

Determinants of Employee Trust in AI Decision-Making: A Bibliometric Literature Review and Conceptual Framework

Aldirafi Gani¹, Puji Lestari Hasibuan², Arifah Mustaqimah³, Siti Rokhanah⁴, Farida Yuliaty⁵, Vip Paramarta⁶, Kosasih Kosasih⁷

Universitas Sangga Buana, Bandung, Indonesia^{1,2,3,4,5,6,7}

Aldiraffi@gmail.com¹, pujilestarihasibuan98@gmail.com², arifahibrahim18@gmail.com³, sitirokhanah85jannah@gmail.com⁴, Farida.yuliaty@usbypkp.ac.id⁵, vip@usbypkp.ac.id⁶, kosasih@usbypkp.ac.id⁷



Article History

Received on 13 April 2026

1st Revision on 16 April 2026

2nd Revision on 06 July 2026

3rd Revision on 04 August 2026

Accepted on 05 August 2026

Abstract

Purpose: This study aimed to identify and categorize factors influencing employee trust in AI-based decision-making within organizations and develop a conceptual framework that explains their relationships.

Research Methodology: A literature review combining bibliometric analysis with an in-depth examination of highly cited papers was conducted. A total of 136 articles from Scopus (2021–2025) were analyzed for this study. The ten most-cited papers were selected for thematic synthesis.

Results: Bibliometric mapping using VOSviewer identified publication growth, country distribution, subject areas, keyword co-occurrence, and author collaborations. The thematic synthesis of the ten most-cited papers grouped determinants of trust into AI system, individual, organizational, and task-context factors and emphasized calibrated trust as the core mechanism of trust-building.

Conclusions: Employee trust in AI is not static but a dynamic spectrum that requires calibrated trust. Self-confidence often outweighs trust in AI in determining the acceptance or rejection of AI recommendations.

Limitations: Reliance on Scopus-indexed articles (2021–2025) may have excluded historical perspectives. Qualitative synthesis has not been tested using meta-analysis. Sector and cultural heterogeneity limit the generalizability of the findings.

Contributions: Theoretically, this study integrates the Trust in Automation Model, the Integrative Model of Organizational Trust, and the SOR Model into a unified, dynamic, multilevel framework. Practically, it offers six recommendations for organizations to build sustainable employee trust in AI-assisted decision-making processes.

Keywords: *AI-assisted Decision Making, Employee Trust in AI, Self-confidence, Trust Calibration, XAI Adoption*

How to Cite: Gani, A., Hasibuan, P. L., Mustaqimah, A., Rokhanah, S., Yuliaty, F., Paramarta, V., & Kosasih, K. (2026). Determinants of Employee Trust in AI Decision-Making: A Bibliometric Literature Review and Conceptual Framework. *Studi Ilmu Manajemen dan Organisasi*, 7(3), 105-125.

1. Introduction

Artificial Intelligence (AI)-based business transformation within the human-centric Industry 5.0 paradigm is not understood as a replacement of human roles, but rather as a reconfiguration of work methods that create new jobs and require adjustments to human resource competencies. [Su, Du, Chen, Tian, Ma, and Zhong \(2025\)](#) emphasizes that value creation in AI-based companies requires digital empowerment, intelligence, and stakeholder network engagement, in which AI drives process redesign and task reallocation, altering job composition and required skill profiles. AI has become the most transformative technology of the 21st century, changing how organizations operate, compete, and create values. However, AI's value of AI to company performance depends heavily on adequate governance and knowledge management, given its opaque or non-transparent nature ([Wamba-Taguimdje & Tchatchouang, 2020](#)).

AI capabilities for Human Resource (HR) functions do not exist in isolation but rather require a combination of human resources, managerial practices, and data management ([Chowdhury, Budhwar, & Wood, 2024](#)). AI innovation is data-driven, requiring organizations to make data and analytics competencies integral to their competitive capabilities ([Cho, Choi, & Choi, 2023](#)). Therefore, the success of AI implementation depends heavily on the quality of human capabilities and organizational governance, not just on technological sophistication ([Wamba-Taguimdje & Tchatchouang, 2020](#)). AI now plays a crucial role in organizational decision-making, not only for automating routine tasks but also for strategic managerial decisions such as recruitment, performance appraisals, promotions, resource allocation, and strategic planning, serving as a "decision-making partner" across organizational levels ([Chowdhury et al., 2024](#)).

Combining human intelligence and AI for strategic organizational decisions requires an appropriate design that encompasses role configuration, procedures, and decision governance, not just technical aspects ([Zárate-Torres, Rey-Sarmiento, Acosta-Prado, Gómez-Cruz, & Camargo, 2025](#)). In recruitment practices, the use of AI impacts decision-making processes and operational routines, requiring organizations to manage its use for routine activities ([Prem, Kharbanda, & Mukherjee, 2023](#)). At the user experience level, explainability is a critical design issue because users need information to understand how AI works and the rationale for its recommendations, in line with criticisms of the opacity of AI models that have led to hesitancy in using AI for critical decisions ([Ehsan, Liao, Muller, Riedl, & Weisz, 2021](#)). The adoption of AI in the workplace significantly increases employee productivity and contributes to competitive advantage and organizational performance, with employee engagement as a mediator ([Gusti, Satrianto, Candrianto, Juniardi, & Fitra, 2024](#)). The value of AI depends heavily on employee interaction and utilization of AI systems; therefore, last-mile integration requires a harmonious integration of technology, work processes, and human needs ([Salwei & Carayon, 2022](#)). This implies that organizations must redesign work and competencies to ensure productive human-AI collaboration.

[Weisz, Muller, and Varshney \(2025\)](#) highlighted the tension between augmentation and replacement in human-AI collaboration, where AI can generate 50,000 lines of code in minutes, a task that would take weeks for a human team, creating existential anxiety among developers who question their roles. However, AI value can emerge through augmentation as developers shift to higher-value tasks that require creativity and business judgment. Therefore, AI-based work design requires the orchestration of roles, competencies, and learning processes to maximize augmentation and mitigate the risk of rejection ([Weisz et al., 2025](#)). [Enholtm, Papagiannidis, Mikalef, and Krogstie \(2022\)](#) emphasized that AI capabilities depend not only on technology but also on organizational factors such as employee skills, an innovative culture, top management support, and trust in AI. Many organizations struggle to realize the value of AI because of an overemphasis on technology. Therefore, organizations must strengthen AI capabilities holistically, encompassing technology, organization, and the environment for sustainable AI adoption.

Maturity models are useful for identifying organizational strengths and weaknesses for tiered development, as AI effectiveness requires a holistic view of enterprise capability design and architecture ([Enholtm et al., 2022](#)). In managerial practice, maturity measurement is related to KPI

management and the utilization of analytics or AI as a critical dimension of an organization's digital maturity ([Yablonsky, 2021](#)). The AI transformation gap lies between adoption claims and implementation readiness, which can be detected through maturity assessment instruments, KPIs, and linking AI to strategic objectives. In cross-system integration sectors, such as the AEC industry, the integration of BIM, AI, and IoT faces issues with interoperability, standardization, and a lack of technological maturity. Therefore, the integration gap cannot be addressed simply by adopting tools ([Ouchtout, Zaki, & Idrissi, 2025](#)). Similarly, ERP implementation requires a phased strategy aligned with the organization's processes and data needs ([Miller, 2019](#)). Therefore, an enterprise-level AI agenda requires strategic alignment, a technology-data-process integration architecture, and performance metrics to assess the progress of the transformation.

A relevant analogy can be drawn from the Enterprise Resource Planning (ERP) implementation literature, as noted by [Ashraf and Ali \(2022\)](#), who point out that enterprise system implementation strategies often need to align with an organization's innovative culture, processes, and data needs, rather than simply installing a comprehensive system at once. In the context of AI-assisted decision-making, trust is a critical variable influencing AI acceptance, with ethics being the most important factor, followed by fairness, accountability, and transparency ([Sides, Kbaier, Farrell, & Third, 2024](#)). Without adequate trust, employees are likely to reject or sabotage AI systems. Trust also mediates the relationship between perceived usefulness and AI adoption, either encouraging or hindering it ([Al & Alsmairat, 2023](#)). The challenges of AI implementation are not only technical but also cultural and organizational in nature. A global survey found that 99% of organizations experience at least one challenge, with fear of job loss, loss of autonomy, and resistance to change being the most significant ([Yu, Tian, Chen, Huang, Pan, & Jang, 2025](#)). Trust in human-AI relationships is influenced by 16 related factors, including task characteristics, system autonomy, transparency, and competence. Regulation can influence six key factors, such as task type and transparency, but has limited influence on personal factors such as emotions and the propensity to trust ([Tamò, Guitton, Mayer, & Lutz, 2024](#)). Increasing AI adoption does not necessarily translate into increased trust; the more experienced an organization becomes, the more complex the trust challenges.

