Journal of Data Analytics and Intelligent Decision-making

Online ISSN: 3115-8161

Journal Homepage: jdaid.qom.ac.ir

University of Qom

Building Trust in Artificial Intelligence Deployment: Designing a Trustworthy Artificial Intelligence Framework Through a Mixed Method Study

Nazanin Mohammadi Haroniyeh[1], Mona Jami Pour[2]✉

1. MA of E-commerce, Management Department, Faculty of Management and Accounting, Hazrat-e Masoumeh University, Qom, Iran. Email: Nazanin.haroniyeh2438@gmail.com
2. Corresponding author, Associate Professor, Management Department, Faculty of Management and Accounting, Hazrat-e Masoumeh University, Qom, Iran. Email: mjami@hmu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 8 July 2026
Received in revised form 30 August 2026
Accepted 15 September 2026
Available online 30 September 2026

Keywords:
artificial intelligence (AI),
trustworthy AI,
user trust,
best–worst method (BWM)

ABSTRACT

With the rapid expansion of artificial intelligence (AI) applications in businesses, the issue of trust in this technology has become one of the fundamental challenges in its adoption and effective utilization. A lack of sufficient trust in AI systems can lead to reduced adoption, increased risk, and weakened competitive advantage in businesses. The present study aims to provide a comprehensive and localized framework for realizing trustworthy AI in Iranian businesses. To achieve the research objective, a mixed-methods approach was adopted. In the first phase, a comprehensive literature review, together with a qualitative approach, was conducted to identify the criteria and components of trustworthy AI. During the qualitative phase, data were collected through semi-structured interviews with experts and analyzed using thematic analysis. In the second (quantitative) phase, the Best-Worst Method (BWM) was employed to prioritize and assign weights to the five main criteria and 35 identified sub-criteria. The BWM calculations were performed using LINGO software to determine the final weights. The findings showed that achieving trustworthy AI in businesses requires simultaneous attention to technical and operational, human, ethical, legal, and social impact dimensions, among which the technical and operational aspects are the most important. These results indicate that despite the rapid growth of AI in businesses, there is still no practical, localized framework in Iran, considering all these aspects in an integrated manner.

Cite this article: Mohammadi Haroniyeh, N., & Jami Pour, M. (2026). Building trust in artificial intelligence deployment: Designing a trustworthy artificial intelligence framework through a mixed method study. *Journal of Data Analytics and Intelligent Decision-Making (JDAID)*, *2*(3), 212-239.
https://doi.org/10.22091/jdaid.2026.16682.1087

© Author(s) retain the copyright. **Publisher**: University of Qom.
DOI: https://doi.org/10.22091/jdaid.2026.16682.1087

Introduction

AI has emerged as one of the primary drivers of the Fourth Industrial Revolution, fundamentally transforming human-technology interactions across a wide range of sectors, including healthcare, financial services, autonomous transportation, tourism, education, and e-commerce (Dang & Li, 2026; Sarun et al., 2026). According to Gartner (2026), global spending on AI is projected to reach USD 2.59 trillion in 2026, representing a 47% year-over-year increase. This rapid growth reflects the expanding adoption of AI technologies across industries and their increasing role in organizational operations. Organizations are increasingly integrating AI into their decision-making processes (Miedema et al., 2026). These developments have created significant opportunities, including enhanced productivity, improved service quality, and more efficient decision-making. However, they have also introduced critical challenges, such as algorithmic bias, the lack of transparency and accountability, privacy concerns, misinformation, and distributional inequalities (Sarun et al., 2026). Furthermore, the widespread deployment of AI systems has raised important concerns regarding governance, ethical considerations, and user trust (Ferdaus et al., 2026).

To gain the trust of users, AI systems must be safe, respectful, and accountable, and transparency in their performance and decisions should enable timely review and evaluation (Smith, 2019). As one of the fundamental concepts of human relations, trust underlies many social, economic, and technological interactions and plays a key role in the acceptance of new technologies (Ryan et al., 2020). Trust is a multidimensional concept that has been decomposed in various studies into factors such as usefulness, reliability, accuracy, understandability, and ease of use. Research shows that users' trust in AI systems is based on their predictability, safety, and efficiency, and the system's competence in performing tasks correctly plays a decisive role in this process (Ashoori & Weiszi, 2019). Trust in AI is not a static phenomenon, but rather evolves with increasing user experience. In this direction, factors such as reliability, predictability, and ethical alignment are considered as crucial elements. On the other hand, challenges such as algorithmic bias, data protection, and fairness, if overlooked, can hinder the widespread adoption of AI in various industries (Regona et al., 2026). Despite the high potential of AI to create business value, many organizations still face challenges in effectively applying it in their operations (Enholm et al., 2022).

In response to these challenges, researchers have proposed several theoretical frameworks to explain the dimensions of trustworthy AI. This situation shows that the development and deployment of AI is not simply a technical issue but also requires attention to ethical, social and organizational dimensions. Therefore, ensuring the responsible design and development of AI systems has become a fundamental necessity (Baldassarre et al., 2024).

Trust is a complex and multidimensional phenomenon that has been the subject of research in various disciplines, including psychology, sociology, economics, management, and computer science. In the classical definition, trust refers to an individual's willingness to rely on another party in conditions of uncertainty and lack of complete control (Thiebes et al., 2021). In the context of AI, trust plays a pivotal role in the acceptance, adoption, and continued deployment of these systems (Regona et al., 2026; Jang, 2026). As AI applications continue to

expand into sensitive domains, increasing attention has been devoted to the trustworthiness and reliability of these systems, which depends not only on the accuracy of their outputs but also on the robustness of their models, the interpretability of their decisions, and the mitigation of data bias (Mortaji & Sadeghi, 2024). Trustworthy AI is based on the idea that trust is the foundation of societies, economies, and sustainable development; only when trust is established can the full potential of AI be realized (Díaz-Rodríguez et al., 2023). This approach proposes a set of principles and requirements for the design, development, and deployment of these systems, with the aim of reducing or preventing AI risks (Kinney et al., 2024). In this regard, several studies have provided frameworks to explain the dimensions of trustworthy AI, including transparency, accountability, fairness, privacy, robustness, and reliability (Regona et al., 2026). Additionally, with the expansion of AI-based products, regulatory bodies have tried to help increase public trust by developing regulations (Nasr-Azadani & Chatelain, 2024).

Trust in and the trustworthiness of AI have become central issues in the AI literature due to the opacity and "black-box" nature of AI decision-making processes (Dang & Li, 2026). Consequently, a growing body of research has investigated the concept of trustworthy AI, with individual studies focusing on specific dimensions of this multifaceted concept. For instance, Chander et al. (2025) primarily emphasize the dimensions of explainability and interpretability and robustness in trustworthy AI, focusing on how these characteristics can be achieved rather than on how they can be operationally assessed or measured. Similarly, Kaur et al. (2022) argue that the increasing demand for trustworthy AI has led to the emergence of numerous methods and frameworks, each addressing a particular stage or aspect of the AI lifecycle. While some frameworks focus on the design phase of AI systems, others emphasize data collection, protection, and preprocessing. A separate group concentrates on model development and evaluation (Indukuri et al., 2026), whereas others address deployment, implementation, and post-deployment monitoring by incorporating auditing and testing mechanisms to ensure accountability and system reliability. Although these studies have advanced the field, they also highlight the need for a more holistic approach to trustworthy AI (Kaur et al., 2022). From another perspective, Li et al. (2024) classify the determinants of trustworthy AI into three categories: the trustor, the trustee, and the interactive context.

Despite these valuable contributions, the existing literature remains fragmented and often reports inconsistent findings. Most studies remain at the conceptual or theoretical level and provide limited operational guidance for measuring trustworthy AI. Consequently, the literature has yet to offer a comprehensive and integrated understanding of the dimensions underlying trustworthiness in AI systems (Dang & Li, 2026). Furthermore, despite the substantial growth of research on trustworthy AI (Vyhmeister & Castañe, 2026), a comprehensive framework that integrates the multidimensional nature of trustworthy AI within organizational contexts is still lacking.

This gap is particularly important in light of the rapid expansion of AI adoption in businesses. According to Statista (2026), the market for AI development tools in Iran is expected to experience significant growth in the coming years, driven by the increasing organizational demand for automation, data analytics, and AI model development. As AI technologies become increasingly embedded across industries, the trustworthiness of AI systems is no longer merely

a technical requirement but has become a critical determinant of organizational effectiveness and business success (Nastoska et al., 2025). Moreover, as Miedema et al. (2026) argue, trustworthiness is inherently context-dependent and is shaped by the social, organizational, regulatory, technological, and political environments in which AI systems are developed, deployed, and used. Accordingly, understanding Iranian users' perceptions of trustworthy AI within the country's unique social, legal, technological, and cultural context can contribute to reducing the risks associated with AI adoption while addressing its ethical and human-related challenges in Iranian businesses.

Therefore, there remains a clear need to develop a comprehensive and context-sensitive framework for assessing trustworthy AI in Iranian businesses. Addressing this gap is essential for the successful adoption and deployment of transformative AI technologies, yet it has received limited attention in the existing literature. Accordingly, this study aims to address this theoretical gap by answering the following research questions:

Research Question 1 (RQ1): What are the key dimensions of trustworthy AI in business?

Research Question 2 (RQ2): What is the relative importance of the dimensions of trustworthy AI for business adoption?

Literature and Research Background

AI and Its Role in Business

AI affects every aspect of our lives. This technology has also revolutionized the business arena (Walia, 2024). The term AI broadly refers to computer systems capable of performing tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation (Machucho & Ortiz, 2025). Alenizi et al. (2023) identified several AI methods, including machine learning, deep learning, natural language processing, computer vision, and expert systems. According to Bozko et al. (2016), in today's modern business environment, the use of AI technology as a method of decision-making will be relevant. Usman (2022) also confirmed that AI is a digital innovation that offers a good transformational solution for products, services, and businesses and companies. AI has proven to be a versatile technology to support multiple applications of business systems. It has been used to transform current business models and provide innovative management strategies (Singh & Chouhan, 2021). Melcher and Vaziri (2025) categorized the applications of AI in business, including chatbots, data analytics, image recognition, robotics, text recognition, pattern recognition, workflow automation, and other AI-enabled applications. Pallathadka et al. (2023) discussed several AI applications in business, including chatbots, image recognition, customer data management, recommendation systems, inventory management, cybersecurity, customer relationship management (CRM), sales, human resource management (HRM), credit scoring, loan underwriting, and portfolio management.

