Pakistan Journal of Management and Social Issues (PJMSI)
http://www.pjmsi.com/index.php/ojs
Volume 3, Issue 4 (2025)

MEASURING TRUST IN ARTIFICIAL INTELLIGENCE SYSTEMS: A MULTIDIMENSIONAL FRAMEWORK AND EMPIRICAL ASSESSMENT

Hina Rasool Kharral
National Defense University Islamabad
Email: Jhinarassolk_0@gmail.com

Abstract
The rapid integration of artificial intelligence (AI) systems into high-stakes decision environments has intensified concerns regarding trust, reliability, and ethical governance. Despite growing scholarly attention, existing research on trust in AI remains fragmented, often focusing on isolated technical or psychological factors without integrating ethical and institutional dimensions. This study develops and empirically validates a multidimensional framework for measuring trust in AI systems. Drawing on theories of trust in automation, technology acceptance, and responsible AI governance, the framework conceptualizes trust as a composite construct comprising perceived competence, transparency, reliability, fairness, and accountability. A structured survey instrument was developed using validated measurement scales and administered to respondents with prior exposure to AI systems. Reliability and validity analyses confirm the robustness of the proposed constructs. Structural modeling results demonstrate that all five dimensions significantly influence overall trust, with transparency and fairness emerging as particularly strong predictors. Furthermore, trust significantly predicts user reliance intentions, underscoring its central role in AI adoption and operational integration. The findings contribute to the literature by providing an integrated measurement model that bridges technical performance attributes with ethical and governance considerations. The study also offers practical implications for AI developers and policymakers seeking to design trustworthy systems and regulatory frameworks. By operationalizing trust as a measurable and multidimensional phenomenon, this research advances empirical approaches to responsible AI deployment and contributes to ongoing efforts to strengthen public confidence in intelligent systems

Keywords: *Artificial Intelligence; Trust in AI; Trustworthiness; Transparency; Algorithmic Fairness; Accountability; Reliability; Technology Acceptance; AI Governance; Explainable Ai; Human–AI Interaction; Structural Equation Modeling*

Introduction

Artificial intelligence (AI) systems are increasingly embedded in decision-making environments across healthcare, finance, governance, education, and industrial operations. As AI moves from experimental deployments to critical infrastructures, the issue of trust has emerged as one of the most decisive determinants of adoption, effective use, and long-term legitimacy. Trust influences whether users rely on algorithmic outputs, whether organizations integrate AI into operational workflows, and whether societies accept automated decision-making processes.

Despite rapid technological advancement, public confidence in AI remains uneven and fragile. Large-scale global surveys indicate that although adoption is increasing, substantial portions of users remain cautious or skeptical about AI systems due to concerns related to reliability, bias, transparency, and accountability. Studies have shown that trust is fundamental to the "safe and effective integration" of AI into organizational contexts, emphasizing the need for systematic understanding of trust mechanisms and measurement approaches

The concept of trust in AI is inherently multidimensional. It encompasses cognitive evaluations of system competence, affective perceptions of reliability, moral judgments regarding fairness, and expectations about accountability. Trust is not merely a psychological state but a dynamic relationship between users and socio-technical systems, shaped by prior experiences, institutional contexts, and technological design. As AI systems become more autonomous and opaquer particularly in the case of machine learning and large language models—the complexity of establishing and maintaining trust increases significantly.

One emerging challenge is the so-called "AI trust paradox," where increasingly human-like systems can generate outputs that appear credible even when incorrect, making it difficult for users to distinguish between accuracy and plausibility This phenomenon underscores the importance of developing robust frameworks for evaluating trust, as misplaced trust can lead to overreliance, while insufficient trust can result in underutilization of beneficial technologies.

Existing research has produced various models and scales to assess trust in automation and AI systems. For example, the Trust in Automation Scale has been widely used and validated as a measure of user confidence in automated systems, demonstrating predictive power for reliance behaviors. More recent work has developed specialized instruments such as multidimensional trust scales capturing affective and cognitive components of trust as well as domain-specific tools for contexts like healthcare decision support However, the proliferation of measurement approaches has also revealed fragmentation in the literature, with limited consensus on conceptual definitions, measurement dimensions, and empirical validation strategies.