The level of trust in AI varies significantly across countries, with developing countries tending to have higher trust and adoption than developed countries ([Gillespie, Lockey, Ward, Macdade, & Hased, 2025](#)). However, high AI adoption does not necessarily translate into adequate governance or responsible use ([Gillespie et al., 2025](#)). In Indonesia, accountability frameworks and algorithmic bias prevention significantly influence trust and responsibility in AI-assisted decision-making; therefore, talent development programs need to consider ethical and accountability dimensions ([Mardiani, Judijanto, & Rukmana, 2023](#)). Therefore, competency development strategies cannot be separated from the design of governance, accountability, and evaluative competencies to ensure that AI truly improves decision quality and productivity.

Based on the background description and research gaps, this study aims to conduct a literature review to identify, analyze, and synthesize the factors that influence employee trust in AI-based decision-making in organizations. Although prior studies have discussed AI adoption, transparency, governance, and human-AI collaboration, the literature remains fragmented in explaining how these determinants interact as a dynamic employee trust framework. This study will answer three questions: What is the publication trend on this topic (by year, country, sector, and keywords)? What factors influence employees' trust in AI in organizational decision-making, and how are they categorized? What conceptual framework can explain the relationship between these factors? This study is expected to provide theoretical contributions by mapping and synthesizing the literature, as well as practical contributions in the form of recommendations for organizations and managers on building and maintaining employee trust in the adoption and implementation of AI for organizational decision-making.

2. Literature Review

2.1 Trust Foundations in Human-AI Relationships

Trust is a multi-dimensional construct. [Mayer et al. \(1995\)](#) defined it as the willingness to accept vulnerability based on positive expectations, with three components: vulnerability, positive expectations, and willingness to take risks. In the context of AI, trust is the belief that an AI system will be reliable, competent, and responsible, even if the user does not fully understand its mechanisms ([Glikson & Woolley, 2020](#)). Unlike human trust, which involves consciousness, intention, and emotion, AI lacks these features, making trust in AI more functional and performance-dependent ([Li et al., 2024](#); [Scharowski et al., 2023](#)). However, advances in AI are blurring these boundaries ([Li et al., 2024](#); [Yam et al., 2023](#)), triggering algorithm appreciation (over-trust in AI) or algorithm aversion, as well as automation bias (over-reliance on AI), depending on the context ([Logg et al., 2019](#); [Parasuraman & Manzey, 2010](#)).

2.2 Trust in Automation and Organizational Trust Models

Researchers have developed various theoretical models to understand the factors that shape trust in AI. One of the most influential is the three-layered model of trust proposed by [Hoff & Bashir \(2015\)](#) in the context of automation. This model distinguishes trust into three layers: dispositional trust, which is an individual's general tendency to trust technology, influenced by personality, culture, and past experiences; situational trust, which depends on the specific task context, such as risk level, task complexity, and workload; and learned trust, which develops from direct experience with the system, including perceptions of its reliability, transparency, and fairness. This model emphasizes that trust is not a static trait but a dynamic construct that changes over time through repeated interactions between users and the system ([Cabiddu et al., 2022](#); [Hoff & Bashir, 2015](#)). In an organizational context, employee trust in AI is likely influenced by all three layers simultaneously, with varying weights depending on individual characteristics, task nature, and organizational experience.

Another relevant model is [Mayer et al. \(1995\)](#) Integrative Model of Organizational Trust, which identifies three key factors influencing perceptions of trust: ability, benevolence, and integrity. Ability refers to the technical competence and performance of an AI system in performing its tasks. Benevolence refers to the extent to which an AI system is perceived to act in the best interest of users rather than harm or replace them. Integrity refers to the consistency and fairness of an AI system in applying the rules and procedures. In the context of AI, integrity is often associated with the absence of algorithmic bias and adherence to ethical principles ([Ferrario et al., 2020](#); [Shin, 2021](#)). Employees who perceive an AI system to be highly capable of supporting their work (rather than replacing it) and operating fairly and consistently tend to have higher levels of trust ([Mayer & Davis, 1999](#); [Shin, 2021](#)).

2.3 Technology Acceptance, Explainability, and Fairness

The Technology Acceptance Model (TAM) emphasizes perceived usefulness and ease of use ([Davis, 1989](#)), but for AI, the trust dimension is more important ([Hoff & Bashir, 2015](#)). AI must demonstrate fair, transparent, and accountable behavior, with explainability and causability as the foundations of trust ([Shin, 2021](#); [Wölker, & Powell, 2021](#); [Shin, 2021](#)). Transparency and explainability are key because black-box systems generate distrust, especially in high-stakes decision-making ([Adadi, & Berrada, 2018](#); [Doshi-Velez, & Kim, 2017](#); [Choubisa, & Choubisa, 2024](#)). Conversely, explainable AI (XAI) increases user trust, although its effectiveness depends on the context and expertise ([Miller, 2019](#); [Poursabzi-Sangdeh, Goldstein, Hofman, & Wallach, 2021](#)). In organizations, employees who receive clear explanations tend to be more trusting ([Vereschak, Bailly, & Caramiaux, 2021](#)).

Trust is highly sensitive to system errors, especially rare but impactful ones ([Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003](#)). A single significant error can drastically reduce trust, necessitating remedial mechanisms (trust violation and repair), such as transparent explanations and assurances of remediation ([Chen, Pan, Li, & Yu, 2022](#)). Perceived fairness is also critical, encompassing distributive, procedural, and interactional justice, particularly for decisions with a high career impact ([Lee, 2018](#); [Robert, Pierce, Marquis, Kim, & Alahmad, 2020](#); [Shin, & Park, 2019](#)). Individual factors, such as positive experiences with AI, enhance initial trust, whereas AI literacy helps calibrate trust,

thereby preventing automation bias or algorithm aversion ([Alruwaili, Alasmari, Niyazi, Youssef, Aifan, & Askland, 2025](#); [Han, & Ko, 2025](#); [Tomsett, Preece, Braines, Cerutti, Chakraborty, Srivastava, Pearson, & Kaplan, 2020](#)). Organizational factors, such as management support, an innovative culture, and AI governance (including human-in-the-loop), increase employee trust ([Birkstedt, Minkkinen, Tandon, & Mäntymäki, 2023](#)). Task context also influences employees' trust in AI for routine, objective, and low-risk tasks, but less for complex, ambiguous, or subjective tasks ([Yang, Li, Qi, Liu, He, & Zhao, 2025](#)). The higher the perceived risk, the lower the trust in AI ([Yu, Tian, Chen, Huang, Pan, & Jang, 2025](#)).

2.4 Psychological Contract and Work Design

Job insecurity and anxiety can undermine trust in AI. When AI is perceived as a threat, employees become defensive and resistant to it. AI-induced human capital obsolescence threatens business sustainability ([Amankwah-Amoah & Appiah, 2025](#)). Conversely, if AI is positioned as an augmentation tool rather than a replacement and is supported by clear communication about role changes and investments in reskilling/upskilling, trust increases. In organizational theory, Social Exchange Theory and Psychological Contract Theory explain that AI adoption alters the psychological contract: employees expect AI to respect their dignity, privacy, and contributions ([Braganza, Chen, Canhoto, & Sap, 2021](#); [Gong, & Sims, 2023](#)). Abuse (excessive monitoring or unfair decisions) undermines trust by [Burrell and Mcandrew \(2023\)](#) and [Gong and Sims \(2023\)](#), whereas transparent, fair, and development-supportive use of AI strengthens the psychological contract and increases trust ([Burrell and Mcandrew, 2023](#); [Soulami, Benchekroun, & Galiulina, 2024](#)).

2.5 Synthesis of Theoretical Contributions

The above theoretical perspectives indicate that employee trust in AI is shaped by technical, cognitive, ethical, and organizational factors. The Trust in Automation model explains trust development through experience, the Organizational Trust model highlights ability, benevolence, and integrity, TAM explains acceptance through usefulness and ease of use, and Social Exchange/Psychological Contract theory describes employee responses to AI-driven workplace changes. Together, these theories show that trust in AI is a multidimensional construct influenced by both technological and organizational factors, although further empirical investigation across different contexts is required.