The Concept of Trustworthiness AI

Today, in the world of AI, trust remains a cornerstone of widespread adoption and integration across sectors. As AI systems become more complex and autonomous, understanding and building trust in these systems is crucial (Paliszkiewicz & Gołuchowski, 2024). The sustainable and effective adoption of AI in business depends on the trustworthiness of AI systems, which extends beyond technical performance to include robust data governance, model validation,

transparency, accountability, bias mitigation, and ethical oversight. In addition, addressing privacy, security, fairness, and explainability concerns while actively engaging stakeholders is essential for fostering user trust and acceptance (Westover, 2025). This is because, given the pervasiveness of these AI systems, breaches of stakeholder trust can lead to severe social consequences. Such breaches can range from biased behavior of automated systems in hiring and lending decisions to the loss of human lives (Li et al., 2023). Furthermore, as businesses use AI to address societal challenges and enhance public well-being, a commitment to ethical AI practices becomes essential to ensure positive and sustainable social impacts. The rapid integration of AI into business operations has ushered in a new era of efficiency, innovation, and productivity. However, this advancement comes with a range of ethical and trust challenges that require careful consideration (Maiti et al., 2025). Trustworthy AI is a term that accurately defines AI as robust, legal, and ethical. In other words, it fully describes the design, intent, concept, and conclusions of AI reaching their full potential when trust can be built at every stage of its lifecycle, from initial proposal to development and efficient positioning (Chander et al., 2025).

Given that AI systems are responsible for making decisions that can have fundamental consequences for both individuals and society, it is crucial that these systems are considered trustworthy by the people they affect. Interaction with AI can affect how users perceive and understand it, potentially affecting their level of trust (Eskandari et al., 2025; Oyekunle et al., 2024). In recent years, several international organizations, including the European Commission, IEEE, and the OECD, have introduced guidelines and principles for trustworthy AI to support the ethical, transparent, and responsible development and use of AI systems (Lewis et al., 2020).

Related Works

In recent years, several frameworks for trustworthy AI have been presented at the international level. These frameworks often emphasize ethical, technical, and legal principles. However, a review of the literature shows that most studies have limited themselves to providing a list of principles and have rarely addressed how to operationalize these principles in real business contexts. Furthermore, domestic studies in this area are limited and there is a significant research gap in providing indigenous models. In Table 1, key findings related to this topic from various articles are reviewed.

Table 1. Studies Related to Trustworthy AI in Business

Authors	Title	Findings
Miedema et al. (2026)	Towards trustworthy AI for decision-making: A lifecycle perspective on knowledge- and data-driven AI systems	Governance and oversight (accountability and security); Information and data management (accuracy, integrity, privacy and quality); System operation (availability, reliability, resilience, robustness and safety); Human interaction (authenticity, controllability, transparency, usability)
Ferdaus et al. (2026)	Towards trustworthy AI: A review of ethical and robust large language models	Toxicity and safety; Bias and fairness; Robustness and security; Explainability; Privacy; Emergent behavior; Temporal consistency
Chander et al. (2025)	Towards trustworthy AI in the context of explainability and robustness	Transparency; Technical robustness and safety; Accountability; Privacy and data governance; Fairness, diversity, and nondiscrimination; Human agency and oversight; Societal and environmental well-being; Additional trustworthy requirement of AI

Table 1. Studies Related to Trustworthy AI in Business

Authors	Title	Findings
Kinney et al. (2024)	Expectation management in AI: A framework for understanding stakeholder trust and acceptance of AI systems	Human agency and oversight; Privacy and data governance; Transparency, explainability, and interpretability; Diversity, non-discrimination, fairness, and bias mitigation; Accountability; Technical robustness and safety; Societal and environmental wellbeing; Performance expectancy; Effort expectancy
Kaur et al. (2022)	Trustworthy AI: A review	Fairness; Explainability; Accountability; Privacy; and Acceptance
Kowald et al. (2024)	Creating and evaluating trustworthy AI: An overview and research challenges	Human agency and oversight; Fairness and non-discrimination; Transparency and explainability; Robustness and accuracy; Privacy and security; Accountability
Smith (2019)	Designing trustworthy AI: A framework for human-machine collaboration	Accountable to humans; Cognizant of speculative risks and benefits; Respectful and secure; Honest and usable
Hohma & Lütge (2023)	From trustworthy principles to trustworthy development process: The need and elements of trustworthy development of AI systems	The trustworthiness of the AI development process consists of transparency, reliability and consistency, objectivity, accountability, and auditability and intervenability
Li et al. (2024)	Developing trustworthy AI: Insights from research on interpersonal trust, human-automation, and human-AI	A three-dimension framework of trust in AI: Trustor, trustee, interactive context
Nastoska et al. (2025)	Evaluating trustworthiness in AI: Risks, metrics, and applications across industries	Fairness; Transparency; Privacy; Accountability; Security
Arrieta et al. (2019)	Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI	Technical factor; Understandable, reliable and responsible AI; Explainability; Accountability; Privacy; System performance (availability, reliability, flexibility, robustness and safety)
Tanveer (2026)	AI governance frameworks: ISO/IEC 42001, NIST AI RMF, and the EU AI Act	Human factor; The extent and type of monitoring should be determined through decision-making process: Human control, higher intervenability, stronger system review, and stop mechanisms
Kandikatla et al. (2025)	AI and human oversight: A risk-based framework for alignment	Human factors: AI applications do not require a level of human oversight. The extent and type of oversight should be decided based on the level of risk. The greater the consequences of an error or harm, the greater the need for human control. Higher intervenability
Lazaros et al. (2026)	Human-in-the-Loop artificial intelligence: A systematic review of concepts, methods, and applications	Human factors: Reliable AI, particularly in high-risk applications, where humans are truly effective, not just for show. Designing effective human-AI collaboration

A review of the literature on trustworthy AI reveals several important theoretical and methodological gaps.

First, the majority of existing studies discuss only a limited set of trustworthy or responsible AI requirements at a conceptual and theoretical level. Most lack an integrated framework operational criterion for measuring and implementing the multidimensional nature of trustworthy AI in business contexts.

For example, Kaur et al. (2022) conducted a comprehensive review of the trustworthy AI literature and conceptually synthesized its key requirements from different research

perspectives (summarized in Table 1). However, the study does not propose operational indicators for these dimensions or address how trustworthy AI can be measured in practice. Similarly, Smith (2019) provide a conceptual review of several characteristics of trustworthy AI (presented in Table 1); however, their proposed framework is not supported by a robust theoretical foundation or empirical validation. Similarly, Kowald et al. (2024) identify six key requirements of responsible AI but limit their contribution to conceptual definitions, focusing primarily on how these requirements can be achieved rather than on how they can be systematically evaluated or measured. In the same vein, Kinney et al. (2024) introduce a conceptual framework that is not grounded in a specific research methodology. Their empirical work is confined to assessing the perceived applicability of the framework among participants from the healthcare and education sectors, rather than developing a validated measurement framework for trustworthy AI.

Second, rather than concentrating on the dimensions and components of trustworthy AI itself, several studies have focused on related issues such as the lifecycle of trustworthy AI system development, evaluation processes, antecedents of trust, or risks that hinder the realization of trustworthy AI.

For instance, Hohma and Lütge (2023) emphasize the development process of trustworthy AI systems rather than the characteristics of the final AI system. Their work adopts a procedural perspective on trustworthiness by identifying desirable features of the AI development process to ensure that the resulting system is trustworthy. Similarly, Miedema et al. (2026) focus on the lifecycle of data-driven AI systems used in decision-making processes. Adopting a process-oriented perspective, they employ the ISO 2002 trustworthiness framework to assess AI trustworthiness and classify its dimensions into four broad categories. Nevertheless, their study remains descriptive, providing definitions and classifications of trustworthiness characteristics without developing an instrument for measuring trustworthy AI. In contrast, Ferdaus et al. (2026) propose a framework for evaluating the trustworthiness of large language models (LLMs), comparing systems such as Gemini, ChatGPT, DeepSeek, and Claude using a set of predefined trustworthiness indicators. Meanwhile, Nastoska et al. (2025) identify the major risks associated with different stages of the AI lifecycle and propose strategies to address challenges related to system reliability, bias mitigation, and model explainability. Likewise, Li et al. (2024) investigate the determinants of trust in AI rather than the dimensions of trustworthy AI itself. Their framework identifies three categories of antecedents—Trustor, Trustee, and Interactive Context—as the primary factors influencing users' trust in intelligent systems.

Third, as emphasized by Miedema et al. (2026), trustworthiness is not merely an objective and static characteristic but is largely shaped by users' perceptions and the contextual environment in which AI systems operate. Consequently, the meaning and evaluation of trustworthy AI may vary across different cultural, social, legal, and technological contexts. This highlights the necessity of developing a context-sensitive framework that reflects the cultural, regulatory, and business environment of Iranian organizations.

Fourth, despite the growing importance of sustainability and social impact—particularly for Generation Z, whose trust in AI systems is significantly influenced by these considerations (Dewalska-Opitek et al., 2024)—existing studies have devoted limited attention to the

sustainability and societal dimensions of trustworthy AI. As a result, these aspects remain largely underexplored in the literature.

In summary, the present study seeks to address these theoretical gaps by adopting a mixed-methods approach to develop and validate an integrated, context-specific framework for trustworthy AI in Iranian businesses.

This approach is expected to provide both a stronger theoretical foundation and a practical measurement framework for assessing trustworthy AI in organizations.

Research Methodology

Research Approach and Data Collection

To achieve the research objective, a mixed-methods approach was adopted, as illustrated in Figure 1. In the first phase, in addition to a comprehensive literature review, semi-structured interviews were conducted with experts to identify the dimensions and indicators of trustworthy AI. Relevant studies on trustworthy AI, trust in AI, and AI ethics were reviewed. Subsequently, to enrich and refine the extracted codes, semi-structured interviews with experts were conducted.

Purposive sampling was used to conduct interviews, and the interviews continued until theoretical saturation was achieved. The interviews were conducted over approximately one and a half months, from November 10, 2025, to December 24, 2025. The qualitative data were then analyzed using thematic analysis, which involved initial coding, the identification of subthemes, and the extraction of main themes.

