SIGNALING TRANSPARENCY IN THE ERA OF ARTIFICIAL INTELLIGENCE

Journal:	*Internet Research*
Manuscript ID	INTR-11-2023-1041.R5
Manuscript Type:	Research Paper
Keywords:	Artificial intelligence, Transparency, Co-citation analysis, Signaling theory, AI

SIGNALING TRANSPARENCY IN THE ERA OF ARTIFICIAL INTELLIGENCE

ABSTRACT

Purpose

This study provides researchers and business practitioners with a comprehensive understanding of artificial intelligence (AI) transparency in the business discipline, enabling them to navigate the evolving digital landscape, where AI transparency is an escalating concern, by identifying the conceptual foundations in the most influential studies.

Design

This study uses bibliometric analysis techniques, including performance and co-citation analyses. These analyses are grounded in data extracted from the Social Sciences Citation Index within the Web of Science, comprising 108 primary articles and 7,459 secondary cited documents.

Findings

AI transparency research is rising with a greater focus on end-users. Six clusters of cited publications serve as the bedrock of AI transparency in the business discipline: trust, AI explanation, bias and power, undesirable usage, user acceptance/aversion, and user heuristics. Analyzing these clusters revealed a framework for signaling AI transparency that can be extended to future research and business strategies.

Originality

This study addresses the following research gaps. First, the nature of AI transparency and its knowledge basis remain elusive. Second, AI transparency in the business discipline is under-explored compared to information sciences and law. Third, there is ambiguity surrounding the implementation strategies for AI transparency, with companies often resorting to simplistic methods such as updating terms and conditions. Fourth, there is a lack of clear future research directions specifically for AI transparency, as opposed to the broader context of AI ethics.

Keywords

Artificial Intelligence

AI

Co-Citation Analysis

Signaling Theory

Transparency

INTRODUCTION

Following multiple accidents, Tesla's AI-driven Autopilot system has faced scrutiny due to growing concerns about safety and transparency (Overberg *et al.*, 2024). People are demanding transparency from delivery apps about how couriers' pay and job access are decided (Booth, 2025). In a world where AI-based products and services are ubiquitous, both virtually (chatbots, search engines, generative AI) and physically (robots in hotels and restaurants, unmanned checkouts, biometric unlocking systems on phones), AI transparency emerges as a critical issue for business and its stakeholders.

AI is "a growing resource of interactive, autonomous, self-learning agency, which enables computational artifacts to perform tasks that otherwise would require human intelligence to be executed successfully" (Taddeo and Floridi, 2018, p.751). Notwithstanding alternative definitions of AI (Haenlein and Kaplan, 2019; Berente *et al.*, 2021), framing AI as a 'resource' emphasizes its business utility, while its 'autonomous, self-learning agency' highlights challenges to transparency.

AI transparency refers to disclosure, clarity, and openness (Ananny and Crawford, 2018; Christensen and Cornelissen, 2015) to enhance stakeholders' grasp of AI systems. Organizations are increasingly expected to embed transparency safeguards in AI systems—for example, to counter bot-generated ad traffic fraud that causes significant losses to advertisers (Fulgoni, 2016; Gordon *et al.*, 2021).

The passage of the first AI Act by the European Parliament on March 13, 2024, with public disclosure obligations for general-purpose AI systems to foster trust and accountability, reflects the growing institutional emphasis on AI transparency (European Parliament, 2024). Despite the recognized significance of AI transparency, understanding the meaning and foundations of the

ongoing AI transparency debate is challenging for business. The primary aim of this paper is to provide a comprehensive review delineating the conceptual underpinnings of AI transparency in the business discipline. The novelty of this study lies in filling the following research gaps:

(1) The elusive nature of AI transparency and its knowledge basis

(2) The underexplored AI transparency in business compared with other disciplines such as information sciences and law

(3) The ambiguity surrounding the implementation strategies for AI transparency, with companies often resorting to simplistic methods like updating terms and conditions (Volkmar *et al.*, 2022)

(4) A lack of clear future research directions specifically for AI transparency, as opposed to the broader context of AI ethics

This study undertakes bibliometric analyses to address the following research questions:

RQ1: What is the research performance regarding AI transparency in the business discipline?

RQ2: What are the intellectual foundations of this domain?

RQ3: What are the possible future research directions and business strategies for AI transparency?

In the following sections, we first provide the background on AI transparency in different fields, situating it within the broader literature on AI ethics. We then explain our bibliometric analysis method and present our findings. Finally, we offer insights and recommendations for researchers and businesses to navigate AI transparency.

THEORETICAL BACKGROUND

Transparency in Different Fields

Transparency has been examined through various lenses across different fields. In computing, it relates to the clarity with which users comprehend algorithms and computer applications, while in information systems management, it revolves around users' access to information about themselves or the extent of information disclosure to them (Turilli and Floridi, 2009). In the legal domain, transparency intertwines with discussions on safeguarding users' legal rights concerning personal data, as evidenced by the European General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) (Bukaty, 2019; European Union, 2016; Felzmann *et al.*, 2019). The recently enacted Artificial Intelligence Act defines transparency as ensuring the traceability and explainability of AI systems, and making humans aware of their interactions with AI systems (European Parliament, 2024, p. 28).

Transparency in business studies is inherent in efforts to monitor and control operational processes to increase organizational performance (Bernstein, 2017). It is contextualized within partners' mutual disclosure of information to foster relational capital and safeguard knowledge (Ho and Wang, 2015). When selectively applied in supply relationships, transparency enhances business value (Lamming *et al.*, 2004). Blockchain technology was created to improve supply chain transparency. Regarding customers' transparency expectations and experiences with AI, there is a call for AI and machine learning integration with marketing processes and customer interactions (Volkmar *et al.*, 2022).

The Theme of Transparency Within Ethics in AI

Transparency is a recurring theme in the discourse on AI ethics. Increasing reliance on data has raised ethical concerns about transparency, including bias, discrimination, and explainability. Bias from societal preconceptions can permeate systems through creators (Mittelstadt *et al.*, 2016). Technical constraints can lead to technical bias, unfairly disadvantaging certain groups. Emerging bias can further surface during usage, driven by differences in user values (Friedman and Nissenbaum, 1996). Moreover, explaining algorithmic decisions post facto is challenging, as machine learning algorithms continually evolve, complicating the distinction between isolated incidents and systemic failures (Mittelstadt *et al.*, 2016). Information on input data bias is rarely disclosed to users (Citron and Pasquale, 2014). Such a lack of transparency can incite user frustration (Eslami *et al.*, 2016). Table 1 summarizes selected review papers on AI ethics.[1] It indicates that AI transparency has garnered significant attention, including its conceptualization, explainability, privacy, security, fairness, and bias, among others, forming a crucial component of AI ethics discussions.

[Insert Table 1 here]

Existing literature reviews on AI transparency

As companies are increasingly scrutinized for lacking AI transparency, there is an urgent need for research that allows businesses and their stakeholders to better understand what AI transparency means, the basis of AI transparency, and what actions should be taken. In the studies listed in Table 1, with the exception of Ashok *et al.* (2022), AI transparency is mentioned as a secondary element, either as one of the AI ethics principles (Birkstedt *et al.*, 2023; Hunkenschroer

[1] Existing reviews on AI ethics can be technical (e.g., computer applications, engineering, ergonomics) or non-technical (e.g., law, business, education). Table 1 only includes non-technical papers in business.

and Luetge, 2022; Laine *et al.*, 2024) or in relation to specific ethical constructs, such as AI explainability (Brasse *et al.*, 2023; Haque *et al.*, 2023; Laato *et al.*, 2022). However, none of the studies provides a comprehensive picture of AI transparency.