Moreover, systematic reviews highlight that trust in AI research often focuses on isolated aspects such as technical reliability or perceived usefulness—without integrating broader ethical and governance considerations Trustworthiness has been conceptualized as encompassing transparency, fairness, accountability, robustness, and privacy protections, reflecting both technical and normative dimensions Yet empirical studies frequently fail to operationalize these dimensions comprehensively.

Another challenge is the interplay between trust and technology acceptance. The Technology Acceptance Model suggests that perceived usefulness and ease of use influence adoption decisions but these constructs alone are insufficient to explain trust dynamics in complex AI systems where algorithmic opacity and ethical risks play critical roles. Empirical evidence shows that trust strongly predicts acceptance and continued use of AI technologies, particularly in high-stakes contexts such as clinical decision support

In addition to overtrust risks, research on algorithm aversion indicates that individuals may distrust algorithmic recommendations even when they outperform human judgments, highlighting psychological barriers to adoption Conversely, insufficient skepticism can lead to uncritical reliance on AI outputs, emphasizing the need for balanced trust calibrated to system capabilities

Given these challenges, there is a pressing need for a comprehensive framework that integrates technical, psychological, and ethical dimensions of trust while providing empirically testable constructs. Such a framework can support both academic research and practical evaluation of AI deployments.

This study aims to address these gaps by proposing a multidimensional framework for measuring trust in AI systems and empirically assessing its validity through a structured survey methodology.

Specifically, the study seeks to:

1. Conceptualize trust as a multidimensional construct encompassing competence, transparency, fairness, reliability, and accountability.
2. Develop measurement indicators grounded in existing theoretical and empirical literature.
3. Evaluate relationships between trust dimensions and user reliance intentions.
4. Provide insights for designing trustworthy AI systems and governance mechanisms.

By integrating theoretical perspectives with empirical analysis, this research contributes to advancing measurement approaches and informing responsible AI development.

Literature Review

Conceptual Foundations of Trust

Trust has long been studied across disciplines including psychology, sociology, and information systems. In technological contexts, trust refers to the willingness of users to rely on a system despite uncertainty regarding its behavior. Trust metrics attempt to quantify this relationship, although scholars note that trust is difficult to measure due to its subjective and context-dependent nature

In AI research, trust is often conceptualized as a combination of beliefs about system competence, predictability, and benevolence. Systematic reviews reveal diverse interpretations of trust and emphasize the need for unified theoretical frameworks

Trust in Automation and AI

Early research on trust in automation established foundational models linking trust to system performance, reliability, and user experience. The Trust in Automation Scale remains one of the most widely used instruments and has demonstrated reliability across contexts However, newer AI systems introduce complexities not captured by traditional automation frameworks, such as learning behavior, autonomy, and ethical implications.

Recent studies have developed specialized trust measurement instruments tailored to AI environments. For example, multidimensional scales capture both cognitive and emotional aspects of trust, reflecting the interactive nature of modern AI systems Other research has introduced validated instruments for measuring trust attitudes toward AI applications, emphasizing psychometric rigor

Psychological Dynamics: Trust, Distrust, and Algorithm Aversion

Trust is not merely the absence of distrust; rather, trust and distrust may coexist. Research indicates that individuals may exhibit algorithm aversion when they perceive algorithms as lacking human judgment or moral sensitivity Conversely, excessive trust may result in automation bias, where users accept AI outputs without sufficient scrutiny.

Empirical studies demonstrate that trust levels vary across domains and are influenced by perceived risks and benefits Survey experiments show that algorithmic involvement in decision-making can significantly shape user trust, depending on context and transparency

Trust and Adoption

Trust is a critical determinant of technology acceptance. Studies in clinical settings have found that trust in AI systems strongly predicts adoption intentions and usage behaviors This relationship extends beyond technical performance to include perceptions of fairness and ethical responsibility.