In the second stage, the BWM, a multi-criteria decision-making technique, was employed to prioritize the identified dimensions and components of trustworthy AI in business. A BWM questionnaire was administered to five experts, purposively selected from the 14 interview participants based on their extensive experience, familiarity with the research topic, accessibility, and willingness to participate. The selected experts included senior business managers and AI specialists. During the first online session, they evaluated the dimensions and criteria using a nine-point rating scale. In the second session, they reviewed the pairwise comparison scores and reached a consensus. Final agreement on the best and worst criteria and their corresponding sub-criteria was achieved through these two online sessions.

To assess the reliability of the interview data, test-retest reliability and inter-coder agreement were employed. For the test-retest assessment, three interviews (P3, P5, and P12) were selected and independently coded by the researcher in two rounds with a 15-day interval between coding sessions. The codes obtained from the two rounds were then compared. Matching codes were considered agreements, whereas non-matching codes were considered disagreements. The percentage of test-retest reliability was calculated using the following formula:

$$\text{Percentage of intra-subject agreement} = \frac{2 \times Number\ agreements}{Total\ number\ of\ codes} \times \%\ 100$$

A total of 57 codes were recorded in the two rounds of coding, 19 of which were in agreement. Therefore, a 66.67% test-retest reliability was obtained. Given that this value exceeded the acceptable limit of 60%, the coding consistency was assessed as acceptable.

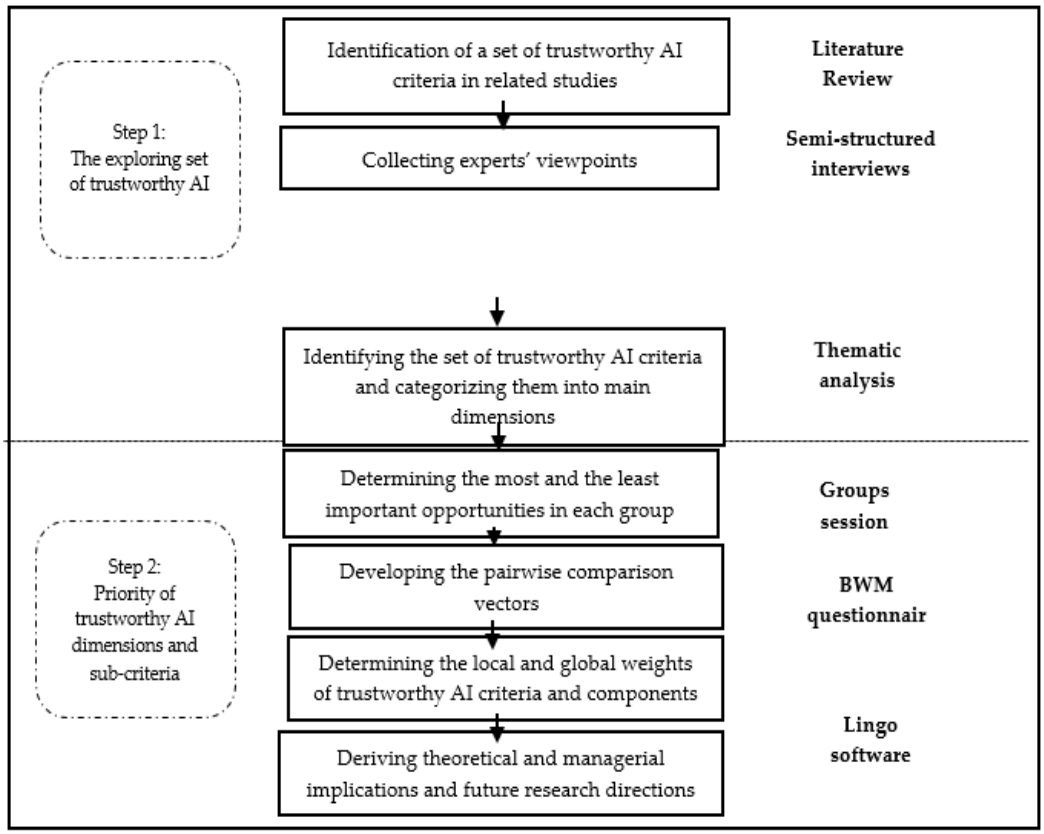

Figure 1. Research Steps

To examine the inter-coder reliability, three interviews with codes P3, P5, and P12 were selected from a total of 14 interviews. These three interviews were coded by an academic researcher familiar with the research topic. Then, the codes extracted by the two coders were compared with one another; similar codes were considered as agreement and dissimilar codes as disagreement. In the three selected interviews, a total of 45 codes were recorded by the two coders, 17 of which were in agreement. Accordingly, the agreement rate between the two coders was 75.56%. Therefore, the reported value was related to the coding of the three selected interviews and not the entire 14 interviews. Given that this value exceeded the acceptable limit of 60%, the reliability of the coding process was confirmed.

To assess the content validity of the interview guide, the initial set of interview questions was submitted to the supervisors and a panel of experts in the fields of AI and business. The questions were evaluated in terms of clarity, comprehensibility, relevance to the research objectives, the coverage of the research domain, and logical sequencing. Based on the feedback obtained from experts, the interview guide was revised accordingly, and the final version was developed. Therefore, the face and content validity of the interview guide was established qualitatively through expert judgment.

BWM

To rank the findings of the BWM presented by Rezaei (2015), the best and worst indicators are determined by the decision maker and a pairwise comparison is made between each of these two criteria (best and worst) and other indicators. Among the outstanding features of this method compared to other MCDM methods, we can mention the need for less comparative data and a more robust comparison. The steps of this method are:

Step 1: Determine a set of decision criteria: In this step, we determine a set of criteria, {Cn, ..., C2, C1}, that should be considered in decision making.

Step 2: Determine the best (most desirable/most important) and worst (least important/least desirable) indicators. No comparison will be made at this stage.

Step 3: The priority of the best criterion over the other criteria is determined using the numbers 1 to 9. The vector of the priority of the best criterion over the other criteria is represented in Equation (1):

$$A_B = \left(a_{B_1}, a_{B_2}, \dots, a_{B_n}\right) \qquad (1)$$

Where, a_{B_j} indicates the preference of the best criterion C_B over criterion C_j.

Step 4: Determine the priority of the other criteria over the worst criterion using the numbers 1 to 9. The vector of the priority of the other criteria over the worst criterion is represented as Equation (2):

$$A_W = (a_{1w}, a_{2w}, \dots, a_{nw}) \qquad (2)$$

In the above vector, a_{jW} is the priority of criterion (j) over the worst criterion (W).

Step 5: Find the optimal weights $(W_1^*, W_2^*, \dots, W_n^*)$. To determine the optimal weight of each of the indices, the pairs are formed: $a_{Bj} = W_B/W_j$ and $a_{jw} = w_j/w_w$. To satisfy these conditions for all j, a solution must be found that maximizes the expressions $\left|{W_B}/{W_j} - a_{B_j}\right|$ and $\left|{W_j}/{W_W} - a_{jW}\right|$ for all j that are minimized. Considering the non-negativity of the weights and the sum of the weights, the model can be formulated as follows:

(3)
$$\text{Min max}_j \left\{\left|\frac{W_B}{W_j} - a_{Bj}\right|, \left|\frac{W_j}{W_W} - a_{jW}\right|\right\}$$
$$\sum W_j = 1$$
$$W_j \geq 0, \text{ for all } j$$

Additionally, the following model is equivalent to the above model:

(4)
$$\text{Min } \boldsymbol{\lambda}$$
$$\text{s.t.}$$
$$\left|\frac{W_B}{W_j} - a_{Bj}\right| \leq \boldsymbol{\lambda}, \text{ for all } j$$
$$\left|\frac{W_j}{W_W} - a_{jW}\right| \leq \boldsymbol{\lambda}, \text{ for all } j$$
$$\sum_j w_j = 1$$
$$W_j \geq 0, \text{ for all } j$$

The linear model of the above function is also presented in the following. In this study, the weights of the indicators are obtained using the linear model.

(5)
$$\text{Min } \boldsymbol{\lambda}$$
$$\text{s.t.}$$
$$|W_B - a_{Bj} w_j| \leq \boldsymbol{\lambda}, \text{ for all } j$$

$$|W_j - a_{jw}w_w| \leq \boldsymbol{\lambda}, \text{ for all } j$$
$$\sum_j w_j = 1$$
$$W_j \geq 0, \text{ for all } j$$

By solving the above model, the optimal values $(W_1^*, W_2^*, \ldots, W_n^*)$ and $\boldsymbol{\lambda}$ are obtained.

Findings

In the first phase of the study, a qualitative approach was adopted, and semi-structured interviews were conducted with experts in the research field, who voluntarily agreed to participate in the study. Table 2 presents the profile of the interview participants. The transcripts of the semi-structured interviews were carefully analyzed to identify the components contributing to the realization of trustworthy AI in business. After the interview data were organized and meaningful units were extracted, the identified components were coded and grouped using thematic analysis.

Table 2. Demographic Characteristics of Experts

Interviewee Code	Education Level	Field/Field of Expertise	Work Experience
P1	PhD	Multimedia systems (E-commerce focus)/NLP researcher	8
P2	PhD	Computer engineering/AI model developer	10
P3	PhD	Business management professional/entrepreneur, founder, and CEO of a smart services company	18
P4	PhD	IT engineering	10
P5	PhD	IT engineering	10
P6	PhD	CEO of advanced technologies development and innovation company/AI specialist	16
P7	Master's degree	Algorithm and computational engineering/AI instructor	12
P8	PhD	Public administration/Researcher in the field of AI in business and director of the marketing department	17
P9	PhD	IT/AI in supply chain	8
P10	PhD	AI and business consultant	13
P11	PhD	Business consultant/Product and project manager/AI	16
P12	Master's degree	AI and robotics/Developer and implementer of machine learning algorithms	5
P13	Master's degree	Data science and AI instructor/Founder of Vitake	9
P14	Master's degree	Senior machine learning engineer and specialist/Business development consultant/Product and logistics manager	5

The transcripts of the semi-structured interviews were carefully analyzed to identify the components contributing to the realization of trustworthy AI in business. After the interview data were organized and meaningful units were extracted, the identified components were coded and grouped using thematic analysis. Examples of these components are presented in Table 3. Following the extraction and analysis of concepts from the interview data, the components identified as contributing to the realization of trustworthy AI in business were compiled and are presented.