Table 2 summarizes the main contributions of this study in comparison to the papers in Table 1. To the best of our knowledge, this is the first comprehensive review of the literature on AI transparency, specifically in the business discipline. Second, this study provides a more stakeholder-oriented definition of AI transparency, while past studies focused more on organizational information disclosure. Third, this study is the only review paper that provides the foundational intellectual structure of AI transparency by applying bibliometric analysis techniques. Finally, this study adopts signaling theory to strategically strengthen AI transparency at end-user and organizational levels while indicating directions for future research.

[Insert Table 2 here]

METHOD

Data

The process of identifying and selecting studies is depicted in Figure 1. We first identified keywords and concepts representing transparency in the AI systems field through brainstorming and from the literature to compile our bibliometric data.[2] We conducted an initial search on Google Scholar utilizing the keywords generated during the brainstorming session and scanned the first hundred results to identify additional, relevant terms commonly used in the research area (Abedin *et al.*, 2013; Khanra and Joseph, 2019). Our keywords included *transparency*, *artificial intelligence*, *AI*, *intelligent system*, *intelligent agent*, *intelligent assistant*, *autonomous system*,

[2] The description and rationale of the bibliometric approach are explained in Web Appendix 1.

autonomous agent, *virtual system*, *virtual agent*, *virtual assistant*, *voice assistant*, *robot*, *chatbot*, *social bot*, *bot, automated system*, and *automated agent*. These keywords were combined with Boolean operators and truncation symbols to retrieve journal articles containing them in titles, abstracts, keywords, and/or reference identifiers.

[Insert Figure 1 here]

We chose the 2023 Social Sciences Citation Index (SSCI) in the Web of Science (WoS) Core Collection (Clarivate, 2024a) as our database. [3] In addition, we employed the 2021 Academic Journal Guide (AJG) by the Chartered Association of Business Schools (2021) to filter business journals. We limited our SSCI search to academic journal articles published until 2023, excluding books, book chapters, editorials, proceeding papers, and letters. This search yielded 674 results. Then, after retaining only journals from AJG (2021), we obtained 206 articles.

The three authors independently screened the results and reviewed each article (Zupic and Čater, 2015). First, we evaluated the articles by reading the title, abstract, and keywords. This process led to the exclusion of non-business-centric articles (ergonomics, education, military science), articles unrelated to AI, or articles focused on modeling, measurement, or programming. Second, we assessed the full texts of the remaining articles, excluding those with marginal coverage of transparency. After this refinement, 108 articles (primary documents) containing 7,459 references (secondary documents) remained for analysis, as illustrated in Figure 2.

[Insert Figure 2 here]

[3] The criteria for database selection are explained in Web Appendix 1.

After cleaning the references for consistency, we ranked them by frequency of local citations (the number of citations received within our sample of primary documents) using BibExcel software to identify the most cited articles. Citation frequency indicates a publication's impact on the field (Ramos-Rodríguez and Ruíz-Navarro, 2004; Samiee and Chabowski, 2012; Stremersch *et al.*, 2007). Consistent with previous bibliometric studies (Wilden *et al.*, 2017; Zupic and Čater, 2015), we applied a citation threshold (the minimum number of local citations of a secondary document) to select the most influential secondary documents and reduce the number of references to a manageable size for analysis. Through an iterative approach involving several trials with different cut-off points (Zupic and Čater, 2015), we determined a threshold of four citations, reducing the sample of secondary documents to 115. Following Samiee and Chabowski (2021), we excluded review articles, meta-analyses, method-related publications, book reviews, and editorials from the dataset, resulting in 90 secondary documents remaining for co-citation analysis.

Analytical Approach

Our bibliometric analysis comprised performance and co-citation analyses. The first analysis used the *bibliometrix* R package (Aria and Cuccurullo, 2017). Performance analysis described the contributions of research constituents, including article productivity per year, the most influential articles, and the most productive authors, journals, institutions, and countries. Co-citation analysis aims to unveil the intellectual foundations of AI transparency systems. Co-citation indicates the frequency of two publications being cited together (Small, 1973). We used the highly cited references as the unit of analysis; the references of our primary articles represent the underlying intellectual base of a research domain (Culnan *et al.*, 1990; Samiee and Chabowski, 2012). The underlying principle of co-citation is that two secondary publications are considered thematically

similar if they are cited together in the same primary document. Thus, the more frequently two documents are cited together, the more similar they are. Co-citation determines interrelationships and proximity between publications (White and Griffith, 1981). We used BibExcel software to generate a co-citation matrix and Gephi software to create a network graph illustrating these relationships, with link strength based on the number of co-citations and node size weighted by the number of local citations received.

We employed the Louvain method for network community detection and Ward's method in hierarchical cluster analysis to identify thematic clusters within the network and determine subfields within the domain (Zupic and Čater, 2015).[4] Given the adjustability of the Louvain algorithm's resolution coefficient, we iteratively sought the optimal clustering solution (Wilden *et al.*, 2017). After experimenting with various cluster solutions with modularity above 0.4 (Blondel *et al.*, 2008) and assessing cluster quality and consistency, we selected a modularity resolution of 0.8, resulting in a six-group solution. We conducted hierarchical cluster analysis using SPSS with Ward's method to identify subgroups and to ensure the study's robustness and credibility of findings. This analysis generated a dendrogram aiding visual analysis and interpretation of clustering results. We obtained a six-cluster solution, broadly consistent with the network community detection method in Gephi, with only eleven differences in the classification. This study interprets and reports the cluster solution identified through the network community detection procedure. We followed a two-step process recommended by Zupic and Čater (2015) for cluster interpretation. The three authors independently examined cluster content by reviewing included publications. They then discussed their interpretations and agreed upon cluster labels.

[4] The justification of the Louvain method and Ward's method is explained in Web Appendix 1.

FINDINGS

Scientific Production

Figure 3 illustrates the chronological distribution of AI and transparency-related publications in our sample.[5] Article numbers remained low until 2019, then steadily increased, with a surge in 2021 and 2022. The surge reflected growing research interest and 2023 saw 33 articles published. This burgeoning field is anticipated to expand significantly, propelled by factors such as AI industry growth, escalating interest in AI systems and transparency research, and increasing funding opportunities.

[Insert Figure 3 here]

Foundational Intellectual Structure

Using the references in the primary articles, we identified six clusters, spatially represented in Figure 4, to reveal the literature's intellectual structure. Table 3 shows the representative papers in each cluster.