3. Research Methodology

This study employed a literature review that combined bibliometric analysis with an in-depth examination of highly cited papers to map the intellectual structure and extract key factors of trust ([Donthu et al., 2021](#); [Zhou & Ma, 2025](#); [Zupic & Čater, 2015](#)). Bibliometric analysis uses quantitative metadata ([Pritchard, 1969](#); [Van Eck & Waltman, 2010](#)). The first step was to formulate research questions ([Resnawita & Hendrik, 2023](#)). Based on the background and identification of research gaps, this study establishes the following three research questions (RQs) as can be seen in Table 1.

Table 1. Research question

Order	Research Question (RQ)	Objective
<i>RQ₁</i>	What are the publication trends on employee trust in AI-based decision-making (by year, country, industry sector, and keyword)?	Mapping the intellectual structure and development of this topic over time.
<i>RQ₂</i>	What factors influence employees' trust in AI in the context of organizational decision-making?	Identifying and categorizing the determinants of trust.
<i>RQ₃</i>	What conceptual framework can be developed to explain the relationships between these factors?	A theoretical synthesis was developed to inform further research and practical recommendations.

The next step was to conduct a literature search using bibliographic data sources from the Scopus database. Scopus was chosen for its comprehensive coverage of peer-reviewed literature in the social sciences, business, management, psychology, and computer science, and its advanced filtering features ([Chadegani et al., 2013](#); [Mongeon & Paul-Hus, 2016](#)). As recommended in the practical guide

to bibliometric analysis, selecting the right database is crucial for ensuring the quality and coverage of the literature analyzed ([Donthu et al., 2021](#)).

A literature search was conducted on March 7, 2026, to ensure the consistency of the literature snapshot at a single point in time. The search strategy used keywords in the title, abstract, and keywords fields (TITLE-ABS-KEY) with the following Boolean string:

TITLE-ABS-KEY (("employee trust" OR "trust in AI" OR "trust in automation" OR "algorithmic trust") AND ("artificial intelligence" OR "ai" OR "machine learning" OR "decision support system") AND ("decision making" OR "organizational decision" OR "hr decision" OR "recruitment") AND NOT ("clinical" OR "diagnosis" OR "patient" OR "medical" OR "health care" OR "mri" OR "imaging" OR "radiology" OR "surgery" OR "disease" OR "therapy" OR "treatment" OR "hospital" OR "physician" OR "nurse")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2024) OR LIMIT-TO (PUBYEAR , 2025)) AND (LIMIT-TO (LANGUAGE , "English"))

Because this research approach is a bibliometric analysis followed by an in-depth review of highly cited papers, article selection was performed in stages. The first stage used inclusion and exclusion criteria ([Resnawita & Hendrik, 2023](#)): Scopus-indexed articles published between 2021 and 2025 that discussed trust as the primary construct and AI as the research object, and in English. The exclusion criteria were clinical/medical contexts, proceedings/book chapters/editorials, or lack of full text. Of the 136 articles found, all met the inclusion criteria because of the limited scope of the search string. In the second stage, the 10 most-cited articles were selected for an in-depth review, as highly cited papers have theoretical influence and serve as pillars of the literature ([Donthu et al., 2021](#); [Hoang, 2025](#); [Van Eck & Waltman, 2010](#); [Zhou & Ma, 2025](#)).

The use of the ten most-cited articles was justified as an influence-based purposive sample because highly cited papers typically represent conceptual anchors and methodological references in the bibliometric literature. To maintain reproducibility, the search date, Boolean string, inclusion and exclusion criteria, Scopus export metadata, VOSviewer mapping parameters, citation ranking, and thematic coding stages were documented so that the selection and synthesis processes could be repeated using the same database snapshot.

Data analysis was performed in two stages ([Hoang, 2025](#); [Li et al., 2024](#)). First, bibliometric analysis mapped publication trends, country distribution, subject areas, keyword co-occurrence (VOSviewer), and author collaboration ([Donthu et al., 2021](#); [Zupic & Čater, 2015](#)). Second, a thematic synthesis by ([Thomas & Harden, 2008](#)) of the 10 most-cited articles involved line-by-line coding, descriptive themes (transparency, reliability, fairness), and analytical themes (AI system, individual, organizational, task factors), leading to a conceptual framework.

4. Results and Discussions

4.1 Bibliometric Analysis

A bibliometric analysis was conducted to map the intellectual structure and publication trends regarding the factors influencing employee trust in AI-based decision-making in organizations. Table 2 presents the distribution of documents by publication year.

Table 2. Distribution of articles by year

Year	Number of Documents
2021	4
2022	10
2023	14
2024	38
2025	70

Year	Number of Documents
Total	136

Table 2 shows the distribution of documents by year in this bibliometric analysis. The year 2025 had the largest number of documents, with 70 articles, followed by 2024, with 38. The years 2023, 2022, and 2021 had 14, 10, and 4 articles relevant to the analysis topic, respectively. A total of 136 articles were processed in this analysis from the Scopus database (both free and paid versions). The increasing number of publications from year to year indicates that the issue of employee trust in AI-based decision-making is gaining increasing attention from researchers, along with AI widespread adoption in various organizational sectors. Table 3 presents the distribution of articles by country.

Table 3. Distribution of articles by country

No.	Country	Number of Documents
1	United States	36
2	China	15
3	India	13
4	Germany	12
5	Canada	9
6	United Kingdom	7
7	South Korea	6
8	Saudi Arabia	6
9	Italy	6
10	Australia	6
11	Spain	5
12	Japan	5
13	Greece	5
14	Vietnam	4
15	Switzerland	4
16	Netherlands	4
17	Czech Republic	4

Table 3 shows that the United States was the largest contributor, with 36 articles, followed by China (15), India (13), Germany (12), Canada (9), and the United Kingdom (7). Countries with moderate contributions (4–6 articles) included South Korea, Saudi Arabia, Italy, Australia, Spain, Japan, Greece, Vietnam, Switzerland, the Netherlands, and the Czech Republic. The remaining 36 countries contributed one to three articles, with Indonesia contributing only one. Indonesia's low contribution indicates that research on employee trust in AI within organizations remains limited, thereby opening broad research opportunities for Indonesian researchers in the future. The dominance of the United States and China reflects the research gap between developed and developing countries on this topic.

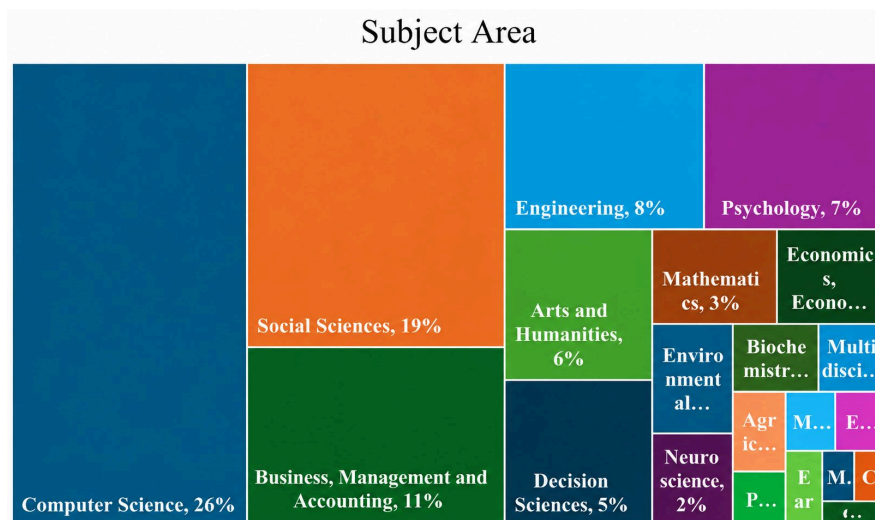


Figure 1. Results by subject area

Figure 1 shows that Computer Science is the subject area with the largest percentage (26%), followed by Social Sciences (19%), and Business, Management, and Accounting (11%). The dominance of Computer Science reflects the importance of the technical aspects of AI, while significant contributions from Social Sciences, Business, and Psychology (7%) confirm the multidisciplinary nature of research on employee trust in AI. The success of AI implementation in organizations depends not only on technological sophistication but also on perceptions, behaviors, and social dynamics in the workplace. Other subject areas, such as Economics, Econometrics, and Finance (2%), Environmental Science (2%), neuroscience (2%), and other subjects under 1%, also make significant contributions.

Overall, the distribution of these subject areas confirms that research on employee trust in AI in organizational decision-making cannot be approached from a purely technical viewpoint. Successful AI implementation in organizations relies heavily on understanding human factors, including perceptions of trustworthiness, fairness, transparency, and accountability. This aligns with the concept of Industry 5.0, which emphasizes a human-centric approach to technology development.