Table 3. Some Verbal Evidence Extracted from Semi-Structured Interviews

Extracted codes	Speech Evidence
Communication encryption and secure data storage	*"…The AI monitors the user's movement behavior and stores the data. One solution to mitigate security concerns is to encrypt the data before presenting it to the AI…"*
Dashboards for monitoring the performance of AI systems based on reliability indicators	*"….Users have repeatedly encountered incorrect or unexpected performance of AI. The AI algorithm responds based on what it has learned; if the learning is incorrect, the performance is also incorrect. But it can be corrected by providing accurate data and giving direction to the responses (feedback mechanism). A periodic monitoring can be used to show the performance of the models and*
Efficiency of model design, development, maintenance, and updating costs relative to the value produced	*"….The AI system must be commensurate with the maintenance and development costs so that the value it creates for the business is cost-effective. The economic, cultural, and environmental conditions of organizations and businesses must be compatible…"*
-Improving data quality and fixing deficiencies -Using diverse, comprehensive, and valid educational data	*"….In large language models (LMMs), they may provide false answers that they think are correct (delusion). Work is currently underway to help models overcome this delusion. Providing training data and correct insights to AI reduces biased decisions…."*
Fair access by reducing the digital divide between AI system target users in different geographical regions	*"…It is important to consider environmental sustainability within the framework of ethical responsibilities regarding AI systems. If digital infrastructure is not balanced in different regions, then justice in the use of AI technologies will not be achieved…"*
Human observers control the final decision in sensitive decisions	*"…Because AI does not have the power to think, is man-made, and acts according to the data it is given, the presence of human observers is necessary as a mechanism against system errors…."*
Accessibility design for different groups with different devices	*"…The existence of a model designed with simple and user-friendly access for different groups of users regarding access to these systems…"*

Following the extraction and analysis of concepts from the interview data, the components identified as contributing to the realization of trustworthy AI in business were compiled and are presented in Table 4.

Table 4. Factors Influencing the Realization of Reliable AI in Business

Dimensions	Items' Codes	Items	Sources	Some Related Studies
	TOA1	Communication encryption and secure data storage	P1, P3, P5, P6, P7, P9, P10, P12, P14	Bellamkonda (2019); Hazra et al. (2024); Victor et al. (2023)

Table 4. Factors Influencing the Realization of Reliable AI in Business

Dimensions	Items' Codes	Items	Sources	Some Related Studies
	TOA2	Protecting privacy and sensitive information	P1, P2, P3, P4, P5, P8, P13, P14	Sargiotis (2024); Stach et al. (2022); Timan & Mann (2021)
	TOA3	Existence of user control mechanisms over privacy settings	P1, P2, P3, P4, P5, P8, P10, P13	Khandelwal et al. (2021); Ahmadinejad & Fong (2013)
	TOA4	Definition of multi-factor authentication methods	P4, P6, P8, P12	Ometov et al. (2018); ALSaleem & Alshoshan (2021); Singh (2017)
	TOA5	Transparency in the source, method of data collection, processing and final decision	P1, P2, P4, P5, P7, P10, P12, P13	Koops (2013); Tanveer (2026)
	TOA6	Ability to trace the decision-making process (path of converting input to output)	P2, P3, P4, P6, P7, P10, P12, P13	Singh et al. (2018); Kroll (2021); Wenzel et al. (2025)
	TOA7	Explainability, understanding and clarity of the system for users	P1, P3, P5, P7, P10, P12, P13, P14	Naveed et al. (2024); Chazette & Schneider (2020); Akhtar et al. (2024)
	TOA8	Dashboards for monitoring the performance of AI systems based on reliability indicators	P1, P4, P6, P7, P8, P9	Karri (2024); Rahman et al. (2025); Tanveer (2026)
	TOA9	Providing a system for receiving user feedback and modifying the model	P1, P3, P4, P12	Madaan et al. (2022); Chen et al. (2021) Honeycutt et al. (2020)
	TOA10	Periodic monitoring of system algorithms and modifying their model	P1, P3, P5, P8, P12, P13	Dvorak & Kuipers (1989)
	TOA11	Definition of a clear mechanism for managing errors and the effect of user feedback	P2, P3, P5, P7, P9, P12	Madaan et al. (2022); Akhtar et al. (2024); McBride et al. (2014)
	TOA12	Existence of a regular program for improving the efficiency and performance of models	P2, P3, P5, P7, P8, P19, P11	Balasubramanian (2020); Vangala (2022)
	HA1	Defining the level of human intervention in AI decisions	P1, P2, P5, P8, P13, P14	Sterz (2024); Lai et al. (2021); Schmidt (2020); Zhu et al. (2026)
	HA2	Human observer control over the final decision in sensitive decisions	P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P12	Schmidt (2020); Santoni et al. (2018)
	HA3	Defining the specific responsibility and role of humans in the AI decision-making cycle	P1, P4, P7, P9, P10, P12	Zhu et al. (2026); Koulu (2020); Nushi et al. (2019)
	HA4	Training and empowering human observers to interact with AI systems	P1, P7, P10, P12	Wienrich et al. (2023); Endsley (2023); Laux (2024)

Table 4. Factors Influencing the Realization of Reliable AI in Business

Dimensions	Items' Codes	Items	Sources	Some Related Studies
	HA5	Periodic assessment of the quality of human interventions on the output of AI decisions	P1, P2, P6, P8, P10	Zhu et al. (2026); Langer et al. (2024)
	HA6	Improving organizational culture and increasing user awareness	P3, P5, P7	Cetindamar et al. (2022); PEOPLE (2022)
	HA7	The possibility of personalizing and modifying the system by the user	P1, P4, P9, P11	Sundar & Marathe (2010)
	HA8	Simple and understandable user interface	P3, P9, P14	Brand et al. (2023); Božić (2023)
	HA9	Accessibility design for different groups with different devices	P3, P6, P9, P14	Kinnula et al. (2021); Bittencourt et al. (2013)
	HA10	The rate of acceptance of intelligent systems by systems and organizational units	P3, P7, P9	Jöhnk et al. (2021); Alsheibani et al. (2018); Dwivedi et al. (2021)
	LA1	Compliance with local laws and regulations and local ethical requirements	P2, P4, P5, P8, P9, P10, P13, P14	Sonani & Govindarajan (2025); Baplawat et al. (2025); Batool et al. (2026)
	LA2	Existence of safety laws and regulations and compliance with international standards	P2, P5, P7, P10, P14	Trager et al. (2023); Batool et al. (2026)
	LA3	Providing infrastructure for legal acceptance of AI in the country	P4, P6, P12, P14	Bhagra & Singh (2024)
	LA4	Collecting native data (such as Persian language data)	P5, P6, P13, P14	Kshetri (2024)
	MA1	Monitoring performance errors of different groups	P1, P3, P5, P7, P8, P13	Langer et al. (2024); Braun et al. (2024)
	MA2	Using diverse, comprehensive, and reliable training data	P1, P2, P3, P4, P5, P6, P7, P8, P9, P12	Petkovic (2023); Sharfuddin & Choudhury (2025)
	MA3	Improving data quality and eliminating their shortcomings	P1, P2, P3, P4, P5, P6, P8, P9, P11, P13, P14	Petkovic (2023); Sharfuddin & Choudhury (2025); Aldoseri (2023)
	MA4	Existence of mechanisms to object to biases by users of diverse groups	P1, P2, P3, P7, P11, P14	Mehrabi et al. (2021); Fazil et al. (2024); Schwartz et al. (2022)
	SSA1	Design and development of green models (AI); Use of data center with renewable energy	P2, P6, P8	Tabbakh et al. (2024); Alloghani (2023)

Table 4. Factors Influencing the Realization of Reliable AI in Business

Dimensions	Items' Codes	Items	Sources	Some Related Studies
Sustainability and social impact aspect	SSA2	Cost efficiency of model design, development, maintenance, and updating relative to the value	P1, P6, P11	Abuzaid (2024)
	SSA3	Human monitoring of the ethical, social and cultural implications of AI deployment	P1, P12, P13, P14	Hagerty & Rubinov (2019); Camilleri (2024); Nosike et al (2024)
	SSA4	Continuous monitoring of environmental impacts and training and deployment of AI models	P6, P8, P13	Penev et al. (2024); Wu et al. (2022); Tettey et al. (2025)
	SSA5	Equitable access by reducing the digital divide between target users of the AI system in different geographical regions	P2, P3, P7	Božić (2023); Kitsara (2022)

Note: P1-P14 denote the interviewee codes

Subsequently, five academic experts in the relevant research field and senior business managers completed the pairwise comparisons required for the BWM. Specifically, the pairwise comparison vector of the best criterion with respect to the other criteria (A_B) and the pairwise comparison vector of the other criteria with respect to the worst criterion (A_W) were determined. The data obtained from the weighting section of the BWM questionnaire were then analyzed using LINGO software to calculate the weights of the identified dimensions, concepts, and components. Table 5 presents these factors along with their corresponding weights and consistency ratio. As shown in the table, all consistency ratios fall within the acceptable range recommended by Liang et al. (2020) for the BWM pairwise comparison matrices, indicating the reliability of the findings.

Table 5. Final Definitive Weight of the Research Criteria and Sub-Criteria Based on the BWM Method

Dimensions	Dimensions' Weights	Items	Criteria Weights	Criteria' Global Weights	Consistency Ratio
		TOA1	0.08582	0.02977	
		TOA2	0.04709	0.01634	
		TOA3	0.09329	0.03237	
		TOA4	0.10849	0.03764	
		TOA5	0.13140	0.04559	
		TOA6	0.10437	0.03621	
		TOA7	0.08582	0.02977	
		TOA8	0.07746	0.02687	
		TOA9	0.05191	0.01801	
		TOA10	0.01145	0.00397	
		TOA11	0.08582	0.02977	
		TOA12	0.11702	0.04060	
		HA1	0.10780	0.02339	
		HA2	0.20289	0.04402	

Table 5. Final Definitive Weight of the Research Criteria and Sub-Criteria Based on the BWM Method

Dimensions	Dimensions' Weights	Items	Criteria Weights	Criteria' Global Weights	Consistency Ratio
		HA3	0.12996	0.02820	
		HA4	0.08885	0.01928	
		HA5	0.10832	0.02350	
		HA6	0.07609	0.01651	
		HA7	0.08885	0.01928	
		HA8	0.07062	0.01532	
		HA9	0.01823	0.00395	
		HA10	0.10832	0.02350	
		LA1	0.45187	0.01943	
		LA2	0.17035	0.00732	
		LA3	0.30863	0.01327	
		LA4	0.06913	0.00297	
		MA1	0.40955	0.08887	
		MA2	0.33928	0.07362	
		MA3	0.18556	0.04026	
		MA4	0.06560	0.01423	
		SSA1	0.05813	0.01005	
		SSA2	0.23840	0.04124	
		SSA3	0.32715	0.05659	
		SSA4	0.13790	0.02385	
		SSA5	0.23840	0.04124	

As shown in Table 5, the results of the BWM analysis indicate that the *technical and operational dimension* is the most important dimension, with a weight of 0.347. The *human dimension* and the *ethical dimension* are jointly ranked second, each with a weight of 0.217. The *sustainability and social impact dimension* rank third, with a weight of 0.173, while the *legal dimension* is the least important dimension, with a weight of 0.043, according to the experts' evaluations.