[Insert Figure 4 here]

[Insert Table 3 here]

Cluster 1: Trust. Trust is a pivotal aspect of transparency in determining behavior toward AI. It can determine willingness to relinquish some degree of control and monitoring to machines, embodying risk-taking (Mayer *et al.*, 1995). It can lead to irrational decisions, such as choosing

[5] The additional bibliometric analysis results appear in Web Appendix 2.

trusted systems that sometimes fail over superior but distrusted ones (Parasuraman and Riley, 1997). Transparent information regarding the design rationale (the reasoning behind automated errors) can bolster trust in automated decision aids (Dzindolet *et al.*, 2003; Mercado *et al.*, 2016). However, excessive trust can diminish users' situational awareness and lead to overcompensation following system errors (Endsley, 2017; Mercado *et al.*, 2016). Trust must be coupled with operators' ability to maintain situational awareness without information overload to effectively improve transparency (Mercado *et al.*, 2016). Robots should assist in reducing operators' mental workload by enhancing situational awareness and providing relevant information (Parasuraman *et al.*, 2000). This cluster underscores business practices that balance trust, situational awareness, and autonomy provided by AI to achieve organizational goals.

Cluster 2: AI Explanation. This cluster investigates methods to clarify algorithmic decision-making processes to the public to improve transparency. The effectiveness of explanations should not be evaluated only by domain experts but also by non-specialists (Miller, 2019). Transparency is essential for creating interpretable models (Guidotti *et al.*, 2019). The most cited paper, Ribeiro *et al.* (2016), showed that revealing the model's underlying text and image classifiers can assist non-specialists in determining whether to trust the model. Another approach is Lundberg's 'SHAP' analysis (Lundberg and Lee, 2017), which explains algorithmic predictions through feature importance measures. Models should ideally be designed with explainability in mind from the outset, not retroactively (Rudin, 2019). This involves improving reliability, causality, robustness, scalability, and generality (Guidotti *et al.*, 2019). Explanations may also have biases (Miller, 2019). Interactive explanations can reduce bias and enhance the inclusivity of discussions on algorithmic decision-making (Mittelstadt *et al.*, 2019). This cluster calls for continued efforts to improve user understanding of AI outcomes.

Cluster 3: Bias and power. Transparency has gained attention amid concerns that machines perpetuate bias by reusing generated data, obscuring sources behind decisions (Pasquale, 2015). A central theme in this cluster revolves around the undisclosed use of personal data, classification bias, and algorithmic discrimination (Burrell, 2016). Accumulating extensive consumer data raises privacy concerns, and personalization creates disparate information ecosystems (Pariser, 2011). Another key theme is power. The advent of big data was accompanied by a perceived aura of truth and accuracy, which conferred 'cultural power' (Boyd and Crawford, 2012). Unchecked, big data's influence perpetuates racial bias (O'Neil, 2016), criminalizes poverty, and impedes upward mobility (Eubanks, 2018). Proposed solutions include advocating for the adoption of similar testing and evaluation technologies by developers and courts to oversee and audit algorithmic decision outcomes (Kroll *et al.*, 2017). Multi-disciplinary efforts are required to develop tools that mitigate information asymmetries and enhance transparency and accountability (Lepri *et al.*, 2018).

Cluster 4: Undesirable usage. This cluster highlights inappropriate contexts for AI use and challenges conventional views of transparency. Ananny and Crawford (2018) cautioned against limiting transparency to mere openness or visibility, as it can inadvertently obscure critical aspects while emphasizing others. Complete disclosure of technical details can also facilitate malfeasance. Scholars called for greater transparency and accountability in algorithm design, emphasizing the ethical risks posed by gaps between design and operation (Martin, 2019; Mittelstadt *et al.*, 2016).

Using algorithms for employee direction, evaluation, and discipline is undesirable (Kellogg *et al.*, 2020). Employees tend to view certain decisions requiring intuition as more suited to human managers than algorithms (Lee, 2018). Employees' awareness of smart technology, AI, robotics, and algorithms can adversely affect commitment and career satisfaction (Brougham and Haar,

2018). Overall, AI systems should augment rather than replace human intelligence (Raisch and Krakowski, 2021; Wilson and Daugherty, 2018).

Cluster 5: User acceptance/aversion. This cluster explores the complex relationship between transparency and user acceptance. Davis' (1989) seminal work examined methods to enhance user acceptance of technology through perceived usefulness and ease of use. Dietvorst *et al.* (2015) highlighted that individuals can paradoxically be more accepting of human errors than machine errors despite machines being more transparent. Kizilcec (2016) showed that excessive or insufficient explanations can engender distrust in the system when users are initially disadvantaged. Users' responses to transparency are nuanced, with varied impacts of explanation types on beliefs about competence, benevolence, and integrity (Wang and Benbasat, 2007). Overall, the behavioral and psychological aspects of human interaction with AI provide deep explanations on user acceptance and aversion.

Cluster 6: User heuristics. This cluster centers on user heuristics, accountability, and explainability. Shin and Park (2019) argued that transparency improves when users are given simple heuristics—like explanations, rationales, and limitations—to help them quickly assess the system's trustworthiness. Rai (2020) advocated shifting from a 'black box' to a 'glass box' approach by focusing on user-oriented explanations. Diakopoulos (2016) emphasized transparency as pivotal in increasing accountability and linked accountability with user heuristics for investigating and disputing the integrity of the source data. Shin (2021) found that explaining recommendations increased user trust and emotional confidence when evaluating the causality of decisions.

DISCUSSION

Our study examined the research development in AI transparency with a business-centric focus. It showed the contributions of research constituents, identifying the annual scientific production, the most impactful publications, and the key authors, journals, institutions, and countries based on their productivity. Thus, we assessed research performance regarding AI transparency, addressing RQ1. Through co-citation analysis and clustering of publications for RQ2, we revealed the popular areas shaping the intellectual foundations of transparency. These include trust, AI explanation, bias and power, undesirable usage, user acceptance/aversion, and user heuristics. Uncovering the intellectual foundations and structure of AI transparency illuminated future research trajectories and business strategies, thereby addressing RQ3.

Theoretical implications

Our analysis of business-focused AI transparency shows rising concern about its unintended effects on users. Exploring AI in terms of underlying algorithms and architecture can be helpful in the field of information sciences, but less so when it comes to resolving business issues that are ethical and policy-related (Taddeo and Floridi, 2018). Transparency in business research was primarily about collecting data to control business processes for enhanced performance (Bernstein, 2017). However, as privacy concerns became more prominent, the focus shifted dramatically. This evolution highlights the importance of developing transparency practices that respect individual rights and foster trust.

Clusters 2, "AI explanation", and 6, "user heuristics", show an appetite for more nuanced transparency research around user engagement supported by stronger ethical standards. For example, the advertising literature explains how transparency enhances the credibility of

sponsored advertising (Krouwer *et al.*, 2020), but how it can backfire if consumers' privacy concerns are triggered (Kim *et al.*, 2018). Users often harbor skepticism toward companies' intentions (Foreh and Grier, 2003) due to the lack of technical expertise and companies' unclear disclosures about data usage (Christensen and Cornelissen, 2015). Although information sciences have pioneered explainable AI research, business researchers can play a pivotal role in uncovering attitudes on AI developments (e.g., agentic AI), analyzing consumer behavior, and developing targeted messaging to address transparency.