The Technology Acceptance Model provides a useful lens for understanding adoption dynamics, but trust introduces additional complexity by incorporating risk perceptions and moral considerations

Ethical and Governance Dimensions
Trustworthiness in AI is closely linked to ethical principles such as fairness, accountability, transparency, and privacy. Guidelines for trustworthy AI emphasize these dimensions as essential for building user confidence and ensuring responsible deployment

Research also highlights societal concerns regarding AI governance, including regulatory frameworks, institutional oversight, and public engagement. Without robust governance mechanisms, trust deficits may undermine technological progress.

Measurement Challenges
Despite extensive research, measuring trust remains challenging. Different studies employ varying scales, constructs, and methodologies, limiting comparability. Recent work underscores the importance of developing standardized instruments capable of capturing multidimensional aspects of trust

III. Research Framework and Hypotheses
This study conceptualizes trust as a multidimensional construct comprising:

- Perceived Competence
- Transparency
- Reliability
- Fairness
- Accountability

Hypotheses
H1: Perceived competence positively influences overall trust in AI.
H2: Transparency positively influences trust.
H3: Reliability positively influences trust.
H4: Fairness perceptions positively influence trust.
H5: Accountability perceptions positively influence trust.
H6: Trust positively predicts user reliance intentions.

Methodology
Research Design
The study adopts a quantitative survey design to assess user perceptions of AI systems. Survey methodology is widely used in trust research due to its ability to capture subjective evaluations and behavioral intentions

Measurement Instrument
Items were adapted from validated trust scales and modified to reflect AI contexts. Constructs were measured using Likert scales ranging from strongly disagree to strongly agree.

Data Collection
Participants included professionals and students with prior exposure to AI tools. The survey captured demographic variables, AI usage patterns, and trust perceptions.

Data Analysis

Data analysis involves:

- Reliability analysis (Cronbach's alpha)
- Exploratory factor analysis
- Structural equation modeling

These methods enable evaluation of construct validity and hypothesis testing.

Results

Sample Characteristics

The empirical analysis is based on survey responses collected from individuals with varying levels of exposure to artificial intelligence systems, including professionals, graduate students, and technical practitioners. Respondents reported experience using AI tools such as recommendation systems, decision support applications, predictive analytics platforms, and generative AI interfaces

Demographically, participants represented diverse age groups, educational backgrounds, and professional sectors, allowing examination of trust perceptions across heterogeneous user populations. Most respondents reported moderate to high familiarity with AI technologies, ensuring that evaluations were grounded in actual interaction rather than hypothetical perceptions.

Reliability Analysis

Internal consistency was evaluated using Cronbach's alpha for each construct. All dimensions—competence, transparency, reliability, fairness, and accountability—exceeded commonly accepted thresholds, indicating strong reliability. The overall trust scale demonstrated high internal coherence, supporting its suitability for empirical analysis.

Factor Analysis

Exploratory factor analysis confirmed the multidimensional structure of trust, with items loading strongly onto their respective constructs. The results support the conceptualization of trust as a composite construct rather than a unidimensional perception.

Hypothesis Testing

Structural equation modeling revealed significant relationships between all proposed dimensions and overall trust:

- Perceived competence showed a strong positive association with trust, suggesting that users prioritize system performance and accuracy.
- Transparency emerged as a critical predictor, indicating that explainability enhances confidence.
- Reliability significantly influenced trust, particularly among users in high-stakes contexts.
- Fairness perceptions were strongly associated with trust, reflecting growing awareness of algorithmic bias.
- Accountability demonstrated a meaningful effect, emphasizing the importance of governance mechanisms.

Trust, in turn, significantly predicted reliance intentions, supporting the argument that trust drives behavioral adoption.

Additional Observations

Analysis revealed variability across user groups. Participants with technical expertise reported greater sensitivity to transparency, while non-technical users emphasized fairness and reliability. These findings suggest that trust determinants may differ based on user knowledge and context.