Among the 35 identified sub-criteria, the global weights obtained through the BWM indicate that '*Monitoring performance errors across different groups,'* with a weight of 0.08887, was ranked as the most important sub-criterion, followed by '*Using diverse, comprehensive, and reliable training data,'* with a weight of 0.07362, and '*Human oversight of the ethical, social, and cultural implications of AI deployment,*' with a weight of 0.05659.

Within the technical and operational dimension, '*Transparency in data sources, data collection methods, data processing, and final decision-making'* was identified as the most important sub-criterion, with a weight of 0.04559. This was followed by the '*existence of a regular program for improving the efficiency and performance of AI models,'* with a weight of 0.04060.

Within the human dimension, '*human oversight of final decisions in sensitive decision-making situations'* was identified as the most important sub-criterion, with a weight of 0.04402.

This was followed by '*clearly defining human responsibilities and roles throughout the AI decision-making process,'* with a weight of 0.02820.

Within the legal dimension, '*Compliance with local laws, regulations, and ethical requirements*' was identified as the most important sub-criterion, with a weight of 0.01943. This was followed by providing the '*necessary infrastructure for the legal acceptance of AI within the country,*' with a weight of 0.01327.

Within the ethical dimension, '*monitoring performance errors across different groups*' was identified as the most important sub-criterion, with a weight of 0.08887. Notably, this sub-criterion also received the highest global weight among all 35 identified sub-criteria. This was followed by '*using diverse, comprehensive, and reliable training data,'* with a weight of 0.07362, which ranked as the second most important sub-criterion overall.

Within the sustainability and social impact dimension, 'Human oversight of the ethical, social, and cultural implications of AI deployment' was identified as the most important sub-criterion, with a weight of 0.05659. This was followed jointly by 'Cost efficiency in the design, development, maintenance, and updating of AI models relative to the value generated' and 'Equitable access through reducing the digital divide among target users of AI systems across different geographical regions,' each with a weight of 0.04124.

Discussion and Conclusion

The primary objective of this study was to address the gap in the existing literature by developing an integrated framework for trustworthy AI dimensions using a mixed-methods approach. In the first phase, a qualitative approach was employed to explore the components and indicators of trustworthy AI in business. In the second phase, the identified dimensions and components were prioritized using the BWM. The findings indicate that the proposed framework for trustworthy AI comprises five main dimensions and 35 sub-criteria. The following sections discuss each of these dimensions in detail.

Technical and Operational Aspect: This dimension refers to the technological infrastructure, capabilities, and control mechanisms that enable these systems to be secure, stable, understandable, and controllable. In the literature, this dimension is usually defined by concepts such as security, privacy, transparency, explainability, data encryption, and continuous monitoring (Kumar et al., 2026). This paper reveals that "trustworthiness" is not just a technical characteristic of the model but also a combination of management standards, risk management, and legal requirements (Tanveer, 2026). Other research has stated that for the responsible use of AI, technical accuracy alone is not enough; we must be able to explain the behavior of the model, regulate trust, and make decisions understandable (Arrieta et al., 2019). Li et al. (2023) focuses on dimensions such as robustness, explainability, transparency, and repeatability. New research also emphasizes the need for transparency and accountability in AI algorithms and assessing the reliability of systems (Mortaji & Sadeghi, 2024). Taken together, from an interpretative perspective, when in the BWM results, it can be argued that the experts studied linked trust primarily to the "security" and "understandability" of the system. This interpretation is also consistent with the literature, because in many trustworthiness frameworks, data protection and the ability to explain decisions are recognized as basic elements of trust. Therefore, in the present study, within the technical and operational dimension, "transparency

in data sources, data collection methods, data processing and final decision-making" was identified as the most important sub-criterion, with a weight of 0.04559.

Human Dimension: This aspect focuses on users and human observers in the formation and maintenance of trust in AI. In the literature, this dimension is usually defined by components such as human oversight and effective human-AI interaction, with the user remaining in the decision-making loop rather than being completely replaced by AI. This suggest that both over-trust in AI and under-trust can undermine performance (Nushi et al., 2019). The human factor should be actively maintained and enhanced by AI systems, emphasizing the critical role of human participation in the responsible deployment of AI, but not all AI applications require the same level of human oversight (Kandikatla & Radeljic, 2025). However, other research suggests that building trustworthy AI is a necessity, especially in high-risk applications. For responsible AI, humans need to have a meaningful presence in the process; this presence varies depending on the application, method, and (Lazaros et al., 2026). Accordingly, in the present study, the human dimension, "human supervision of final decisions in sensitive decision-making situations," with a weight of 0.04402, was identified as the most important sub-criterion.

Legal Dimension: As AI systems increasingly influence decision-making processes and operational efficiency, the need for comprehensive and forward-looking regulations becomes essential. Therefore, in ethical aspects, algorithms are used to protect rights or prohibit discrimination and ensure the transparency of AI decisions (Bhagra & Singh, 2024). Trustworthy AI, on the other hand, refers to the extent to which the system complies with laws, regulations, legal requirements, and formal governance mechanisms. In the new literature, this dimension is recognized as one of the key pillars of Trustworthy AI; because even if a model is technically accurate and ethically desirable, its trustworthiness will not be complete without legal backing and institutional compliance (Smuha, 2025). Therefore, it can be said that the present study adds to the existing literature by revealing the fact that the legal dimension in trustworthy AI is not only “compliance with the law” but also includes the design of legal mechanisms, accountability, and adaptation to the local context. In practice, the paper consolidates the idea that regulation should not only see trustworthiness as “risk management” but also should consider it as a multidimensional concept that includes factors such as: transparency, accountability, fairness, and safety (Laux et al., 2024). Contrary to what said, recent studies argue that AI regulation is not simply a legal issue, but a structural determinant of technological development (Vujicic, 2024). The present study identified “compliance with local laws, regulations, and ethical requirements” as the most important sub-criterion in measuring the legal dimension criteria, with a weight of 0.01943.

Ethical Dimension: As AI advances, it is crucial to consider ethical considerations surrounding its implementation. Bias reduction, transparency, and accountability are essential for the responsible deployment of AI. Bias reduction is crucial to ensure that AI systems do not perpetuate or reinforce existing biases in society (Mensah, 2023). In this study, indicators such as monitoring performance errors, the use of diverse data, and mechanisms for challenging bias were included in the ethical dimension. This finding is consistent with the literature, as several studies have shown that the lack of diversity in educational data can reproduce or exacerbate

bias, thereby undermining the trust of individuals and stakeholder groups (Mehrabi et al., 2021). The important innovation of the present study in this dimension is that it takes ethics beyond the general level and redefines it as a set of enforcement and monitoring mechanisms. In the ethical dimension, "monitoring performance errors in different groups" was identified as the most important sub-criterion, with a weight of 0.08887.

Sustainability and Social Impact Dimension: The rapid development of AI must be supported by the necessary insight and regulatory oversight for AI-based technologies to enable sustainable development (Vinuesa et al., 2020). Discriminatory AI in various domains can exacerbate existing racial inequalities from education to law enforcement (Min, 2023). However, the sub-criterion of the cost-effectiveness of model design, development, maintenance, and updating relative to the value produced is almost entirely discussed, and it is the first time that the cost of maintaining, updating, and developing a model is directly critiqued in a framework (Ozili, 2025). This is valuable because a significant portion of recent trustworthy AI studies have focused more on fairness, transparency, and privacy, and have less incorporated sustainability as a central dimension of trustworthiness into the model. While recent research on Green AI has clearly shown that computational cost and environmental footprint of models should be part of responsible technology evaluation (Schwartz et al., 2020; Strubell et al., 2020), by weighting the sustainability and social impact dimensions, “human oversight of ethical, social, and cultural implications of AI deployment” was identified as the most important sub-criterion, with a weight of 0.05659.

Practical and Theoretical Implication

The present study is among the first to develop an integrated framework for trustworthy AI in business applications within the socio-cultural context of Iran. The main theoretical contributions of this study are summarized as follows.

First, to the best of the authors' knowledge, previous studies have primarily focused either on the technical and operational aspects of AI adoption or exclusively on its ethical dimensions. In contrast, this study adopts a holistic perspective by integrating the technical, operational, human, ethical, legal, and sustainability dimensions of trustworthiness into a comprehensive framework for trustworthy AI in business environments.

Second, although previous research has identified various dimensions of trustworthy AI, limited attention has been paid to prioritizing these dimensions. This study addresses this gap by prioritizing the identified dimensions and components, thereby providing valuable guidance for researchers, decision-makers, and practitioners in allocating limited organizational resources more effectively and focusing on the most critical aspects of trustworthy AI implementation.

Finally, despite the growing recognition of the social and environmental implications of AI, sustainability has received relatively limited attention in the trustworthy AI literature. This study contributes to the literature by incorporating sustainability and social impact alongside the technical, operational, human, ethical, and legal dimensions, thereby offering a more comprehensive perspective on achieving trustworthy AI in business contexts. This integrated approach provides a stronger theoretical foundation for the responsible and sustainable deployment of AI technologies.

In addition to advancing the literature on trustworthy AI, the findings of this study provide several practical implications for policymakers and managers seeking to adopt AI technologies.

First, the findings indicate that users' trust in AI systems is shaped by multiple dimensions, including technical, operational, human, ethical, legal, and sustainability-related factors. Therefore, organizations should adopt a multidimensional approach when developing and implementing AI systems to enhance users' trust and facilitate successful AI adoption.

Second, because the technical and operational dimension was identified as the highest priority, organizations should pay particular attention to information security and privacy, transparency in data sources and AI models, model explainability, and the continuous monitoring, evaluation, and updating of AI models to ensure their reliability and trustworthiness.