Clusters 1, "trust", and 5, "user acceptance/aversion", show the need to manage transparency to foster user engagement by improving situational awareness and clarifying AI decision-making. These clusters tend to present an information-sciences perspective as a starting point. From a business-centric perspective, authenticity, brand trust, virtual liberty, and users' opt-out behavior are areas where AI transparency can contribute. There are already studies on algorithmic literacy and trust (Shin *et al.*, 2022) as well as consumer trust in AI services, for instance, voice-based AI (Pitardi and Marriott, 2021), travel and tourism AI technologies (Wong *et al.*, 2024), and even romantic affection for AI (Song *et al.*, 2022). Further research on AI transparency can complement these studies.

Clusters 3, "bias and power", and 4, "undesirable usage", represent the need for more research on user perceptions of AI transparency. Users can lack awareness of issues relevant to social responsibility such as online price discrimination (Pandey and Caliskan, 2021) and facial recognition software inaccurately categorizing darker-skinned female faces (Buolamwini and Gebru, 2018). Moreover, business studies exploring AI ethics and data privacy are growing (Andrew and Baker, 2021; Fainmesser *et al.*, 2023; Grewal *et al.*, 2020; Martin and Murphy, 2017; Huang and Rust, 2021; Volkmar *et al.*, 2022). Scholars argued that enhanced transparency

diminishes users' perceptions of data vulnerability (Martin and Murphy, 2017; Waseem *et al.*, 2024).

Signaling theory is appropriate for situations of information asymmetry, such as users' limited expertise in machine learning and incomplete information on AI decisions (Spence, 2022). Developing tools to enhance transparency is vital in mitigating information asymmetries (Lepri *et al.*, 2018). In the context of AI, information asymmetry can disadvantage users during data collection and decision-making. Automated content moderation often occurs without transparency or public acknowledgment (Gorwa *et al.*, 2020). Even when users attempt to delete their accounts, they remain unaware of how algorithmic decisions influence account removal (Eslami *et al.*, 2016). Given the pervasive information asymmetry, users rely on unintended signals, like indications of AI profiling, to inform their perceptions. For example, a female BBC reporter questioned Netflix's AI algorithm's ability to infer her sexuality before she was aware of it (House, 2023). Although signaling has been extensively investigated across various contexts, including innovation, share pricing (Gomulya and Mishina, 2017), board membership (Certo, 2003), and strategic alliances (Ozmel *et al.*, 2013), its application in AI remains underexplored. A lack of understanding regarding which aspects of transparency should be signaled to stakeholders highlights a significant gap in business literature.

Practical implications

Should companies signal transparency? From an ethical perspective, signaling transparency is the right course of action. Companies wielding power should bear responsibility for enhancing user transparency. The AI Act (European Parliament, 2024) represents this ethos. Companies exacerbate information asymmetry by collecting excessive user data and transforming it into

behavioral data (Zuboff, 2019). This raises serious ethical concerns, exemplified by Meta Platforms' €405 million fine for Facebook (McCarthy, 2023).

Companies proactively considering transparency and committing resources to signaling transparency will be better equipped to navigate stricter user protection regulations. Like corporate social responsibility or ESG reporting, transparency is emerging as an area of compliance. Microsoft already publishes transparency reports online, covering topics from digital safety to content removal (Microsoft, 2024).

Transparency helps companies to cultivate trust among external stakeholders and enhance performance (Hyken, 2023). If companies opt to signal transparency, a starting point can involve choosing between low-cost and high-cost signaling strategies outlined in Table 4.

[Insert Table 4 here]

Organizations should explore diverse avenues for signaling transparency rather than assuming a one-size-fits-all approach is optimal (Laato *et al.*, 2022). Transparency signals can vary in clarity (Fox, 2007) and be provided in real-time or retrospectively (Heald, 2006). Building user confidence is another crucial aspect of transparency signaling. Based on Gomulya and Mishina (2017), stakeholders are attuned to both positive and negative signals, and companies can rehabilitate their reputation by taking and signaling remedial actions following breaches of stakeholder expectations. Market research often prioritizes product-focused inquiries, potentially overlooking consumer trust in AI. Low-cost signaling can involve addressing issues reactively through public relations. High-cost signaling can entail dedicating more resources to market research to understand consumer trust in AI and fostering algorithmic literacy among end-users to

enhance transparency (Ananny, 2016). Finally, a rapid diffusion of generative AI tools opens a new venue for transparency research. How to prevent misinformation and disinformation driven by harmful deepfakes has become an urgent agenda for scholars, practitioners, and policymakers. Such issues should be addressed in the context of transparency as a single overarching principle of disclosure, clarity, and openness in AI-generated content.

Organizations may signal transparency with specific risk levels to address their ethical obligations and maximize user value. According to the EU AI Act, if an AI system violates or impacts EU fundamental rights and values, it can be classified as having 'unacceptable risk' or 'high risk.' There can be a 'transparent risk' if an AI system leads to manipulated or deceptive outcomes (deepfakes and generative AI) (European Parliament, 2024). For example, the proliferation of toxic online content led platforms like YouTube to communicate the use of algorithmic moderation methods to establish digital content governance (Gorwa *et al.*, 2020).

Limitations

This study draws on the WoS SSCI database, which may not index all articles in the research area. Our data collection focused exclusively on academic journal articles, excluding books, book chapters, editorials, and letters. Hence, our analysis and findings only partially cover the literature on AI transparency. Our compilation and analysis of bibliometric data relied on selected keywords, and alternative search terms can influence the findings. The co-citation analysis, clustering of papers, and proposed research agenda were based on the most cited references and their interrelationships. Although highly cited references represent impactful and significant works, they constitute only a sample of the literature in this research area. Moreover, co-citation analysis primarily looks backward (Chabowski *et al.*, 2013), selecting articles that have accumulated many citations over time and excluding recent and emerging literature.

Future research

Figure 5 proposes research directions of AI transparency at end-user and organizational levels by incorporating signaling theory as an overarching framework. Signaling can be used in situations of information asymmetry to mitigate uncertainty (Spence, 2002). The information asymmetry between end-users and organizations utilizing AI systems can be examined regarding low-cost versus high-cost signaling strategies, with high-cost signaling often being the more effective approach (Kotha *et al.*, 2018, see Table 4). To the best of current knowledge, signaling theory has rarely been applied in the context of AI ethics in general and AI transparency in particular. Through this signaling transparency framework, future research can address diverse business challenges related to AI transparency at both the end-user and organizational levels.

[Insert Figure 5 here]

The suggested topics are not limited to those presented in Figure 5. AI transparency can be further investigated at the end-user level in contexts such as human-machine workplace configurations (Seeber *et al.*, 2020) and employee work displacement. Researchers can explore how AI transparency shapes employee perceptions of job roles, or how organizational guidelines can clarify AI's role in decision-making processes and facilitate employee adaptation to AI integration. Beyond workplace considerations, AI transparency can be analyzed within the customer journey. Researchers can analyze pre-purchase, purchase, and post-purchase transparency requirements and draw on information visualization (Wang *et al.*, 2023) and interactive tools to evaluate different strategies. Little is also known about how AI transparency can resolve the personalization-privacy paradox (Sutanto *et al.*, 2013) in online marketing or digital products (wearable health monitors) whereby consumers enjoy the benefits of using their personal data, but are similarly distrustful about privacy issues. A greater understanding of

transparency needs can start with determining privacy aspects consumers are willing to sacrifice, and sacrifice thresholds based on profiles and context. Spamming news bots that redirect users to advertising pages and interface interference in websites that incite or deter users from making certain decisions without the user's knowledge are also issues to resolve (Lokot and Diakopoulos, 2016;Mathur *et al.*, 2019). Future research can explore user detectability of transparency problems, and the impact on opt-out behavior and perceptions of authenticity. Research in employee and consumer contexts can help organizations develop ethical standards in AI transparency regarding user heuristics and AI explanation.