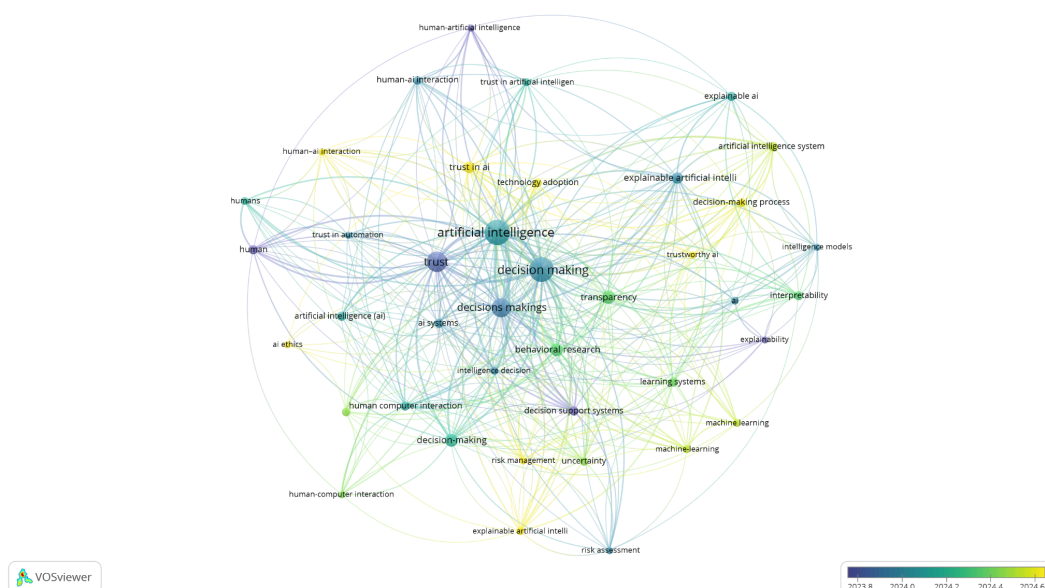


Figure 2. VOSviewer results based on keywords

Figure 2 presents a VOSviewer bibliometric network map illustrating the relationships among key concepts in the research on employee trust, artificial intelligence, and organizational decision-making. AI and trust emerge as central concepts connected with explainability, transparency, fairness, reliability, job insecurity, organizational culture, and human-AI collaboration. The visualization also highlights emerging topics such as explainable AI (XAI), perceived fairness, and trust repair, providing insights into the factors shaping employee trust in AI-based decision-making. The VOSviewer results by author are shown in Figure 3.

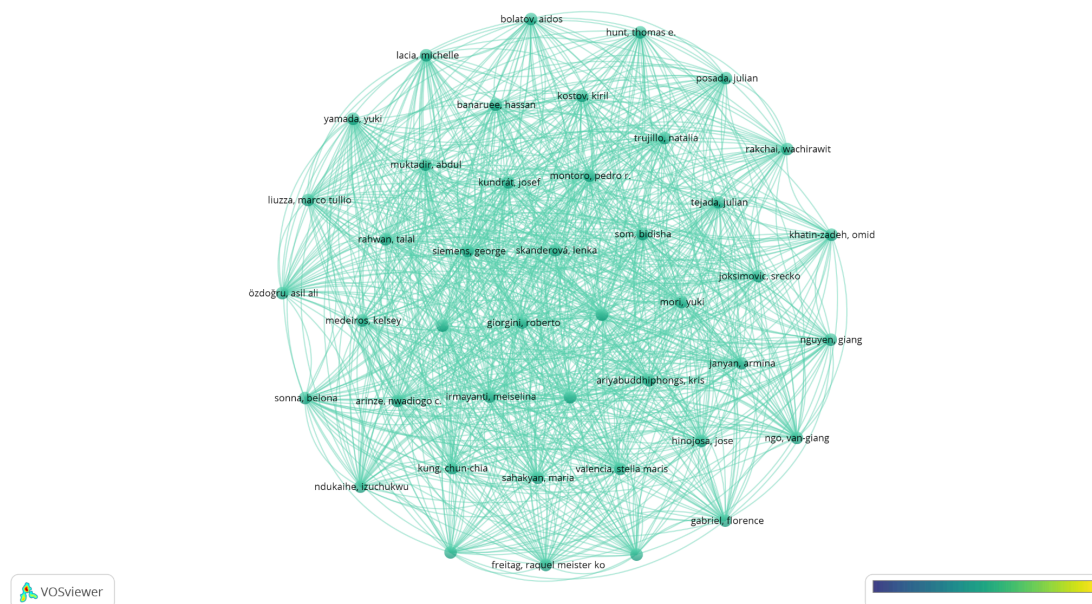


Figure 3. VOSviewer results based on author

Figure 3 presents a VOSviewer bibliometric network map that illustrates the author collaboration patterns in the reviewed literature. The interconnected network indicates strong collaboration among researchers, with centrally positioned authors as key contributors. The dominant turquoise color suggests that most publications are relatively recent, reflecting the growing collective research efforts in artificial intelligence, trust, and decision-making. The main findings of this bibliometric analysis are discussed in the Discussion section.

Analytically, these bibliometric patterns indicate that employee trust in AI has shifted from a primarily technical discussion to an integrated socio-technical issue. The growth of publications after 2023, the dominance of computer science together with social science and management, and the keyword links among explainability, fairness, reliability, job insecurity, and human-AI collaboration imply that trust must be theorized as a multilevel construct connecting system design, employee psychology, organizational governance, and task risk.

Table 4. Ten most cited documents from 136 articles on employee trust, trust in AI and decision making (2021-2025)

No	Article Title	Writer	Year	Number of Citations
1	Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence	Ali, Sajid; Abbod, Tamer; El-Sappagh, Shaker; Muhammad, Khan; Alonso-Moral, Jose M.; Confalonieri, Roberto; Guidotti, Riccardo; Del Ser, Javier; Diaz-Rodriguez, Natalia; Herrera, Francisco	2023	1,241

No	Article Title	Writer	Year	Number of Citations
2	Transparency and the Black Box Problem: Why We Do Not Trust AI	Von Eschenbach, Warren J.	2021	570
3	Human confidence in artificial intelligence in itself: The evolution and impact of confidence on adoption of AI advice	Chong, Leah; Zhang, Guanglu; Goucher-Lambert, Kosa; Kotovsky, Kenneth; Cagan, Jonathan	2022	199
4	How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies	Vereschak, Oleksandra; Bailly, Gilles; Caramiaux, Baptiste	2021	167
5	Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI	Klingbeil, Artur; Grüttner, Cassandra; Schreck, Philipp	2024	163
6	Analyzing the adoption challenges of the internet of things (IoT) and artificial intelligence (AI) for smart cities in China	Wang, Ke; Zhao, Yafei; Gangadhari, R.K.; Li, Zixing	2021	110
7	Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation	Zhang, Guanglu; Chong, Leah; Kotovsky, Kenneth; Cagan, Jonathan	2023	104
8	Rise of the machines: Delegating decisions to autonomous AI	Candrian, Cindy; Scherer, Anne	2022	99
9	The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending	Dikmen, Murat; Burns, Catherine	2022	97
10	Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort	Yu, Liangru; Li, Yi	2022	96

Table 4 presents the 10 articles with the most citations on employee trust, trust in AI, and decision-making published between 2021 and 2025. The article titled "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence" by [Ali, Abuhmed, El-Sappagh, Muhammad, Alonso-Moral, Confalonieri, Guidotti, Díaz-Rodríguez, and Herrera \(2023\)](#) ranked first with 1,241 citations. This article comprehensively examines 410 critical articles to map the development of XAI as a solution to the black box problem in AI. [Ali et al. \(2023\)](#) structured the XAI taxonomy into four main axes: data explainability, model explainability, post-hoc explainability, and assessment of the explanations. The main results show that despite the rapid development of various XAI methods, there is no standard protocol for evaluating the quality of the resulting explanations, and transparency alone is insufficient to build trust without guarantees of fairness, accountability, and robustness. [Ali et al. \(2023\)](#) emphasized that explainability is a primary prerequisite for trustworthy AI, especially in critical and high-risk decision-making contexts.

The second place is held by [Von \(2021\)](#), with 570 citations for "Transparency and the Black Box Problem: Why We Do Not Trust AI," which argues philosophically that transparency is a necessary condition for trust in AI. The article contends that black-box deep-learning AI systems cannot be morally trusted because they fail to meet transparency requirements; moral trust demands understanding how and why a decision is made, not merely knowing that it is reliable. [Von \(2021\)](#) proposed a sociotechnical systems approach, situating AI within a broader framework of designers, users, and oversight bodies to form a distributed web of trust.