Third, given the importance of the human dimension, organizations should place greater emphasis on user-related factors, including AI literacy and user training, clearly defining the level of human involvement and oversight in AI-driven processes, specifying users' roles in decision-making, providing personalized user experiences, and designing intuitive and user-friendly interfaces, including those that clearly communicate AI-detected risks (Eskandari et al., 2026. These measures can strengthen users' confidence in AI systems and improve their willingness to adopt and use them.

Limitations and Future Research Suggestions

Despite its attempt to develop a framework for the realization of trustworthy AI in Iranian businesses, the present study, like any research, has several limitations that provide opportunities for future research. First, the proposed framework was developed based on the perspectives of experts and the contextual requirements of Iranian businesses. Therefore, cultural, social, and legal conditions, as well as the country's level of data maturity, may influence the formulation of its dimensions and components. Accordingly, future studies are encouraged to examine and compare this framework across different countries or business environments with varying levels of technological, data, and analytical maturity to facilitate the development of a more comprehensive and generalizable model.

Second, this study primarily focused on identifying, explaining, and prioritizing the dimensions of trustworthy AI. However, the interrelationships among the identified components, the extent of their mutual influence, and the underlying causal structure of the proposed framework were not independently analyzed. Therefore, future research is recommended to employ the Decision-Making Trial and Evaluation Laboratory (DEMATEL) method to examine the causal relationships and the degree of influence among the identified dimensions and indicators. In addition, Social Network Analysis (SNA) could be used to reveal the interaction structure, network density, and the centrality of key components within the proposed framework.

Finally, the application of quantitative techniques such as Structural Equation Modeling (SEM) could provide empirical validation of the relationships among the antecedents, dimensions, and outcomes of trustworthy AI implementation in Iranian businesses, thereby enhancing the explanatory validity of the proposed framework.

Acknowledgement

The authors used ChatGPT to assist with language editing, grammar correction, proofreading, and the translation of portions of the manuscript from Persian into English. ChatGPT was used solely to improve the linguistic quality of the manuscript and did not contribute to the scientific content, data analysis, or interpretation of the results. The authors take full responsibility for the accuracy, originality, integrity, and scientific content of the manuscript.

References

Abuzaid, A. N. (2024, April). AI-driven business processes: Efficiency gains, cost reductions, and implications for economic performance. In 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS) (Vol. 1, pp. 1-5). IEEE. https://doi.org/10.1109/ICKECS61492.2024.10617370

Ahmadinejad, S. H., & Fong, P. W. (2013, May). On the feasibility of inference attacks by third-party extensions to social network systems. In Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security (pp. 161-166). https://doi.org/10.1145/2484313.2484333

Akhtar, M. A. K., Kumar, M., & Nayyar, A. (2024). Transparency and accountability in explainable AI: Best practices. In Towards ethical and socially responsible explainable ai: Challenges and opportunities (pp. 127-164). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-66489-2_5

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for artificial intelligence: concepts, opportunities, and challenges. Applied Sciences, 13(12), 7082. https://doi.org/10.3390/app13127082

Alenizi, F. A., Abbasi, S., Mohammed, A. H., & Rahmani, A. M. (2023). The artificial intelligence technologies in Industry 4.0: A taxonomy, approaches, and future directions. Computers & Industrial Engineering, 185, 109662. https://doi.org/10.1016/j.cie.2023.109662

Alloghani, M. A. (2023). Architecting green artificial intelligence products: Recommendations for sustainable ai software development and evaluation. In Artificial Intelligence and Sustainability (pp. 65-86). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-45214-7_4

ALSaleem, B. O., & Alshoshan, A. I. (2021, March). Multi-factor authentication to systems login. In 2021 National Computing Colleges Conference (NCCC) (pp. 1-4). IEEE. https://doi.org/10.1109/NCCC49330.2021.9428806

Alsheibani, S., Cheung, Y., & Messom, C. (2018). Artificial intelligence adoption: AI-readiness at firm-level. In Pacific Asia conference on information systems 2018 (p. 37). Association for Information Systems. https://aisel.aisnet.org/pacis2018/37

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gillopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115. https://doi.org/ 10.1016/j.inffus.2019.12.012.

Ashoori, M., & Weisz, J. D. (2019). In AI we trust? Factors that influence trustworthiness of AI-infused decision-making processes. arXiv preprint arXiv:1912.02675. https://doi.org/10.48550/arXiv.1912.02675

Balasubramanian, A. (2020). End-to-end model lifecycle management: An MLOPS framework for drift detection, root cause analysis, and continuous retraining. Int. J. Multidiscip. Res. Growth Eval, 1(1), 92-102. https://doi.org/10.54660/.IJMRGE.2020.1.1-92-102

Baldassarre, M. T., Gigante, D., Kalinowski, M., & Ragone, A. (2024, April). POLARIS: A framework to guide the development of Trustworthy AI systems. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI (pp. 200-210). https://doi.org/10.1145/3644815.364494

Baplawat, A., Shete, J. M., Kurup, S. J., Thomas, S. A., Bhosale, Y. H., & Priya, A. (2025). AI-Driven policy enforcement in local governance: Legal frameworks and ethical boundaries. Vascular and Endovascular Review, 8(6s), 442-449. https://verjournal.com/index.php/ver/article/view/589

Batool, A., Lee, S., Liu, Y., & Dong, L. (2026). The anatomy of AI policies: A systematic comparative analysis of AI policies across the globe. AI and Ethics, 6(1), 55 .https://doi.org/10.1007/s43681-025-00886-3

Bellamkonda, S. (2019). Securing data with encryption: A comprehensive guide. International Journal of Communication Networks and Security, 11, 248-254.

Bhagra, L., & Singh, T. (2024). Leveraging artificial intelligence in business activation: Legal and IPR perspectives. DME Journal of Management, 5(01), 24-33. https://doi.org/10.53361/dmejm.v5i01.04

Bittencourt, I. I., Baranauskas, M. C., Dermeval, D., Braga, J., & Pereira, R. (2013, September). Designing for different users and multiple devices: A roadmap towards inclusive environments. In IFIP Conference on Human-Computer Interaction (pp. 605-622). :اشئ×Springer Berlin Heidelberg. https://dx.doi.org/10.1007/978-3-642-40483-2_43

Božić, V. (2023a). Application of artificial intelligence in user interface design. Researchgate. https://doi.org/10.13140/RG.2.2.26105.34401.

Božić, V. (23 July, 2023b). Risks of digital divide in using artificial intelligence (AI). Researchgate. https://www.researchgate.net/publication/371126516. https://doi. org/10.13140/RG

Brand, L., Humm, B. G., Krajewski, A., & Zender, A. (2023, June). Towards improved user experience for artificial intelligence systems. In International Conference on Engineering Applications of Neural Networks (pp. 33-44). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34204-2_4

Braun, M., Greve, M., Brendel, A. B., & Kolbe, L. M. (2024). Humans supervising artificial intelligence–investigation of designs to optimize error detection. Journal of Decision Systems, 33(4), 674-699. https://doi.org/10.1080/12460125.2023.2260518

Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. Expert Systems, 41(7), e13406. https://doi.org/10.1111/exsy.13406

Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI literacy of employees at digital workplaces. IEEE Transactions on Engineering Management, 71, 810-823. https://doi.org/10.1109/TEM.2021.3138503

Chander, B., John, C., Warrier, L., & Gopalakrishnan, K. (2025). Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness. ACM Computing Surveys. https://doi.org/10.1145/3675392

Chazette, L., & Schneider, K. (2020). Explainability as a non-functional requirement: challenges and recommendations. Requirements Engineering, 25(4), 493-514. https://doi.org/10.1007/s00766-020-00333-1

Chen, M., Chang, B., Xu, C., & Chi, E. H. (2021, March). User response models to improve a reinforce recommender system. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining (pp. 121-129). https://doi.org/10.1145/3437963.3441764

Dang, Q., & Li, G. (2026). Unveiling trust in AI: The interplay of antecedents, consequences, and cultural dynamics. AI & SOCIETY, 41(1), 669-692. https://doi.org/10.1007/s00146-025-02477-6

Dewalska-Opitek, A., Witczak, O., Szostak, A., Dziura, M., & Wroniszewska-Drabek, B. (2024). Generation Z's trust toward artificial intelligence: attitudes and opinions. European Research Studies Journal, XXVII(3), 33-52. https://doi.org/10.35808/ersj/3475

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

Dvorak, D., & Kuipers, B. (1989, August). Model-Based monitoring of dynamic systems. In IJCAI (pp. 1238-1243). Los Altos, CA: Morgan Kauffmann.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, V., Janssen, M., Jones, P., Kumar Kar, A., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi:10.1016/j.ijinfomgt.2019.08.002

Endsley, M. R. (2023). Supporting human-AI teams: Transparency, explainability, and situation awareness. Computers in Human Behavior, 140, 107574. https://doi:10.1016/j.chb.2022.107574

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. Information Systems Frontiers, 24(5), 1709-1734. https://doi:10.1007/s10796-021-10186-w

Eskandari, M., Indukuri, M. K. V., Lukin, S. M., & Matuszek, C. (2026). Autonomous VR-Based Risk Detection for Situational Awareness in Dangerous Settings. arXiv preprint arXiv:2607.16582. https://doi.org/10.48550/arXiv.2607.16582

Eskandari, M., Indukuri, M., Lukin, S. M., & Matuszek, C. (2025). LLM-supported safety annotation in high-risk environments. In HRI 2025 Workshop VAM-HRI.