At an organizational level, how companies signal AI transparency and how these signals are interpreted are key questions to be explored. Researchers can examine how companies effectively communicate their ethical standards, authenticity, and social responsibility to stakeholders. This includes analyzing the types of signals that are most effective for different user groups and the situational factors influencing the interpretation of these signals. For example, researchers can investigate how AI transparency signals, such as customized explanations of AI decisions, compliance with ethical guidelines, or accountability through third-party audits, impact stakeholder perceptions. In addition, the role of visual and interactive elements in enhancing the effectiveness of transparency signals can be explored. Integrating signaling theory into AI transparency research also enables researchers to identify unintended consequences of transparency efforts. For instance, excessive explanations of AI transparency can be irrelevant or irritating to users while exposing the company to intellectual property infringement risks. Research on signaling the appropriate amount of information while maintaining user engagement will enhance our understanding of AI transparency management.

CONCLUSION

Given the increasing significance of AI transparency and the dearth of quantitative bibliographic studies in the domain, our bibliometric analysis highlights the performance of AI transparency research constituents and the research areas forming the intellectual bedrock of the domain. Our findings revealed a rising number of journal articles in the domain in recent years, potentially reflecting an uptick in standards on the ethical use of AI, such as the Ethics Guidelines for Trustworthy AI (European Commission, 2019), the Ten Principles for Ethical AI (UNI Global Union, 2017), and regulations related to AI transparency, such as the GDPR and the AI Act (European Parliament, 2024). Our suggested future research directions can be a springboard for further exploration of AI transparency, particularly in low-cost versus high-cost signaling. They also represent areas where organizations should consider fulfilling their ethical obligations in AI-based decisions. Organizations are strongly encouraged to adopt a multistakeholder approach to signaling AI transparency to have a broader societal impact. Finally, our bibliometric study and resulting lines of research can encourage multidisciplinary collaboration on AI transparency. For example, finance, accounting, and human resource researchers can shed light on their discipline-specific views that can complement other business areas, such as marketing and organizational behavior. Such multidisciplinary collaborations can advance and deepen our knowledge about the role of transparency in an AI-driven economy.

ACKNOWLEDGEMENTS

The first two authors contributed equally to this work.

REFERENCES

Abedin, B., Talaei-Khoei, T. and Ghapanchi, A. (2013), "A review of critical factors for communicating with customers on social networking sites", *International Technology Management Review,* Vol. 3, pp. 208-218.

Alcaide-Muñoz, L. and Rodríguez Bolívar, M.P. (2015), "Understanding e-government research", *Internet Research,* Vol. 25, No. 4, pp. 633-673.

Ananny, M. (2016), "Toward an ethics of algorithms: convening, observation, probability, and timeliness", *Science, Technology, & Human Values,* Vol. 41, No. 1, pp. 93-117.

Ananny, M. and Crawford, K. (2018), "Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability", *New Media & Society,* Vol. 20, No. 3, pp. 973-989.

Andrew, J. and Baker, M. (2021), "The General Data Protection Regulation in the age of surveillance capitalism", *Journal of Business Ethics,* Vol. 168, No. 3, pp. 565-578.

Aria, M. and Cuccurullo, C. (2017), "Bibliometrix: an R-tool for comprehensive science mapping analysis", *Journal of Informetrics,* Vol. 11, No. 4, pp. 959-975.

Ashok, M., Madan, R., Joha, A. and Sivarajah, U. (2022), "Ethical framework for artificial intelligence and digital technologies", *International Journal of Information Management,* Vol. 62, p. 102433.

Berente, N., Gu, B., Recker, J. and Santhanam, R. (2021), "Managing artificial intelligence", *MIS Quarterly,* Vol. 45, No. 3, pp.1433-1450.

Bernstein, E.S. (2017), "Making transparency transparent: the evolution of observation in management theory", *Academy of Management Annals,* Vol. 11, No. 1, pp. 217-266.

Birkstedt, T., Minkkinen, M., Tandon, A. and Mäntymäki, M. (2023), "AI governance: themes, knowledge gaps and future agendas", *Internet Research,* Vol. 33, No. 7, pp. 133-167.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. (2008), "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment,* Vol. 2008, No. 10, p. P10008.

Booth, R. (2025), "Delivery apps urged to lift lid on 'black-box algorithms' affecting UK couriers", available at: https://www.theguardian.com/business/2025/jan/20/food-delivery-apps-ubereats-deliveroo-justeat-urged-to-reveal-how-algorithms-affect-uk-courierss-work (accessed 3 February 2025)

Boyd, D. and Crawford, K. (2012), "Critical questions for big data", *Information, Communication & Society,* Vol. 15, No. 5, pp. 662-679.

Brasse, J., Broder, H.R., Förster, M., Klier, M. and Sigler, I. (2023), "Explainable artificial intelligence in information systems: a review of the status quo and future research directions", *Electronic Markets,* Vol. 33, No. 26, pp. 1-30.

Brougham, D. and Haar, J. (2018), "Smart technology, artificial intelligence, robotics, and algorithms (STARA): employees' perceptions of our future workplace", *Journal of Management & Organization,* Vol. 24, No. 2, pp. 239-257.

Bukaty, P. (2019), *The California Consumer Privacy Act (CCPA): An Implementation Guide,* IT Governance Publishing.

Buolamwini, J. and Gebru, T. (2018), "Gender shades: intersectional accuracy disparities in commercial gender classification", in Sorelle, A.F. & Christo, W. (Ed.s), *Conference on*

Fairness, Accountability and Transparency, Proceedings of Machine Learning Research, pp. 77-91.

Burrell, J. (2016), "How the machine 'thinks': understanding opacity in machine learning algorithms", *Big Data & Society,* Vol. 3, No. 1, p. 2053951715622512.

Certo, S.T. (2003), "Influencing initial public offering investors with prestige: signaling with board structures", *Academy of Management Review,* Vol. 28, No. 3, pp. 432-446.

Chabowski, B.R., Hult, G.T.M. and Mena, J.A. (2011), "The retailing literature as a basis for franchising research: using intellectual structure to advance theory", *Journal of Retailing,* Vol. 87, No. 3, pp. 269-284.

Chabowski, B.R., Samiee, S. and Hult, G.T.M. (2013), "A bibliometric analysis of the global branding literature and a research agenda", *Journal of International Business Studies,* Vol. 44, No. 6, pp. 622-634.

Chartered Association of Business Schools (2021), "Academic journal guide", available at: https://charterdabs.org/academic-journal-guide/academic-journal-guide-2021 (accessed 15 January 2024)

Christensen, L.T. and Cornelissen, J. (2015), "Organizational transparency as myth and metaphor", *European Journal of Social Theory,* Vol. 18, No. 2, pp. 132-149.

Citron, D.K. and Pasquale, F. (2014), "The scored society: due process for automated predictions", *Washington Law Review,* Vol. 89, pp. 1-34.