Third place went to [Chong, Zhang, Goucher-Lambert, Kotovsky, and Cagan \(2022\)](#) with 199 citations for "Human Confidence in Artificial Intelligence and in Themselves," a cognitive study with quantitative modeling. Using a chess puzzle task with dynamically varying AI performance, they found that human self-confidence, rather than trust in AI, was the primary determinant of whether to accept or reject AI recommendations. This reveals a psychological dynamic in which humans often misattribute blame to themselves when AI performs poorly, entering a vicious cycle of dependence on the incompetent AI ([Chong et al., 2022](#)).

[Vereschak, Bailly, and Caramiaux \(2021\)](#), in a systematic review of 83 empirical papers, highlighted the lack of standardized trust evaluation protocols (167 citations) and identified three key overlooked elements: vulnerability, positive expectations and attitude. They also found significant design and measurement variability, compromising validity and replicability. Meanwhile, [Klingbeil, Grützner, and Schreck \(2024\)](#), with 163 citations, experimentally demonstrated using a modified trust game that merely knowing that advice comes from AI causes overreliance, even when it contradicts contextual information and harms third parties, confirming that overtrust can be as harmful as distrust.

[Klingbeil et al. \(2024\)](#) received 163 citations for their experimental study using a modified trust game with 402 participants, comparing AI (ChatGPT) advice versus expert advice. They found that simply knowing that advice came from AI led to overreliance, even when it contradicted contextual information and harmed third parties, resulting in inefficient outcomes for both decision-makers and their partners. These findings confirm that overtrust can be as harmful as distrust ([Klingbeil et al., 2024](#)).

Furthermore, the sixth article by [Wang, Zhao, Gangadhari, and Li \(2021\)](#) received 110 citations for their study, which identified 10 major AI and IoT adoption challenges in China's smart cities using PRISMA and DEMATEL methods. The main causal challenges were a lack of infrastructure, inadequate funding, cybersecurity risks, and a lack of trust in AI and IoT, which significantly compounded other issues, such as waste management and a shortage of professionals. [Wang et al. \(2021\)](#) emphasized that a lack of trust is a causal factor that slows overall AI adoption.

The seventh article by [Zhang, Chong, Kotovsky, and Cagan \(2023\)](#) received 104 citations for their study on teammate identity ("human" vs. AI) and performance (high vs. low) in human-AI cooperation. The results showed that humans trusted AI more than they trusted one another, accepting AI decisions more often. Teammate performance significantly affected joint performance, but identity did not. Participants deceived into believing that AI was human showed lower behavioral trust but higher confidence in low-performing AI due to social pressure ([Zhang et al., 2023](#)).

The eighth article by [Candrian and Scherer \(2022\)](#) received 99 citations for their three experiments comparing delegation to AI versus human agents. They found that people prefer to delegate decisions to AI rather than to humans, especially in loss domains (e.g., potential harm). The perception that humans can act intentionally creates uncertainty about agents' intentions, leading users to demand a higher reward premium for delegating to humans. Process transparency increased delegation to humans but not to AI ([Candrian & Scherer, 2022](#)).

The ninth article by [Dikmen and Burns \(2022\)](#) received 97 citations for their peer-to-peer lending simulator experiment with 40 participants, which examined the effects of domain knowledge on trust, reliance, and task performance. Participants with domain knowledge relied less on the AI assistant

when it made mistakes and reported lower trust; however, the overall investment performance was unaffected. [Dikmen and Burns \(2022\)](#) concluded that domain knowledge helps nonexpert users develop appropriate levels of trust and reliance.

The final article with 96 citations is by [Yu and Li \(2022\)](#) and is titled "Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort." [\(Yu & Li, 2022\)](#) used an experimental vignette method with 235 participants with work experience to investigate how transparency in AI decision-making affects employee trust. The results showed that AI transparency (vs. non-transparency) led to increased perceptions of transparency, which in turn increased both perceptions of effectiveness (which promotes trust) and discomfort (which hinders trust). [Yu and Li \(2022\)](#) found a parallel multiple-mediating effect, which may partially explain the inconsistent findings in previous research on the relationship between AI transparency and trust. This study also identified a chain mediation pathway through perceived transparency and effectiveness, as well as through perceived transparency and discomfort.

Overall, Table 4 shows that research on trust in AI in organizational decision-making is growing rapidly and is centered around three main clusters: first, explainability as a solution to the black box problem, which is the foundation for transparency and a prerequisite for trustworthy AI; second, methodological evaluation and calibration of trust, including the development of standard protocols and an understanding of the dynamics of self-confidence and overreliance; and third, behavioral and contextual factors that shape employee trust, such as agent identity (human vs. AI), user domain knowledge, system transparency, the context of technology adoption (gain vs. loss), and the dual mediating effects of effectiveness and discomfort. The synthesis of these ten articles strengthens the argument that building calibrated trust in AI in organizational environments requires a holistic approach that goes beyond mere technical sophistication to encompass transparency, humane interface design, responsible governance, and improved user literacy and experiences.

4.2 Discussion

RQ₂: Factors influencing employee trust in AI in the context of organizational decision making

Based on the synthesis of the 10 highest-cited articles summarized above, the factors influencing employees' trust in AI in the context of organizational decision-making can be grouped into four main categories: AI system-related factors, individual employee-related factors, organization-related factors, and task-context-related factors.

First, the factors related to AI systems include transparency, explainability, reliability, accuracy, perceived fairness, and the absence of algorithmic bias. [Ali, Abuhmed, El-Sappagh, Muhammad, Alonso-Moral, Confalonieri, Guidotti, Díaz-Rodríguez, and Herrera \(2023\)](#) assert that explainability is a key prerequisite for trustworthy AI; without the ability to explain how and why a decision is made, users cannot develop calibrated trust. [Von \(2021\)](#) philosophically reinforces this argument by stating that transparency is a necessary condition for trust in the moral context. [Yu and Li \(2022\)](#) found that transparency has a dual effect: it increases perceived effectiveness (which promotes trust) while simultaneously increasing discomfort (which hinders trust). [Klingbeil, Grützner, and Schreck \(2024\)](#) showed that perceived reliability is often more influential than objective reliability, such that the knowledge that advice comes from AI is sufficient to lead to overreliance, even when the advice is incorrect.

Second, employee-related factors include self-confidence, domain knowledge, prior experience with AI, and a disposition toward technology. [Chong, Zhang, Goucher-Lambert, Kotovsky, and Cagan \(2022\)](#) reported counterintuitive findings: human self-confidence, rather than trust in AI, was the primary determinant of whether individuals accepted or rejected AI recommendations. [Dikmen and Burns \(2022\)](#) demonstrated that domain knowledge helps employees critically evaluate AI recommendations and reduces overreliance on faulty AI. [Zhang, Chong, Kotovsky, and Cagan \(2023\)](#) added that initial expectations of AI capabilities (based on prior knowledge) influence the level of disappointment when AI performs poorly, which, in turn, influences trust.

Third, organizational factors include top management support, organizational culture, AI governance policies, communication quality, and training. [Wang, Zhao, Gangadhari, and Li \(2021\)](#) identified a lack of infrastructure, a shortage of skilled professionals, and a collective lack of trust as key barriers to AI adoption at the organizational level. [Vereschak, Bailly, and Caramiaux \(2021\)](#) highlighted that the absence of standardized trust evaluation protocols makes it difficult for organizations to systematically measure and intervene in employee trust. [Candrian and Scherer \(2022\)](#) added that transparency in organizational processes can improve the delegation of decisions to human agents but not to AI, indicating that organizational interventions need to be tailored to the type of agent involved.

Fourth, task context-related factors include perceived risk level, task complexity, decision domain (gain vs. loss), and the desired level of autonomy. [Candrian and Scherer \(2022\)](#) found that employees were more willing to delegate potentially loss-inducing decisions (loss domain) to AI than to humans, as in loss situations, the desire to maintain control is lower when dealing with AI perceived as neutral. [Dikmen and Burns \(2022\)](#) showed that in complex tasks such as credit scoring, domain knowledge helps employees identify discrepancies between AI recommendations and their own independent judgment. [Klingbeil et al. \(2024\)](#) added that in situations involving third parties (such as in trust games), overreliance on AI can be detrimental not only to decision-makers but also to their interaction partners.