Fazil, A. W., Hakimi, M., & Shahidzay, A. K. (2024). A comprehensive review of bias in AI algorithms. Nusantara Hasana Journal, 3(8), 1-11. https://doi:10.59003/nhj.v3i8.1052

Ferdaus, M. M., Abdelguerfi, M., Loup, E., N. Niles, K., Pathak, K., & Sloan, S. (2026). Towards trustworthy AI: A review of ethical and robust large language models. ACM Computing Surveys, 58(7), 1-43. https://doi.org/10.1145/3777382

Gartner. (2026). Gartner forecasts worldwide AI spending to grow 47% in 2026. https://www.gartner.com/en/newsroom/press-releases/2026-05-19-gartner-forecasts-worldwide-ai-spending-to-grow-47-percent-in-2026 (Accessed 21 July 2026)

Hagerty, A., & Rubinov, I. (2019). Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. arXiv preprint arXiv:1907.07892 .https://doi.org/10.48550/arXiv.1907.07892

Hazra, R., Chatterjee, P., Singh, Y., Podder, G., & Das, T. (2024). Data encryption and secure communication protocols. In Strategies for e-commerce data security: Cloud, blockchain, AI, and machine learning (pp. 546-570). IGI Global Scientific Publishing. https://doi:10.4018/979-8-3693-6557-1.ch022

Hohma, E., & Lütge, C. (2023). From trustworthy principles to a trustworthy development process: The need and elements of trusted development of AI systems. AI, 4(4), 904-925. https://doi.org/10.3390/ai4040046

Honeycutt, D., Nourani, M., & Ragan, E. (2020, October). Soliciting human-in-the-loop user feedback for interactive machine learning reduces user trust and impressions of model accuracy. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (Vol. 8, pp. 63-72). https://doi.org/10.48550/arXiv.2008.12735

Indukuri, M., Eskandari, M., Kollu, S. N., Lukin, S., & Matuszek, C. (2026). Hazard or Anomaly? Evaluating VLMs for Understanding Dangers and Discrepancies. arXiv preprint arXiv:2607.18325. https://doi.org/10.48550/arXiv.2607.18325

Jang, M. (2026). Building trust in AI: The role of technical capacity, social risk, and corporate institutional accountability. Information, 17(2), 212. https://doi.org/10.3390/info17020212

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes—An interview study of organizational AI readiness factors. Business & Information Systems Engineering, 63(1), 5-20. https://doi:10.1007/s12599-020-00676-7

Kandikatla, L., & Radeljic, B. (2025). AI and human oversight: A risk-based framework for alignment. arXiv preprint arXiv:2510.09090 .https://doi.org/10.48550/arXiv.2510.09090

Karri, N. (2024). Real-Time performance monitoring with AI. International Journal of Emerging Trends in Computer Science and Information Technology, 5(1), 102-111. https://doi:10.63282/3050-9246.IJETCSIT-V5I1P111

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: a review. ACM computing surveys (CSUR), 55(2), 1-38. https://doi.org/10.1145/3491209

Khandelwal, R., Linden, T., Harkous, H., & Fawaz, K. (2021). PriSEC: A privacy settings enforcement controller. In 30th USENIX Security Symposium (USENIX Security 21) (pp. 465-482). University of Wisconsin–Madison

Kinney, M., Anastasiadou, M., Naranjo-Zolotov, M., & Santos, V. (2024). Expectation management in AI: A framework for understanding stakeholder trust and acceptance of artificial intelligence systems. Heliyon, 10. https://doi.org/10.1016/j.heliyon.2024.e28562

Kinnula, M., Iivari, N., Sharma, S., Eden, G., Turunen, M., Achuthan, K., Nedungadi, P., Avellan, T., Thankachan, B., & Tulaskar, R. (2021, June). Researchers' toolbox for the future: Understanding and designing accessible

and inclusive artificial intelligence (AIAI). In Proceedings of the 24th international academic Mindtrek Conference (pp. 1-4). Mindtrek '21. https://doi.org/10.1145/3464327.3464965

Kitsara, I. (2022). Artificial intelligence and the digital divide: From an innovation perspective. In Platforms and artificial intelligence: The next generation of competences (pp. 245-265). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-90192-9_12

Koops, B. J. (2013). On decision transparency, or how to enhance data protection after the computational turn. In M. Hildebrandt & K. De Vries (eds), Privacy, due process and the computational turn, Abingdon (pp. 196-220). Routledge

Koulu, R. (2020). Proceduralizing control and discretion: Human oversight in artificial intelligence policy. Maastricht Journal of European and Comparative Law, 27(6), 720-735. https://doi.org/10.1177/1023263X2097864

Kowald, D., Scher, S., Pammer-Schindler, V., Müllner, P., Waxnegger, K., Demelius, L., Fessl, A., Toller, M., Estrada, I., Simic, I., Sabol, V., Trugler, A., Veas, E., Kern, R., Nad, T., & Kopeinik, S. (2024). Establishing and evaluating trustworthy AI: overview and research challenges. Frontiers in Big Data, 7, 1467222 . https://doi.org/10.3389/fdata.2024.1467222

Kroll, J. A. (2021, March). Outlining traceability: A principle for operationalizing accountability in computing systems. In Proceedings of the 2021 ACM Conference on fairness, accountability, and transparency (pp. 758-771) .https://doi.org/10.1145/3442188.3445937

Kshetri, N. (2024). Linguistic challenges in generative artificial intelligence: Implications for low-resource languages in the developing world. Journal of Global Information Technology Management, 27(2), 95-99. https://doi.org/10.1080/1097198X.2024.2341496

Kumar, D., Hamidi, P., Rana, S. K., Rana, A. K., & Ansarullah, S. I. (2026, April). A Unified Security Framework for SDN-Enabled 5G Core Networks with Secure Network Slicing Enforcement. *In 2026 IEEE 15th International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 7-12). IEEE. https://doi.org/10.1109/CSNT69054.2026.11502394

Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). Towards a science of human-ai decision making: A survey of empirical studies. arXiv preprint arXiv:2112.11471 . https://doi.org/10.48550/arXiv.2112.11471

Langer, M., Baum, K., & Schlicker, N. (2024). Effective human oversight of AI-based systems: A signal detection perspective on the detection of inaccurate and unfair outputs. Minds and Machines, 35(1), 1. https://doi:10.1007/s11023-024-09701-0

Laux, J. (2024). Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. AI & Society, 39(6), 2853-2866. https://doi:10.1007/s00146-023-01777-z

Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance, 18(1), 3-32. https://doi.org/10.1111/rego.12512

Lazaros, K., Vrahatis, A. G., & Kotsiantis, S. (2026). Human-in-the-Loop artificial intelligence: A systematic review of concepts, methods, and applications. Entropy (Basel, Switzerland), 28(4), 377-377. https://doi.org/10.3390/e28040377

Lewis, D., Hogan, L., Filip, D., & Wall, P. J. (2020). Global challenges in the standardization of ethics for trustworthy AI. Journal of ICT Standardization, 8(2), 123-150. https://doi:10.13052/jicts2245-800X.823

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023). Trustworthy AI: From principles to practices. ACM Computing Surveys, 55(9), 1-46. https://doi.org/10.1145/3555803

Li, Y., Wu, B., Huang, Y., & Luan, S. (2024). Developing trustworthy artificial intelligence: Insights from research on interpersonal, human-automation, and human-AI trust. Frontiers in Psychology, 15, 1382693. https://doi:10.3389/fpsyg.2024.1382693

Liang, F., Brunelli, M., & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. Omega, 96, 102175. https://doi:10.1016/j.omega.2019.102175

Machucho, R., & Ortiz, D. (2025). The impacts of artificial intelligence on business innovation: A comprehensive review of applications, organizational challenges, and ethical considerations. Systems, 13(4), 264. https://doi.org/10.3390/systems13040264

Madaan, A., Tandon, N., Clark, P., & Yang, Y. (2022, December). Memory-assisted prompt editing to improve GPT-3 after deployment. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 2833-2861). Abu Dhabi, United Arab Emirates. https://doi:10.18653/v1/2022.emnlp-main.183

Maiti, M., Kayal, P., & Vujko, A. (2025). A study on ethical implications of artificial intelligence adoption in business: challenges and best practices. Future Business Journal, 11(1), 34. https://doi:10.1186/s43093-025-00462-5

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2014). Understanding human management of automation errors. Theoretical Issues in Ergonomics Science, 15(6), 545-577. https://doi.org/10.1080/1463922X.2013.817625

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM computing surveys (CSUR), 54(6), 1-35. https://doi.org/10.1145/3457607

Melcher, P. R., & Vaziri, D. D. (2025). AI competencies for companies: definition, taxonomy an regulatory classification. World Journal of Artifical Intelligence and Robotics Research, 2(3), 01-09. https://doi.org/10.63620/MK.WJAIRR.2025.1012

Mensah, G. B. (2023). Artificial intelligence and ethics: a comprehensive review of bias mitigation, transparency, and accountability in AI Systems. Africa Journal For Regulatory Affairs, Corpus ID: 270231462. https://doi 10.13140/RG.2.2.23381.19685/1

Miedema, E., Waschull, S., & Emmanouilidis, C. (2026). Towards trustworthy artificial intelligence for decision-making: A lifecycle perspective on knowledge-and data-driven artificial intelligence systems. Computers in Industry, 174, 104409. https://doi.org/10.1016/j.compind.2025.104409

Min, A. (2023). Artificial intelligence and bias: Challenges, implications, and remedies. Journal of Social Research, 2(11). https://doi.org/10.55324/josr.v2i11.1477

Mortaji, S. T. H., & Sadeghi, M. E. (2024). Assessing the reliability of artificial intelligence systems: Challenges, metrics, and future directions. International Journal of Innovation in Management, Economics and Social Sciences, 4(2), 1-13. https://doi.org/10.59615/ijimes.4.2.1

Nasr-Azadani, M. M., & Chatelain, J. L. (2024). The journey to trustworthy AI-part 1: Pursuit of pragmatic frameworks. arXiv preprint arXiv:2403.15457 .https://doi.org/10.48550/arXiv.2403.15457

Nastoska, A., Jancheska, B., Rizinski, M., & Trajanov, D. (2025). Evaluating trustworthiness in AI: Risks, metrics, and applications across industries. Electronics, 14(13), 2717. https://doi.org/10.3390/electronics14132717

Naveed, S., Stevens, G., & Robin-Kern, D. (2024). An overview of the empirical evaluation of explainable AI (XAI): A comprehensive guideline for user-centered evaluation in XAI. Applied Sciences, 14(23), 11288. https://doi.org/10.3390/app142311288

Nosike, C. J., Ojobor, O. S. N., Nosike, U. C., & Nosike, R. D. C. J. (2024). Exploring the ethical and societal implications of artificial intelligence. Science, 30(2), 169-195.