Clarivate (2024a), "Social Sciences Citation Index", available at: https://mjl.clarivate.com/search-results (accessed 15 January 2024)

Clarivate (2024b), "Web of Science platform: summary of coverage", available at: https://clarivate.libguides.com/librarianresources/coverage (accessed 26 February 2024)

Culnan, M.J., O'Reilly III, C.A. and Chatman, J.A. (1990), "Intellectual structure of research in organizational behavior, 1972–1984: a cocitation analysis", *Journal of the American Society for Information Science,* Vol. 41, No. 6, pp. 453-458.

Davis, F.D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly,* Vol. 13, No. 3, pp. 319-340.

Diakopoulos, N. (2016), "Accountability in algorithmic decision making", *Communications of the ACM,* Vol. 59, No. 2, pp. 56–62.

Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015), "Algorithm aversion: people erroneously avoid algorithms after seeing them err", *The Journal of Experimental Psychology: General,* Vol. 144, No. 1, pp. 114-26.

Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G. and Beck, H.P. (2003), "The role of trust in automation reliance", *International Journal of Human-Computer Studies,* Vol. 58, No. 6, pp. 697-718.

Endsley, M.R. (2017), "From here to autonomy: Lessons learned from human–automation research", *Human Factors,* Vol. 59, No. 1, pp. 5-27.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K. and Sandvig, C. (2016), "'I always assumed that I wasn't really that close to [her]': reasoning about invisible algorithms in news feeds", *Proceedings of the 33rd Annual Association for Computing Machinery Conference on Human Factors in Computing Systems*, Seoul.

Eubanks, V. (2018), *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor,* St. Martin's Press, New York.

European Commission (2019), "Ethics guidelines for trustworthy AI", available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 7 April 2025)

European Parliament (2024), "EU artificial intelligence act", available at: https://artificialintelligenceact.eu/ai-act-explorer/ (accessed 7 April 2025)

European Union (2016), "General data protection regulation", available at: https://gdpr-info.eu (accessed 7 April 2025)

Fainmesser, I.P., Galeotti, A. and Momot, R. (2023), "Digital privacy", *Management Science,* Vol. 69, No. 6, pp. 3157-3173.

Felzmann, H., Villaronga, E.F., Lutz, C. and Tamò-Larrieux, A. (2019), "Transparency you can trust: transparency requirements for artificial intelligence between legal norms and contextual concerns", *Big Data & Society,* Vol. 6, No. 1, p. 2053951719860542.

Ferreira, F.A.F. (2018), "Mapping the field of arts-based management: bibliographic coupling and co-citation analyses", *Journal of Business Research,* Vol. 85, pp. 348-357.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018), "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations", *Minds & Machines*, Vol. 28, pp. 689–707.

Foreh, M.R. and Grier, S. (2003), "When is honesty the best policy? The effect of stated company intent on consumer skepticism", *Journal of Consumer Psychology,* Vol. 13, No. 3, pp. 349-356.

Fortunato, S. (2010), "Community detection in graphs", *Physics Reports,* Vol. 486, No. 3, pp. 75-174.

Fox, J. (2007), "The uncertain relationship between transparency and accountability", *Development in Practice,* Vol. 17, No. 4/5, pp. 663-671.

Friedman, B. and Nissenbaum, H. (1996), "Bias in computer systems", *ACM Transactions on Information Systems,* Vol. 14, No. 3, pp. 330-347.

Fulgoni, G.M. (2016), "Fraud in digital advertising: a multibillion-dollar black hole. How marketers can minimize losses caused by bogus web traffic", *Journal of Advertising Research,* Vol. 56, No. 2, pp. 122-125.

Glikson, E. and Woolley, A.W. (2020), "Human trust in artificial intelligence: review of empirical research", *Academy of Management Annals,* Vol. 14, No. 2, pp. 627-660.

Gomulya, D. and Mishina, Y. (2017), "Signaler credibility, signal susceptibility, and relative reliance on signals: how stakeholders change their evaluative processes after violation of expectations and rehabilitative efforts", *Academy of Management Journal,* Vol. 60, No. 2, pp. 554-583.

Gordon, B.R., Jerath, K., Katona, Z., Narayanan, S., Shin, J. and Wilbur, K.C. (2021), "Inefficiencies in digital advertising markets", *Journal of Marketing,* Vol. 85, No. 1, pp. 7-25.

Gorwa, R., Binns, R. and Katzenbach, C. (2020), "Algorithmic content moderation: technical and political challenges in the automation of platform governance", *Big Data & Society,* Vol. 7, No. 1, p. 2053951719897945.

Grewal, D., Hulland, J., Kopalle, P.K. and Karahanna, E. (2020), "The future of technology and marketing: a multidisciplinary perspective", *Journal of the Academy of Marketing Science,* Vol. 48, No. 1, pp. 1-8.

Guidotti, R., Monreale, A., Turini, F., Pedreschi, D. and Giannotti, F. (2019), "A survey of methods for explaining black box models", *ACM computing surveys (CSUR),* Vol. 51, pp. 1 - 42.

Haenlein, M. and Kaplan, A. (2019), "A brief history of artificial intelligence: on the past, present, and future of artificial intelligence", *California Management Review,* Vol. 61, No. 4, pp. 5-14.

Haque, A.K.M.B., Islam, A.K.M.N. and Mikalef, P. (2023), "Explainable Artificial Intelligence (XAI) from a user perspective: a synthesis of prior literature and problematizing avenues for future research", *Technological Forecasting and Social Change,* Vol. 186, p. 122120.

Heald, D.A. (2006), "Varieties of transparency", *Transparency: the key to better governance? Proceedings of the British Academy,* Vol. 135, pp. 25-43.

Ho, M.H.-W. and Wang, F. (2015), "Unpacking knowledge transfer and learning paradoxes in international strategic alliances: contextual differences matter", *International Business Review,* Vol. 24, No. 2, pp. 287-297.

House, E. (2023), "Netflix: How did it know I was bi before I did?", available at: https://www.bbc.co.uk/news/technology-66472938 (accessed 29 January 2023).

Huang, M.-H. and Rust, R.T. (2021), "A strategic framework for artificial intelligence in marketing", *Journal of the Academy of Marketing Science,* Vol. 49, No. 1, pp. 30-50.

Hunkenschroer, A.L. and Luetge, C. (2022), "Ethics of AI-Enabled recruiting and selection: a review and research agenda", *Journal of Business Ethics,* Vol. 178, No. 4, pp. 977-1007.

Hyken, S. (2023), "Radical transparency: a key to a better customer experience", available at: https://www.forbes.com/sites/shephyken/2023/01/29/radical-transparency-a-key-to-a-better-customer-experience/?sh=272d5a4e16ba (accessed 29 January 2023)

Jarrahi, M.H. (2018), "Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making", *Business Horizons*, Vol. 61, No. 4, pp.577-586. https://doi.org/10.1016/j.bushor.2018.03.007.

Kellogg, K.C., Valentine, M.A. and Christin, A. (2020), "Algorithms at work: the new contested terrain of control", *Academy of Management Annals,* Vol. 14, No. 1, pp. 366-410.

Khanra, S. and Joseph, R.P. (2019), "E-governance maturity models: a meta-ethnographic study", *The International Technology Management Review,* Vol. 8, No. 1, pp. 1-9.