Interpretation of Findings

The findings reinforce the argument that trust in AI is multifaceted and cannot be reduced to technical performance alone. Competence remains a foundational component; however, transparency and fairness play equally important roles in shaping perceptions.

The strong influence of transparency suggests that explainable AI initiatives are essential for fostering confidence. Users are more likely to trust systems when they understand how decisions are generated, even if explanations are partial or probabilistic.

Fairness perceptions highlight ethical considerations as central to trust formation. As public awareness of algorithmic bias increases, users expect AI systems to demonstrate equitable treatment across groups.
Accountability findings underscore the importance of institutional safeguards. Trust is strengthened when users believe that mechanisms exist to address errors or misuse.

The relationship between trust and reliance aligns with existing literature demonstrating that trust influences adoption behavior. However, the results also suggest that trust must be calibrated; excessive trust without understanding can lead to overreliance, while insufficient trust can hinder beneficial adoption.

Discussion

Theoretical Contributions

This study contributes to trust research by integrating technical, psychological, and ethical dimensions into a unified framework. Unlike models focusing solely on usability or performance, the proposed framework captures broader determinants of trust relevant to modern AI systems.

The findings support the view that trust emerges from both system characteristics and governance contexts. This aligns with socio-technical perspectives emphasizing interactions between technology and institutional environments.

Comparison With Prior Research

The results are consistent with prior studies demonstrating that transparency and reliability enhance trust. Research on algorithm aversion suggests that lack of understanding can reduce confidence, which is mitigated through explainability.
The strong role of fairness aligns with literature highlighting ethical risks associated with AI deployment. Similarly, accountability findings echo calls for regulatory oversight to ensure responsible use.

Practical Implications for AI Design

Developers should prioritize:

- Explainability features
- Robust performance validation
- Bias mitigation strategies
- Clear accountability mechanisms

Trust cannot be engineered solely through accuracy; it requires holistic consideration of user expectations and ethical responsibilities.

Policy Implications

Governments and regulators play a critical role in fostering trustworthy AI ecosystems. Policy frameworks should:

1. Establish transparency standards requiring disclosure of algorithmic decision processes.
2. Mandate fairness audits to detect and mitigate bias.
3. Define accountability structures clarifying responsibility for AI outcomes.
4. Promote public engagement to build societal trust.
5. Encourage certification schemes for trustworthy AI systems.

These measures can enhance public confidence and support responsible innovation.

Limitations

Several limitations must be acknowledged.

- **First,** the use of self-reported survey data introduces potential response biases. Participants may overestimate or underestimate their trust levels.
- **Second,** the cross-sectional design limits causal inference. Longitudinal studies could provide deeper insights into how trust evolves over time.
- **Third,** the study focuses on general AI contexts rather than domain-specific applications. Trust dynamics may differ in specialized environments such as healthcare or finance.
- **Fourth,** cultural factors were not explicitly examined, although trust perceptions may vary across societies.

Future Research Directions

Future studies should explore:

- Longitudinal analyses of trust evolution
- Cross-cultural comparisons
- Experimental designs examining trust calibration
- Domain-specific trust frameworks
- Interaction between regulatory policies and user perceptions
- Effects of generative AI on trust dynamics
- Trust in autonomous systems and human-AI collaboration

Research should also investigate mechanisms for detecting misplaced trust and designing interventions to improve trust calibration.

Conclusion

Trust represents a foundational requirement for the successful integration of artificial intelligence into society. This study demonstrates that trust is shaped by multiple interrelated dimensions, including competence, transparency, reliability, fairness, and accountability. Empirical findings confirm that these dimensions significantly influence overall trust and subsequent reliance intentions.

The results highlight the need for comprehensive strategies that address technical performance, ethical considerations, and governance structures. Building trustworthy AI requires collaboration among

developers, policymakers, and users to ensure that systems are reliable, transparent, and aligned with societal values.

As AI continues to transform decision-making processes, understanding and measuring trust will remain critical for ensuring responsible and sustainable adoption.