RQ₃: Conceptual framework of the relationship between factors influencing employee trust in AI in organizations

Based on a synthesis of the 10 most highly cited articles, this study develops a multilevel, dynamic, integrative conceptual framework that explains the relationships among factors influencing employee trust in AI-based decision-making. This framework integrates three main theories: Trust in Automation Model, which distinguishes between dispositional, situational, and learned trust; Integrative Model of Organizational Trust ([Mayer et al., 1995](#)), which emphasizes ability, benevolence, and integrity; and the Stimulus-Organism-Response (SOR) Model adapted by [Yu and Li \(2022\)](#) to explain how external stimuli (transparency) influence internal states (perception, effectiveness, discomfort), which in turn generate responses (trust). This conceptual framework consists of four layers that are causally and dynamically interrelated: the antecedent layer (external and contextual triggers), the psychological mechanism layer (cognitive and affective mechanisms), the trust calibration layer, and the behavioral outcomes and feedback loop layer.

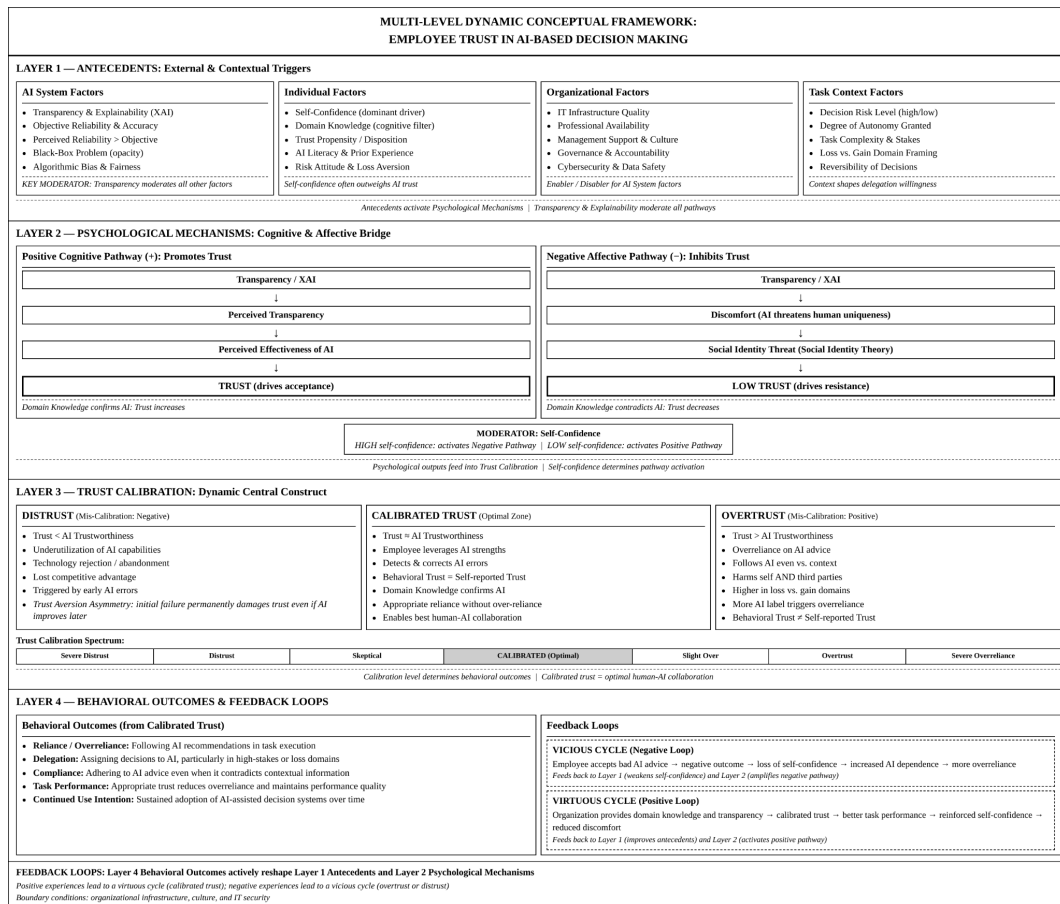


Figure 4. Multi-level dynamic conceptual framework: Employee trust in AI-based decision making

Processed by researchers based on a synthesis of [Ali et al. \(2023\)](#), [Candrian and Scherer \(2022\)](#), [Chong et al. \(2022\)](#), [Dikmen and Burns \(2022\)](#), [Klingbeil et al. \(2024\)](#), [Vereschak et al. \(2021\)](#), Von [Von \(2021\)](#), [Wang et al. \(2021\)](#), [Yu and Li \(2022\)](#), and [Zhang et al. \(2023\)](#).

Figure 4 shows the relationships between the layers. The first layer, the antecedents (external and contextual triggers), consists of the four groups of factors identified in RQ_2 . However, unlike a simple linear model, this conceptual framework emphasizes that these four groups of factors do not operate independently; they interact with each other and are moderated by situation-specific characteristics. [Ali et al. \(2023\)](#) and Von [Von \(2021\)](#) consistently emphasized that transparency and explainability are essential prerequisites that moderate the influence of other factors. Without transparency, factors such as reliability and organizational support are insufficient to build calibrated trust. [Klingbeil et al. \(2024\)](#) add that perceived reliability is often more influential than objective reliability, especially when employees lack mechanisms to verify AI accuracy. [Wang et al. \(2021\)](#) show that organizational factors, such as infrastructure and professional availability, serve as enablers or disablers that moderate the extent to which AI system factors can be implemented effectively.

The second layer, psychological mechanisms (cognitive and affective), serves as a bridge between antecedents and trust calibration. [Yu and Li \(2022\)](#) identified two parallel, opposing pathways (parallel multiple mediating effects). The first pathway is positive and cognitive: transparency increases perceived transparency, which in turn increases perceived AI effectiveness, thereby encouraging trust. The second pathway is a negative affective pathway: transparency also increases discomfort because AI is perceived as too human-like and threatens the uniqueness of human work (social identity theory), thereby hindering trust in AI. The self-confidence mechanism discovered by [Chong et al. \(2022\)](#) functions as a moderator in both pathways: employees with high self-confidence tend to activate the negative affective pathway (rejecting AI), whereas those with low self-confidence tend to activate the positive cognitive pathway (accepting AI usage). [Dikmen and Burns \(2022\)](#)

demonstrated that domain knowledge serves as a cognitive filter, helping employees assess whether AI recommendations align with their understanding of the task. When domain knowledge confirms the AI, trust increases; when it contradicts, trust decreases. [Candrian and Scherer \(2022\)](#) added that the perceived intentional capacity of agents, which is lower for AI than for humans, influences perceptions of social risk and control, which in turn influences willingness to delegate decisions.

The third layer, trust calibration, is a central and dynamic construct rather than a static one. Unlike previous models that treat trust as a dichotomous variable (trust vs. distrust), this framework adopts the concept of calibrated trust from [Vereschak et al. \(2021\)](#) and [Zhang et al. \(2023\)](#). Calibrated trust is defined as the alignment between an employee's level of trust and an AI system's objective trustworthiness. Well-calibrated trust exists when employees are sufficiently trusting to utilize AI capabilities but skeptical enough to detect and correct AI errors. [Zhang et al. \(2023\)](#) distinguished between behavioral trust (the act of following AI recommendations) and self-reported trust (subjective assessments of AI competence and helpfulness) and found that these two dimensions do not always align. For example, employees who are deceived into thinking that AI is human exhibit lower behavioral trust but higher self-reported trust toward low-performing AI due to social pressure. [Klingbeil et al. \(2024\)](#) showed that overtrust is as harmful as distrust, as both lead to miscalibration, which is detrimental to performance. [Chong et al. \(2022\)](#) added that trust calibration is influenced by the accumulation of experience over time, where early AI errors can permanently damage trust, even if AI performance later improves (loss aversion asymmetry).

Finally, the fourth layer, behavioral outcomes and feedback loops, includes reliance, delegation, compliance, task performance, and continued use intentions. [Klingbeil et al. \(2024\)](#) found that overreliance on AI is often detrimental to both employees and third parties, as they follow AI's incorrect advice despite contextual information indicating otherwise. [Candrian and Scherer \(2022\)](#) showed that decision delegation to AI is higher in the loss domain than in the gain domain, as employees feel more comfortable relinquishing control when the risk is negative (prospect theory). [Dikmen and Burns \(2022\)](#) found that overall task performance does not always improve when trust is high; rather, appropriate trust rooted in domain knowledge can reduce overreliance without diminishing performance.