Khare, A. and Jain, R. (2022), "Mapping the conceptual and intellectual structure of the consumer vulnerability field: a bibliometric analysis", *Journal of Business Research,* Vol. 150, pp. 567-584.

Kim, T., Barasz, K. and John, L.K. (2018), "Why am I seeing this ad? The effect of ad transparency on ad effectiveness", *Journal of Consumer Research,* Vol. 45, No. 5, pp. 906-932.

Kizilcec, R.F. (2016), "How much information? Effects of transparency on trust in an algorithmic interface", *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2390-2395.

Kotha, R., Crama, P. and Kim, P.H. (2018), "Experience and signaling value in technology licensing contract payment structures", *Academy of Management Journal,* Vol. 61, No. 4, pp. 1307-1342.

Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G. and Yu, H. (2017), "Accountable algorithms", *University of Pennsylvania Law Review*, Vol. 165, No. 3, pp. 633-705.

Krouwer, S., Poels, K. and Paulussen, S. (2020), "Moving towards transparency for native advertisements on news websites: a test of more detailed disclosures", *International Journal of Advertising,* Vol. 39, No. 1, pp. 51-73.

Laato, S., Tiainen, M., Najmul Islam, A.K.M. and Mäntymäki, M. (2022), "How to explain AI systems to end users: a systematic literature review and research agenda", *Internet Research,* Vol. 32, No. 7, pp. 1-31.

Laine, J., Minkkinen, M. and Mäntymäki, M. (2024), "Ethics-based AI auditing: a systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders", *Information & Management,* Vol. 61, No. 5, p. 103969.

Lamming, R., Caldwell, N. and Harrison, D. (2004), "Developing the concept of transparency for use in supply relationships", *British Journal of Management,* Vol. 15, No. 4, pp. 291-302.

Lee, M.K. (2018), "Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management", *Big Data & Society,* Vol. 5, No. 1, p. 2053951718756684.

Lepri, B., Oliver, N., Letouzé, E., Pentland, A. and Vinck, P. (2018), "Fair, transparent, and accountable algorithmic decision-making processes", *Philosophy & Technology,* Vol. 31, No. 4, pp. 611-627.

Liu, X., Glänzel, W. and De Moor, B. (2012), "Optimal and hierarchical clustering of large-scale hybrid networks for scientific mapping", *Scientometrics,* Vol. 91, No. 2, pp. 473-493.

Lokot, T. and Diakopoulos, N. (2016), "News bots", *Digital Journalism,* Vol. 4, No. 6, pp. 682-699.

Lundberg, S.M. and Lee, S.-I. (2017), "A unified approach to interpreting model predictions", *Advances in Neural Information Processing Systems,* Vol. 30.

Martin, K. (2019), "Ethical implications and accountability of algorithms", *Journal of Business Ethics,* Vol. 160, No. 4, pp. 835-850.

Martin, K.D. and Murphy, P.E. (2017), "The role of data privacy in marketing", *Journal of the Academy of Marketing Science,* Vol. 45, No. 2, pp. 135-155.

Mathur, A., Acar, G., Friedman, M.J., Lucherini, E., Mayer, J., Chetty, M. and Narayanan, A. (2019), "Dark patterns at scale: findings from a crawl of 11K shopping websites", *Proceedings of the ACM on Human-Computer Interaction,* Vol. 3, No. CSCW, pp. 1-32.

Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995), "An integrative model of organizational trust", *Academy of Management Review,* Vol. 20, No. 3, pp. 709-734.

McCarthy, N. (2023), "The biggest GDPR fines of 2022", available at: https://www.eqs.com/compliance-blog/biggest-gdpr-fines/ (accessed 31 January 2023)

Mercado, J.E., Rupp, M.A., Chen, J.Y.C., Barnes, M.J., Barber, D. and Procci, K. (2016), "Intelligent agent transparency in human–agent teaming for multi-UxV management", *Human Factors,* Vol. 58, No. 3, pp. 401-415.

Microsoft (2024), "Transparency reports", available at: https://www.microsoft.com/en/digitalsafety/transparency-reports (accessed 19 April 2024)

Miller, T. (2019), "Explanation in artificial intelligence: insights from the social sciences", *Artificial Intelligence,* Vol. 267, pp. 1-38.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016), "The ethics of algorithms: mapping the debate", *Big Data & Society,* Vol. 3, No. 2, p. 2053951716679679.

Mittelstadt, B.D., Russell, C. and Wachter, S. (2019), "Explaining explanations in AI", *Proceedings of the Conference on Fairness, Accountability, and Transparency,* Association for Computing Machinery, Atlanta, pp. 279-288.

O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* Crown Publishing Group, New York.

Overberg, O., Scott, E. and Matt, F. (2024), "Inside the WSJ's investigation of Tesla's autopilot crash risks", available at: https://www.wsj.com/business/autos/tesla-autopilot-crash-investigation-997b0129 (accessed 25 January 2025)

Ozmel, U., Reuer, J.J. and Gulati, R. (2013), "Signals across multiple networks: How venture capital and alliance networks affect interorganizational collaboration", *Academy of Management Journal,* Vol. 56, No. 3, pp. 852-866.

Pandey, A. and Caliskan, A. (2021), "Disparate impact of artificial intelligence bias in ridehailing economy's price discrimination algorithms", *Proceedings of Conference on AI, Ethics, and Society,* available at: https://dl.acm.org/doi/pdf/10.1145/3461702.3462561 (accessed 7 April 2025)

Parasuraman, R. and Riley, V. (1997), "Humans and automation: use, misuse, disuse, abuse", *Human Factors,* Vol. 39, No. 2, pp. 230-253.

Parasuraman, R., Sheridan, T.B. and Wickens, C.D. (2000), "A model for types and levels of human interaction with automation", *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* Vol. 30, No. 3, pp. 286-297.

Pariser, E. (2011), *The Filter Bubble: What the Internet Is Hiding from You,* Penguin Press, New York.

Pasquale, F. (2015), *The Black Box Society: The Secret Algorithms That Control Money and Information,* Harvard University Press, Cambridge, MA.

Pitardi, V. and Marriott, H.R. (2021), "Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice-based artificial intelligence", *Psychology & Marketing,* Vol. 38, No. 4, pp. 626-642.

Podsakoff, P.M., MacKenzie, S.B., Bachrach, D.G. and Podsakoff, N.P. (2005), "The influence of management journals in the 1980s and 1990s", *Strategic Management Journal,* Vol. 26, No. 5, pp. 473-488.

Rai, A. (2020), "Explainable AI: from black box to glass box", *Journal of the Academy of Marketing Science,* Vol. 48, No. 1, pp. 137-141.

Raisch, S. and Krakowski, S. (2021), "Artificial intelligence and management: the automation–augmentation paradox", *Academy of Management Review,* Vol. 46, No. 1, pp. 192-210.

Ramos-Rodríguez, A.-R. and Ruíz-Navarro, J. (2004), "Changes in the intellectual structure of strategic management research: a bibliometric study of the Strategic Management Journal, 1980–2000", *Strategic Management Journal,* Vol. 25, No. 10, pp. 981-1004.

Ribeiro, M.T., Singh, S. and Guestrin, C. (2016), "'Why should I trust you?' Explaining the predictions of any classifier", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135-1144.