Nushi, B., Iqbal, S. T., Bennett, P., Teevan, J., Horvitz, E., Weld, D., Amershi, S., Norvoreanu, M., Fourney, A., Nushi, B., Collison, P., Inkpen, K., Kikin-Gil, R., & Suh, J. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 chi conference on human factors in computing systems. https://doi.10.1145/3290605.3300233

Ometov, A., Bezzateev, S., Mäkitalo, N., Andreev, S., Mikkonen, T., & Koucheryavy, Y. (2018). Multi-factor authentication: A survey. Cryptography, 2(1), 1. https://doi.org/10.3390/cryptography2010001

Oyekunle, D. O. T., Okwudili Matthew, U., Preston, D., & Boohene, D. (2024). Trust beyond technology algorithms: A theoretical exploration of consumer trust and behavior in technological consumption and AI projects. Journal of Computer and Communications, 12(06), 10-4236. https://doi.10.4236/jcc.2024.126006

Ozili, P. K. (2025). Artificial Intelligence and the sustainable development goals: AI applications for each SDG. MPRA Paper 127371, University Library of Munich, Germany. https://doi 10.2139/ssrn.5271772

Paliszkiewicz, J., & Gołuchowski, J. (2024). Trust and artificial intelligence. Developmen t and Application of AI Technology, 1, 344. https://doi.org/10.4324/9781032627236

Pallathadka, H., Ramirez-Asis, E. H., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J. M., & Naved, M. (2023). Applications of artificial intelligence in business management, e-commerce and finance. Materials Today: Proceedings, 80, 2610-2613. https://doi 10.1016/j.matpr.2021.06.419

Penev, K., Gegov, A., Isiaq, O., & Jafari, R. (2024). Energy efficiency evaluation of artificial intelligence algorithms. Electronics, 13(19), 3836. https://doi.org/10.3390/electronics13193836

PEOPLE, C. O. G. T. E. (2022). How to enable sovereign human-AI interactions at work? Practicing sovereignty. Interventions for Open Digital Futures, 209. https://doi 10.34669/wi.cp/4.20

Petkovic, D. (2023). It is not "Accuracy vs. Explainability"—we need both for trustworthy AI systems. IEEE Transactions on Technology and Society, 4(1), 46-53. https://doi.10.1109/TTS.2023.3239921

Rahman, M. A., Alam, M. S., & Mrida, M. S. H. (2025). How interactive dashboards improve managerial decision-making in operations management. American Journal of Advanced Technology and Engineering Solutions, 1(01), 122-146. https://doi.10.63125/cqm5jk84

Regona, M., Yigitcanlar, T., Hon, C., & Teo, M. (2026). Building trust in artificial intelligence: A systematic review through the lens of trust theory. ACM Computing Surveys. org/10.1145/3789256

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. Omega, 53, 49-57. https://doi.10.1016/j.omega.2014.11.009

Ryan, M. (2020). In AI we trust: ethics, artificial intelligence, and reliability. Science and Engineering Ethics, 26(5), 2749-2767. https://doi.10.1007/s11948-020-00228-y

Santoni de Sio, F., & Van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. Frontiers in Robotics and AI, 5, 323836. https://doi. 10.3389/frobt.2018.00015

Sargiotis, D. (2024). Data security and privacy: Protecting sensitive information. In D. Sargiotis (Ed.). Data governance: A guide (pp. 217-245). Springer Nature Switzerland. https://doi.10.1007/978-3-031-67268-2_6

Sarun, H., Rotana, S., & Chhunla, C. (2026). The Role and Significance of Artificial Intelligence in Transforming Modern Society: Opportunities, Challenges, and Future Directions. Journal of Agriculture and Environment, 3(1), 133-141. https://doi.10.5281/zenodo.18258452

Schmidt, A. (2020, September). Interactive human centered artificial intelligence: a definition and research challenges. In Proceedings of the 2020 International Conference on Advanced Visual Interfaces (pp. 1-4). https://doi.10.1145/3399715.3400873

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. Communications of the ACM, 63(12), 54-63. https://doi.org/10.1145/3381831

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a standard for identifying and managing bias in artificial intelligence [Unpublished doctoral dissertation, National Institute of Standards and Technology]. https://doi.org/10.6028/NIST.SP.1270

Sharfuddin, M., & Choudhury, P. (2025). Enhancing data reliability in management information systems through artificial intelligence driven validation and error detection models. Journal of Ai ML DL, 1(1), 1-8. https://doi.10.25163/ai.1110442

Singh, J., Cobbe, J., & Norval, C. (2018). Decision provenance: Harnessing data flow for accountable systems. IEEE Access, 7, 6562-6574 .https://doi.org/10.48550/arXiv.1804.05741

Singh, N., & Chouhan, S. S. (2021, December). Role of artificial intelligence for development of intelligent business systems. In 2021 IEEE International Symposium on Smart Electronic Systems (iSES) (pp. 373-377). IEEE. https://doi.10.36227/techrxiv.16676794.v1

Singh, S. (2017). Multi-factor authentication and their approaches. International Research Journal of Management, IT and Social Sciences, 4(3), 68-81.

Smith, C. J. (2019). Designing trustworthy AI: A human-machine teaming framework to guide development. arXiv preprint arXiv:1910.03515 .https://doi.org/10.48550/arXiv.1910.03515

Smuha, N. A. (2025). Regulation 2024/1689 of the Eur. Parl. & Council of June 13, 2024 (EU artificial intelligence act). International Legal Materials, 64(5), 1234-1381 . https://doi.10.36227/techrxiv.16676794.v110.1017/ilm.2024.46

Sonani, R., & Govindarajan, V. (2025). Cloud integrated governance driven reinforcement framework for ethical and legal compliance in ai based regulatory enforcement. Journal of Selected Topics in Academic Research, 1(1). https://orcid.org/0009-0009-4072-0479

Stach, C., Gritti, C., Bräcker, J., Behringer, M., & Mitschang, B. (2022). Protecting sensitive data in the information age: State of the art and future prospects. Future Internet, 14(11), 302. https://doi.org/10.3390/fi14110302

Statista. (2026). AI development tool software in Iran. https://www.statista.com/outlook/tmo/software/enterprise-software/ai-development-tool-software/iran/ (Accessed 21 July 2026).

Sterz, S. (2024, December). Effective human intervention in AI and beyond: A high-level framework. In 2024 IEEE Smart World Congress (SWC) (pp. 1987-1992). IEEE. https://doi.10.1109/SWC62898.2024.00305

Strubell, E., Ganesh, A., & McCallum, A. (2020, April). Energy and policy considerations for modern deep learning research. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 09, pp. 13693-13696). https://doi.10.1609/aaai.v34i09.7123

Sundar, S. S., & Marathe, S. S. (2010). Personalization versus customization: The importance of agency, privacy, and power usage. Human Communication Research, 36(3), 298-322. https://doi.10.1111/j.1468-2958.2010.01377.x

Tabbakh, A., Al Amin, L., Islam, M., Mahmud, G. I., Chowdhury, I. K., & Mukta, M. S. H. (2024). Towards sustainable AI: A comprehensive framework for Green AI. Discover Sustainability, 5(1), 408. https://doi.10.1007/s43621-024-00641-4

Tanveer, R. (9 May, 2026). AI Governance Frameworks: ISO/IEC 42001, NIST AI RMF, and the EU AI Act. NIST AI RMF, and the EU AI Act. http://dx.doi.org/10.2139/ssrn.6740438

Tettey, D. J., Chrisben, D., Victoria, M. C., Chukwubuikem, E. P., & Abdullahi, A. (2025). Responsible AI deployment in sustainable project execution: Ensuring transparency, carbon efficiency and regulatory alignment. Int. J. Sci. Res. Arch, 14(3), 1686-1705. https://doi.org/10.30574/ijsra.2025.14.3.0920

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. Electronic Markets, 31, 447-464. https://doi.10.1007/s12525-020-00441-4

Timan, T., & Mann, Z. (2021). Data protection in the era of artificial intelligence: Trends, existing solutions and recommendations for privacy-preserving technologies. In The elements of big data value: Foundations of the research and innovation ecosystem (pp. 153-175). Cham: Springer International Publishing. https://doi.10.1007/978-3-030-68176-0_7

Trager, R., Harack, B., Reuel, A., Carnegie, A., Heim, L., Ho, L., Kreps, S., Lall, R., Larter, O., Ó hÉigeartaigh, S., Staffell, S., & Villalobos, J. J. (2023). International governance of civilian AI: A jurisdictional certification approach. arXiv preprint arXiv:2308.15514 .https://doi.org/10.48550/arXiv.2308.15514

Usman, M. (2022). Synergy between professional human resources and artificial intelligence: a review of the consequences of technological innovation. International Journal of Innovation and Business Strategy (IJIBS), 16(1).

Vangala, V. (2022). MLOps in practice: A framework for scalable AI model deployment, monitoring, and retraining. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 13. https://doi.10.5281/zenodo.15570286

Victor, M., Praveenraj, D. D. W., R, S., Alkhayyat, A., & Shakhzoda, A. (2023, January). Cryptography: Advances in secure communication and data protection. In E3S web of conferences (Vol. 399, p. 07010). EDP Sciences. https://doi.10.1051/e3sconf/202339907010

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. Nature communications, 11(1), 233. https://doi.org/10.1038/s41467-019-14108-y

Vujicic, J. (2024). Global AI regulation and its impact on technology business: A comparative legal framework analysis. World Journal of Advanced Research and Reviews, 24(3), 3457-3463. https://doi.10.30574/wjarr.2024.24.3.3911

Vyhmeister, E., & G. Castane, G. (2026). When Industry meets trustworthy AI: A systematic review of AI for Industry 5.0. AI and Ethics, 6(2), 200. https://doi.10.1007/s43681-026-01026-1

Walia, P. (2024). Role and application of artificial intelligence in business. Shodh Sari-An International Multidisciplinary Journal, 3(02), 244-252. https://doi.10.59231/SARI7700

Wenzel, J., Alam, S. U., Schmidt, A., Zhang, H., & Hermanns, H. (2025). A workflow for full traceability of AI decisions. arXiv preprint arXiv:2511.11275. https://doi.10.48550/arXiv.2511.11275

Westover, J. H. (2025). Navigating the AI revolution: A critical framework for organizational transformation. https://doi.10.20944/preprints202508.0310.v1

Wienrich, C., Carolus, A., & Latoschik, M. E. (2023). How to enable sovereign human-AI interactions at work? Concepts of graspable testbeds empowering people to understand and competently Use AI-systems. In Weizenbaum Conference" Practicing Sovereignty: Interventions for Open Digital Futures" (pp. 209-220). DEU. https://doi.org/10.34669/wi.cp/4.20

Wu, C. J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Aga, F., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H., & Hazelwood, K. (2022). Sustainable AI: Environmental implications, challenges and opportunities. In Proceedings of Machine Learning and Systems, 4, 795-813. https://doi.10.48550/arXiv.2111.00364

Zhu, L., Lu, Q., Ding, M., Lee, S. U., & Wang, C. (2026). Designing meaningful human oversight in ai. AI and Ethics, 6(3), 286. https://doi.10.13140/RG.2.2.16624.11528