Dynamic feedback loops are a key characteristic of the framework. Unlike linear models, behavioral outcomes in the fourth layer actively influence antecedents and psychological mechanisms, thereby creating cycles that can be virtuous or vicious. [Chong et al. \(2022\)](#) identified a vicious cycle in which employees who follow poor AI advice and receive negative feedback lose confidence, thereby increasing their reliance on AI use. Conversely, providing domain knowledge and transparency ([Dikmen and Burns, 2022](#); [Yu and Li, 2022](#)) enables calibrated trust, better performance, and strengthened self-confidence, thus creating a virtuous cycle. [Vereschak et al. \(2021\)](#) added that delayed feedback in contexts such as investment and recruitment complicates trust calibration.

The integration of organizational and cultural contexts adds another dimension. [Wang et al. \(2021\)](#) showed that in developing countries, infrastructure, staffing, and cybersecurity risks act as boundary conditions that moderate interlayer relationships; weak IT infrastructure can breed distrust, even for transparent AI. [Candrian and Scherer \(2022\)](#) found that perceptions of agents' intentions vary across cultures, affecting the social risk premium for delegating decisions.

Overall, this conceptual framework is both prescriptive and dynamic. To build appropriate employee trust, organizations must simultaneously increase AI transparency and explainability ([Ali et al., 2023](#); [Von, 2021](#)), provide domain knowledge and AI literacy training ([Dikmen and Burns, 2022](#); [Yu and Li, 2022](#)), manage trust through constructive feedback ([Chong et al., 2022](#)), design tasks to minimize overreliance ([Candrian and Scherer, 2022](#); [Klingbeil et al., 2024](#)), build infrastructure and governance ([Wang et al., 2021](#)), and adopt standardized trust evaluation protocols ([Vereschak et al., 2021](#)). The framework also identifies emotional discomfort and self-confidence as overlooked psychological mechanisms, opening a new research agenda and contributing to a more holistic theory of Human-AI trust.

5. Conclusions

5.1 Conclusion

This literature review found that employee trust in AI-based decision-making is influenced by four interacting factor groups: AI system factors (transparency, explainability, and reliability), individual factors (self-confidence, domain knowledge, and trust disposition), organizational factors (infrastructure, governance, and management support), and task context factors (risk, autonomy, and complexity). Transparency and explainability emerge as central prerequisites that shape the effectiveness of other trust-related factors.

As a theoretical contribution, this study proposes a four-layer framework comprising antecedents, psychological mechanisms, trust calibration, and behavioral outcomes by integrating the Trust in Automation Model, Integrative Model of Organizational Trust, and SOR Model. The findings demonstrate that trust in AI is a dynamically calibrated spectrum rather than a static dichotomy, where both over-trust and distrust can negatively influence performance, and employee self-confidence represents a critical psychological mechanism in shaping AI-related decision-making.

As a managerial contribution, this study highlights that organizations should develop appropriate trust in AI through six integrated interventions: transparency, AI literacy training, constructive feedback, task design to minimize overreliance, infrastructure and governance support, and standardized evaluation protocols. Therefore, successful AI implementation requires a holistic, sustainable, and human-oriented trust ecosystem that extends beyond technological sophistication.

5.2 Research Limitations

This study has five limitations: reliance on Scopus-indexed articles (2021–2025) potentially overlooks historical perspectives and other databases; the four-factor framework is still qualitative and has not been meta-analytically tested; sector and cultural heterogeneity limits generalizability because specific contexts influence trust; most articles focus more on resistance factors than measuring the impact of specific interventions such as human-in-the-loop or AI literacy; the dominance of cross-sectional surveys limits long-term causal inference.

5.3 Suggestions & Directions for Future Research

Based on the findings and limitations of this study, several future research agendas should be developed. First, causal evaluation designs are needed that link trust-building programs, such as AI literacy training or explainable AI policies, to organizational outcomes through quasi-experimental and longitudinal studies. Second, given the dominance of self-confidence over trust in AI, multi-level models need to be tested to understand the interactions between talent management, organizational culture, and individual characteristics. Third, research needs to validate the use of people analytics across industries and examine AI decision-making appeal mechanisms and accountability policies. Fourth, cross-cultural comparative studies between high- and low-trust countries are needed to identify the best practices for developing countries such as Indonesia. Finally, exploration of feedback design and task framing is needed to appropriately calibrate trust, prevent over-reliance, and build stronger ethical guidelines.

Acknowledgement

The authors would like to thank all parties who supported and contributed to the completion of this study. Their assistance and encouragement are greatly appreciated.

Author Contributions

AG, PLH, AM, SR, FY, VP and KK contributed to the conceptualization, methodology, analysis, writing, and review of the manuscript.

References

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>

- Al Qahtani, E., & Alsmairat, M. (2023). Assisting artificial intelligence adoption drivers in human resources management: A mediation model. *Acta Logistica*, 10(1), 141-150. <https://doi.org/10.22306/al.v10i1.371>
- Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99. <https://doi.org/10.1016/j.inffus.2023.101805>
- Alruwaili, R. F., Alasmari, A. A., Niyazi, H. T., Youssef, I. K., Aifan, H., & Asklany, S. A. (2025). Modeling public trust in AI cognitive capabilities using statistical and machine learning approaches. *Scientific Reports*, 15(1), 39856. <https://doi.org/10.1038/s41598-025-23447-4>
- Amankwah-Amoah, J., & Appiah, G. (2025). Unmasking the silent threat: AI-induced human capital obsolescence and business failure. *Journal of Managerial Psychology*, 66(1), 5-22. <https://doi.org/10.1108/JMP-05-2024-0385>
- Ashraf, Y., & Ali, F. (2022). How innovative organization culture affects firm performance in the wake of enterprise resource planning? Evidence from energy- and non-energy-sector firms in Pakistan. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.991319>
- Birkstedt, T., Minkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, 33(7), 133-167. <https://doi.org/10.1108/INTR-01-2022-0042>
- Braganza, A., Chen, W., Canhoto, A., & Sap, S. (2021). Productive employment and decent work: The impact of AI adoption on psychological contracts, job engagement and employee trust. *Journal of Business Research*, 131, 485–494. <https://doi.org/10.1016/j.jbusres.2020.08.018>
- Burrell, D. N., & Mcandrew, I. (2023). Exploring the ethical dynamics of the use of Artificial Intelligence (AI) in hiring in healthcare organizations. *Land Forces Academy Review*, 28(4), 309-321. <https://doi.org/10.2478/raft-2023-0037>
- Cabiddu, F., Moi, L., Patriotta, G., & Allen, D. G. (2022). Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. *European Management Journal*, 40(5), 685-706. <https://doi.org/10.1016/j.emj.2022.06.001>
- Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*, 134. <https://doi.org/10.1016/j.chb.2022.107308>
- Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ebrahim, N. A. (2013). A comparison between two main academic literature collections: Web of science and scopus databases. *Asian Social Science*, 9(5), 18. <https://doi.org/10.5539/ass.v9n5p18>
- Chen, A., Pan, Y., Li, L., & Yu, Y. (2022). Are you willing to forgive AI? Service recovery from medical AI service failure. *Industrial Management & Data Systems*, 122(11), 2540-2557. <https://doi.org/10.1108/IMDS-12-2021-0801>
- Cho, W., Choi, S., & Choi, H. (2023). Human resources analytics for public personnel management: Concepts, cases, and caveats. *Administrative Sciences*, 13(2), 41. <https://doi.org/10.3390/admsci13020041>
- Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., & Cagan, J. (2022). Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior*, 127. <https://doi.org/10.1016/j.chb.2021.107018>
- Choubisa, S. L., & Choubisa, P. (2024). Mortality in neonate bats infested by ectoparasite bugs in India; humans are at risk too. *Environment and Public Health Research*, 2(2), 1778. <https://doi.org/10.59400/ephr1778>
- Chowdhury, S., Budhwar, P., & Wood, G. (2024). Generative artificial intelligence in business: Towards a strategic human resource management framework. *British Journal of Management*, 35(4), 1680-1691. <https://doi.org/10.1111/1467-8551.12824>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. <https://doi.org/10.2307/249008>