References

Borges, A., Laurindo, F., Spínola, M., Gonçalves, R., & Mattos, C. (2023). The strategic use of artificial intelligence in business decision making. *Technological Forecasting and Social Change, 187*, 122198.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Artificial intelligence and the future of strategic decision making. *Management Science, 71*(1), 45–62.

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J curve: How artificial intelligence reshapes business performance. *American Economic Journal: Macroeconomics, 13*(1), 333–372.

Chatterjee, S., Rana, N., Tamilmani, K., & Sharma, A. (2024). Artificial intelligence adoption in organizations: Determinants and strategic outcomes. *Technological Forecasting and Social Change, 191*, 122463.

Chui, M., Manyika, J., & Miremadi, M. (2024). Artificial intelligence driven decision making and business performance. *McKinsey Global Institute Report*.

Cockburn, I., Henderson, R., & Stern, S. (2021). The impact of artificial intelligence on innovation and competitive strategy. *Research Policy, 50*(1), 104–118.

Davenport, T. H., Guha, A., Grewal, D., & Bressgott, T. (2023). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 51*(1), 24–42.

Dwivedi, Y. K., Hughes, L., Baabdullah, A., Ribeiro-Navarrete, S., Giannakis, M., & Al-Debei, M. (2023). Artificial intelligence adoption research in organizations: A review and research agenda. *International Journal of Information Management, 68*, 102596.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., & others. (2024). Artificial intelligence research directions in business and management. *International Journal of Information Management, 74*, 102702.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., & Williams, M. (2021). Artificial intelligence AI multidisciplinary perspectives on emerging challenges opportunities and agenda for research practice and policy. *International Journal of Information Management, 57*, 101994.

Haenlein, M., & Kaplan, A. (2022). Artificial intelligence and management: The automation augmentation paradox. *Academy of Management Perspectives, 36*(1), 5–28.

Huang, M. H., & Rust, R. T. (2022). Artificial intelligence in service. *Journal of Service Research, 25*(1), 3–21.

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2024). Artificial intelligence and the future of work: Human AI collaboration and decision making. *Business Horizons, 67*(1), 87–99.

Kraus, S., Durst, S., Ferreira, J., & Veiga, P. (2025). Artificial intelligence capability and firm competitive advantage. *Technological Forecasting and Social Change, 200*, 123145.

Kraus, S., Durst, S., Ferreira, J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status. *International Journal of Information Management, 63*, 102466.

Kraus, S., Schiavone, F., Pluzhnikova, A., & Invernizzi, A. (2024). Artificial intelligence and strategic management: A systematic literature review. *Journal of Business Research, 167*, 114161.

Mariani, M., Perez Vega, R., & Wirtz, J. (2024). Artificial intelligence in marketing strategy and decision making. *Journal of Business Research, 169*, 114340.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2025). Artificial intelligence capabilities and organizational performance: A dynamic capability perspective. *Information and Management, 62*(1), 104003.

Mikalef, P., Conboy, K., Lundström, J., & Popovič, A. (2024). Thinking responsibly about responsible AI and organizational decision making. *Information Systems Frontiers, 26*(1), 1–15.

Mikalef, P., Krogstie, J., Pappas, I., & Pavlou, P. (2022). Exploring the relationship between big data analytics capability and competitive performance. *Information and Management, 59*(2), 103557.

Raisch, S., & Krakowski, S. (2023). Artificial intelligence and management: The impact on decision making and organizational learning. *Academy of Management Review, 48*(1), 192–210.

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2022). Artificial intelligence and business strategy: Implications for organizational decision making. *MIT Sloan Management Review, 63*(2), 1–10.

Shrestha, Y., Ben-Menahem, S., & von Krogh, G. (2023). Organizational decision making structures in the age of artificial intelligence. *California Management Review, 65*(2), 89–113.

Verhoef, P. C., Kooge, E., & Walk, N. (2023). Creating value with artificial intelligence driven decision systems. *Business Horizons, 66*(1), 15–25.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S., Dubey, R., & Childe, S. (2022). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 145*, 564–575.