniam, S., Kauppinen, M. and Hiekkanen, K. (2023) 'Transparency and explainability of AI systems: From ethical guidelines to requirements', *Information and Software Technology*, 155, 107–121.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

3. Batool, A., Gill, S.S., Rao, A.S. and Ouda, A. (2025) 'AI governance: A systematic literature review', *AI and Ethics*. doi:10.1007/s43681-024-00653-w.
4. Blanco, A. (2025) 'Human trust in artificial intelligence: Beyond reliance', *AI & Society*. doi:10.1007/s00146-025-01789-3.
5. Chang, V. (2025) 'Explainable artificial intelligence (XAI) in business management research: A systematic review', *Technological Forecasting and Social Change*, 200, 123456.
6. Chaudhary, S., et al. (2025) 'Deploying explainable AI in entrepreneurial organizations: A conceptual framework', *Journal of Business Research*, 180, 114–126.
7. Chen, L. and Wen, Z. (2021) 'Impacts of attitudes toward government and corporations on public trust in artificial intelligence', *Computers in Human Behavior*, 122, 106–121.
8. Cetinkaya, A. and Krämer, N.C. (2025) 'Between transparency and trust: Identifying key factors in AI system perception', *Computers in Human Behavior*, 158, 107–118.
9. Dahlin, E. (2025) 'Trust in AI: A philosophical and practical perspective', *AI & Society*, 40(2), pp. 233–245.
10. Daly, M., Robinson, E. and Sutin, A.R. (2025) 'Shifting attitudes and trust in artificial intelligence: Influences on organizational adoption', *Computers in Human Behavior*, 160, 107–122.
11. Do Khac, L.T., et al. (2025) 'Towards an integrative model of organizational human–AI trust', *Technology in Society*, 72, 102–118.
12. Dwivedi, Y.K., Hughes, L., Ismagilova, E., et al. (2021) 'Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research', *International Journal of Information Management*, 57, 101994.
13. Figueroa-Armijos, M., Fink, J. and Kluemper, D. (2023) 'Ethical perceptions of artificial intelligence in hiring and organizational trust', *Journal of Business Ethics*, 186(1), pp. 123–139.
14. Floridi, L., Cowls, J., Beltrametti, M., et al. (2021) 'An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations', *Nature Machine Intelligence*, 3(1), pp. 1–9.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

15. Glikson, E. and Woolley, A.W. (2020) 'Human trust in artificial intelligence: Review of empirical research', *Academy of Management Annals*, 14(2), pp. 627–660.
16. Henrique, M., et al. (2024) 'Trust in artificial intelligence: A systematic literature review and research agenda', *Journal of Business Research*, 165, 113–130.
17. Jackson, K. (2023) 'Trust or mistrust in algorithmic grading? A study of fairness perceptions', *Assessment & Evaluation in Higher Education*, 48(6), pp. 875–889.
18. Jobin, A., Ienca, M. and Vayena, E. (2019) 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence*, 1(9), pp. 389–399.
19. Lahusen, C., et al. (2024) 'Trust, trustworthiness and artificial intelligence governance', *AI & Society*. doi:10.1007/s00146-024-01678-2.
20. Lee, J., Kim, H. and Park, S. (2025) 'Integrating ESG and AI: A responsible AI framework for organizations', *Business Strategy and the Environment*, 34(2), pp. 456–472.
21. Mehrabi, N., Morstatter, F., Saxena, N., et al. (2021) 'A survey on bias and fairness in machine learning', *ACM Computing Surveys*, 54(6), pp. 1–35.
22. Mikalef, P., Krogstie, J., Pappas, I.O. and Pavlou, P. (2021) 'Exploring the relationship between big data analytics capability and competitive performance', *Information & Management*, 58(2), 103–112.
23. Papagiannidis, E., Mikalef, P. and Conboy, K. (2025) 'Responsible artificial intelligence governance: A review and research framework', *International Journal of Information Management*, 82, 102–118.
24. Phan, T., et al. (2025) 'AI, trust, and risk: Drivers of adoption in financial platforms', *Journal of Financial Technology*, 12(1), pp. 45–63.
25. Rai, A. (2020) 'Explainable AI: From black box to glass box', *Journal of the Academy of Marketing Science*, 48(1), pp. 137–141.
26. Ransbotham, S., Kiron, D., Gerbert, P. and Reeves, M. (2020) 'Expanding AI's impact with organizational learning', *MIT Sloan Management Review*, 61(4), pp. 1–10.
27. Reinhardt, K. (2023) 'Trustworthy artificial intelligence: The role of governance in organizations', *AI and Ethics*, 3(2), pp. 345–356.

CC BY

JIRRS
Journal International Review of Research Studies
Received: 12/07/2026 - **Accepted:** 08/09/2026
Vol: 01.10
DOI: 10.66104/5erm9b42
Pages: 1-23

28. Sadeghi, A., et al. (2024) 'Explainable artificial intelligence and agile decision-making', *Decision Support Systems*, 175, 114–129.
29. Schilke, O. and Reimann, M. (2025) 'The transparency dilemma: How AI disclosure erodes trust', *Journal of Marketing*, 89(1), pp. 45–62.
30. Schlicker, N., et al. (2025) 'How do we assess the trustworthiness of AI? Introducing the TrAM model', *AI & Society*, 40(3), pp. 567–582.
31. Shin, D. (2021) 'The effects of explainability and causability on perception, trust, and acceptance of AI', *Computers in Human Behavior*, 118, 106–118.
32. Siau, K. and Wang, W. (2020) 'Artificial intelligence (AI) ethics: Ethics of AI and ethical AI', *Journal of Database Management*, 31(2), pp. 74–87.
33. Vanneste, B., Puranam, P. and Kretschmer, T. (2024) 'Artificial intelligence, trust, and perceptions of agency', *Organization Science*, 35(1), pp. 123–140.
34. Vassilakopoulou, P., et al. (2022) 'Responsible AI: Concepts, critical perspectives, and an IS research agenda', *Information Systems Journal*, 32(4), pp. 708–736.
35. Vining, A., et al. (2025) 'Trustworthy artificial intelligence and organizational trust: A scoping review', *Information and Organization*, 35(1), 100–120.
36. Wamba, S.F., Gunasekaran, A., Akter, S., et al. (2021) 'Big data analytics and firm performance', *International Journal of Production Economics*, 232, 107–118.