- Dikmen, M., & Burns, C. (2022). The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human-Computer Studies*, 162. <https://doi.org/10.1016/j.ijhcs.2022.102792>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning (version 2). *arXiv*. <https://doi.org/10.48550/ARXIV.1702.08608>
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697–718. [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)
- Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in AI systems. Paper presented at Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.
- 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.org/10.1007/s10796-021-10186-w>
- Ferrario, A., Loi, M., & Viganò, E. (2020). In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions. *Philosophy & Technology*, 33(3), 523–539. <https://doi.org/10.1007/s13347-019-00378-3>
- Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025). Trust, attitudes and use of artificial intelligence: A global study 2025 (p4974511 Bytes). The University of Melbourne. <https://doi.org/10.26188/28822919>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Gong, B., & Sims, R. L. (2023). Psychological contract breach during the pandemic: How an abrupt transition to a work from home schedule impacted the employment relationship. *Journal of Business Research*, 154. <https://doi.org/10.1016/j.jbusres.2022.08.023>
- Gusti, M. A., Satrianto, A., Candrianto, Juniardi, E., & Fitra, H. (2024). Artificial intelligence for employee engagement and productivity. *Problems and Perspectives in Management*, 22(3), 174–184. [https://doi.org/10.21511/ppm.22\(3\).2024.14](https://doi.org/10.21511/ppm.22(3).2024.14)
- Han, J., & Ko, D. (2025). Trust formation, error impact, and repair in human–AI financial advisory: A dynamic behavioral analysis. *Behavioral Sciences*, 15(10), 1370. <https://doi.org/10.3390/bs15101370>
- Hoang, A.-D. (2025). Evaluating bibliometrics reviews: A practical guide for peer review and critical reading. *Evaluation Review*, 49(6), 1074–1102. <https://doi.org/10.1177/0193841X251336839>
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>
- Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI: An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, 160. <https://doi.org/10.1016/j.chb.2024.108352>
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. <https://doi.org/10.1177/2053951718756684>
- 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*. <https://doi.org/10.3389/fpsyg.2024.1382693>

- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Mardiani, E., Judijanto, L., & Rukmana, A. Y. (2023). Improving trust and accountability in AI systems through technological era advancement for decision support in Indonesian manufacturing companies. *West Science Interdisciplinary Studies*, 1(10), 1019-1027. <https://doi.org/10.58812/wsis.v1i10.301>
- Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, 84(1), 123-136. <https://doi.org/10.1037/0021-9010.84.1.123>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, 20(3), 709-734. <https://doi.org/10.2307/258792>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213-228. <https://doi.org/10.1007/s11192-015-1765-5>
- Ouchtout, W., Zaki, S., & Idrissi Gartoumi, K. (2025). Advancing digital innovation in the AEC industry: A bibliometric analysis on the BIM, AI and IoT integration. *IOP Conference Series: Earth and Environmental Science*. <https://doi.org/10.1088/1755-1315/1499/1/012010>
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(3), 381-410. <https://doi.org/10.1177/0018720810376055>
- Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W., & Wallach, H. (2021). Manipulating and measuring model interpretability. Paper presented at Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21.
- Prem, R., Kharbanda, P., & Mukherjee, N. (2023). A review paper: Will Artificial Intelligence (AI) replace the human recruiter?. *International Journal For Multidisciplinary Research*, 05(04). <https://doi.org/10.36948/ijfmr.2023.v05i04.4425>
- Pritchard, A. (1969). Statistical Bibliography or Bibliometrics?. *Journal of Documentation*, 25, 348–349.
- Resnawita, & Hendrik, B. (2023). Penggunaan metode systematic literatur review untuk menganalisis artikel sistem pakar metode forward chaining. *Journal of Information System and Education Development*, 1(2).
- Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human-Computer Interaction*, 35(5–6), 545–575. <https://doi.org/10.1080/07370024.2020.1735391>
- Salwei, M. E., & Carayon, P. (2022). A sociotechnical systems framework for the application of artificial intelligence in health care delivery. *Journal of Cognitive Engineering and Decision Making*, 16(4), 194-206. <https://doi.org/10.1177/15553434221097357>
- Scharowski, N., Perrig, S. A. C., Svab, M., Opwis, K., & Brühlmann, F. (2023). Exploring the effects of human-centered AI explanations on trust and reliance. *Frontiers in Computer Science*, 15. <https://doi.org/10.3389/fcomp.2023.1151150>
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98, 277–284. <https://doi.org/10.1016/j.chb.2019.04.019>

- Sides, T., Kbaier, D., Farrell, T., & Third, A. (2024). Exploring Trust In Artificial Intelligence Among Primary Care Stakeholders: A Mixed-Methods Study. *In Review*. <https://doi.org/10.21203/rs.3.rs-4945818/v1>
- Soulami, M., Benchekroun, S., & Galiulina, A. (2024). Exploring how AI adoption in the workplace affects employees: A bibliometric and systematic review. *Frontiers in Artificial Intelligence*, 7. <https://doi.org/10.3389/frai.2024.1473872>
- Su, H., Du, X., Chen, Y., Tian, S., Ma, Q., & Zhong, W. (2025). Value co-creation mechanism for AI-based enterprises with systems thinking: Case study of Baidu. *Systems Research and Behavioral Science*, 42(4), 1142-1162. <https://doi.org/10.1002/sres.3065>
- Tamò-Larriex, A., Guitton, C., Mayer, S., & Lutz, C. (2024). Regulating for trust: Can law establish trust in artificial intelligence?. *Regulation & Governance*, 18(3), 780–801. <https://doi.org/10.1111/rego.12568>
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8(1), 45. <https://doi.org/10.1186/1471-2288-8-45>
- Tomsett, R., Preece, A., Braines, D., Cerutti, F., Chakraborty, S., Srivastava, M., Pearson, G., & Kaplan, L. (2020). Rapid trust calibration through interpretable and uncertainty-aware AI. *Patterns*, 1(4), 100049. <https://doi.org/10.1016/j.patter.2020.100049>
- Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538. <https://doi.org/10.1007/s11192-009-0146-3>
- Vereschak, O., Bailly, G., & Caramiaux, B. (2021). How to evaluate trust in ai-assisted decision making? a survey of empirical methodologies. Paper presented at Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2).
- Von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, 34(4), 1607-1622. <https://doi.org/10.1007/s13347-021-00477-0>
- Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of Artificial Intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893-1924. <https://doi.org/10.1108/BPMJ-10-2019-0411>
- Wang, K., Zhao, Y., Gangadhari, R. K., & Li, Z. (2021). Analyzing the adoption challenges of the Internet of Things (IoT) and Artificial Intelligence (AI) for smart cities in China. *Sustainability*, 13(19), 10983. <https://doi.org/10.3390/su131910983>
- Weisz, J. D., Muller, M., & Varshney, K. R. (2025). Story arena: A multi-agent environment for envisioning the future of software engineering (version 1). *arXiv*. <https://doi.org/10.48550/ARXIV.2511.05410>
- Wölker, A., & Powell, T. E. (2021). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, 22(1), 86-103. <https://doi.org/10.1177/1464884918757072>
- Yablonsky, S. (2021). AI-driven platform enterprise maturity: From human led to machine governed. *Kybernetes*, 50(10), 2753-2789. <https://doi.org/10.1108/K-06-2020-0384>
- Yam, K. C., Tan, T., Jackson, J. C., Shariff, A., & Gray, K. (2023). Cultural differences in people's reactions and applications of robots, algorithms, and artificial intelligence. *Management and Organization Review*, 19(5), 859-875. <https://doi.org/10.1017/mor.2023.21>
- Yang, R., Li, S., Qi, Y., Liu, J., He, Q., & Zhao, H. (2025). Unveiling users' algorithm trust: The role of task objectivity, time pressure, and cognitive load. *Computers in Human Behavior Reports*, 18. <https://doi.org/10.1016/j.chbr.2025.100667>
- Yu, L., & Li, Y. (2022). Artificial intelligence decision-making transparency and employees' trust: The parallel multiple mediating effect of effectiveness and discomfort. *Behavioral Sciences*, 12(5), 127. <https://doi.org/10.3390/bs12050127>

- Yu, T., Tian, Y., Chen, Y., Huang, Y., Pan, Y., & Jang, W. (2025). How do ethical factors affect user trust and adoption intentions of AI-generated content tools? Evidence from a risk-trust perspective. *Systems*, 13(6), 461. <https://doi.org/10.3390/systems13060461>
- Zhang, G., Chong, L., Kotovsky, K., & Cagan, J. (2023). Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation. *Computers in Human Behavior*, 139. <https://doi.org/10.1016/j.chb.2022.107536>
- Zhou, C., & Ma, L. (2025). Bibliometric insights into the top 100 most-cited annual studies on digital health in nursing education (2020–2024). *Digital Health*. <https://doi.org/10.1177/20552076251342165>
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429-472. <https://doi.org/10.1177/1094428114562629>
- Zárate-Torres, R., Rey-Sarmiento, C. F., Acosta-Prado, J. C., Gómez-Cruz, N. A., Rodríguez Castro, D. Y., & Camargo, J. (2025). Influence of leadership on human–Artificial Intelligence collaboration. *Behavioral Sciences*, 15(7), 873. <https://doi.org/10.3390/bs15070873>