Rudin, C. (2019), "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", *Nature Machine Intelligence,* Vol. 1, No. 5, pp. 206-215.

Samiee, S. and Chabowski, B.R. (2012), "Knowledge structure in international marketing: a multi-method bibliometric analysis", *Journal of the Academy of Marketing Science,* Vol. 40, No. 2, pp. 364-386.

Samiee, S. and Chabowski, B.R. (2021), "Knowledge structure in product-and brand origin–related research", *Journal of the Academy of Marketing Science,* Vol. 49, No. 5, pp. 947-968.

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I. and Lowry, P.B. (2020), "Collaborating with technology-based autonomous agents", *Internet Research,* Vol. 30, No. 1, pp. 1-18.

Serenko, A. and Bontis, N. (2024), "Dancing with the devil: the use and perceptions of academic journal ranking lists in the management field", *Journal of Documentation,* Vol. 80, No. 4, pp. 773-792.

Shin, D. (2021), "The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI", *International Journal of Human-Computer Studies,* Vol. 146, No. 102551, pp.1-10.

Shin, D. and Park, Y.J. (2019), "Role of fairness, accountability, and transparency in algorithmic affordance", *Computers in Human Behavior,* Vol. 98, pp. 277-284.

Shin, D., Rasul, A. and Fotiadis, A. (2022), "Why am I seeing this? Deconstructing algorithm literacy through the lens of users", *Internet Research,* Vol. 32, No. 4, pp. 1214-1234.

Small, H. (1973), "Co-citation in the scientific literature: a new measure of the relationship between two documents", *Journal of the American Society for information Science,* Vol. 24, No. 4, pp. 265-269.

Song, X., Xu, B. and Zhao, Z. (2022), "Can people experience romantic love for artificial intelligence? An empirical study of intelligent assistants", *Information & Management,* Vol. 59, No. 2, pp. 103595.

Spence, M. (2002), "Signaling in retrospect and the informational structure of markets", *The American Economic Review,* Vol. 92, No. 3, pp. 434-459.

Stremersch, S., Verniers, I. and Verhoef, P.C. (2007), "The quest for citations: drivers of article impact", *Journal of Marketing,* Vol. 71, No. 3, pp. 171-193.

Sutanto, J., Palme, E., Tan, C.-H. and Phang, C.W. (2013), "Addressing the personalization-privacy paradox: an empirical assessment from a field experiment on smartphone users", *MIS Quarterly,* Vol. 37, No. 4, pp. 1141–1164.

Taddeo, M. and Floridi, L. (2018), "How AI can be a force for good", *Science,* Vol. 361, No. 6404, pp. 751-752.

Tambe, P., Cappelli, P. and Yakubovich, V. (2019), "Artificial intelligence in human resources management: challenges and a path forward", *California Management Review*, Vol. 61, No. 4, pp. 15-42.

Tranfield, D., Denyer, D. and Smart, P. (2003), "Towards a methodology for developing evidence-informed management knowledge by means of systematic review", *British Journal of Management,* Vol. 14, No. 3, pp. 207-222.

Turilli, M. and Floridi, L. (2009), "The ethics of information transparency", *Ethics and Information Technology,* Vol. 11, No. 2, pp. 105-112.

UNI Global Union (2017), "10 principles for ethical artificial intelligence", available at: https://uniglobalunion.org/report/10-principles-for-ethical-artificial-intelligence/ (accessed 20 April 2024)

Volkmar, G., Fischer, P.M. and Reinecke, S. (2022), "Artificial intelligence and machine learning: exploring drivers, barriers, and future developments in marketing management", *Journal of Business Research,* Vol. 149, pp. 599-614.

Wang, R., Bush-Evans, R., Arden-Close, E., Bolat, E., McAlaney, J., Hodge, S., Thomas, S. and Phalp, K. (2023), "Transparency in persuasive technology, immersive technology, and online marketing: facilitating users' informed decision making and practical implications", *Computers in Human Behavior,* Vol. 139, pp. 107545.

Wang, W. and Benbasat, I. (2007), "Recommendation agents for electronic commerce: effects of explanation facilities on trusting beliefs", *Journal of Management Information Systems,* Vol. 23, No. 4, pp. 217-246.

Waseem, D., Chen, S., Xia, Z., Rana, N.P., Potdar, B. and Tran, K.T. (2024), "Consumer vulnerability: understanding transparency and control in the online environment", *Internet Research,* Vol. 34 No. 6, pp. 1992-2030.

White, H.D. and Griffith, B.C. (1981), "Author cocitation: a literature measure of intellectual structure", *Journal of the American Society for information Science,* Vol. 32, No. 3, pp. 163-171.

Wilden, R., Akaka, M.A., Karpen, I.O. and Hohberger, J. (2017), "The evolution and prospects of service-dominant logic: an investigation of past, present, and future research", *Journal of Service Research,* Vol. 20, No. 4, pp. 345-361.

Wilson, H.J. and Daugherty, P.R. (2018), "Collaborative intelligence: humans and AI are joining forces", *Harvard Business Review,* Vol. 96, No. 4, pp. 114-123.

Wong, L.-W., Tan, G.W.-H., Ooi, K.-B. and Dwivedi, Y. (2024), "The role of institutional and self in the formation of trust in artificial intelligence technologies", *Internet Research,* Vol. 34, No. 2, pp. 343-370.

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power,* Profile books.

Zupic, I. and Čater, T. (2015), "Bibliometric methods in management and organization", *Organizational Research Methods,* Vol. 18, No. 3, pp. 429-472.

Table 1: Selected review papers on ethics in AI

Authors (year)	Review Focus	Review Type	Topics Covered				No. of Articles	Review Period	Themes Identified
			T	*F&B*	*P&S*	*X*			
Ashok *et al.* (2022)	Ethical use of AI in digital technologies	Systematic		✓	✓		59	2000–2020	Explicability, justice, beneficence, non-maleficence, and governance
Hunkenschroer and Luetge (2022)	Ethicality in AI-enabled recruiting	Systematic	✓	✓	✓	✓	51	2016–2020	Ethics in AI-enabled recruiting, differentiating between ethical opportunities, ethical risks (including lack of transparency and explainability), and ethical ambiguities
Laato *et al.* (2022)	AI systems and explainability for end users	Systematic	✓	✓	✓	✓	25	2008–2020	Understandability, trustworthiness, transparency, controllability, and fairness, along with recommendations for designing AI system explanations for end users
Birkstedt *et al.* (2023)	Organizational-level AI governance	Systematic	✓	✓	✓	✓	68	2010–2021	Technology, stakeholders and context, regulation, and processes
Brasse *et al.* (2023)	Explainable AI in information systems	Structured	✓			✓	180	2010–2021	Methods to reveal the functioning of specific black box applications for experts, developers, domain experts, and lay users; explaining decisions and functioning of arbitrary black boxes; investigating the impact of explanations on lay users; investigating the effect of explanations on domain experts; investigating employment of AI explainability in practice

Haque *et al.* (2023)	AI users' explanation needs	Systematic	✓	✓	✓	✓	58	Up to 2021	Dimensions shaping AI explanation: format, completeness, accuracy, and currency; outcomes of AI explanation: trust, transparency, understandability, usability, and fairness
Laine *et al.* (2024)	Ethical principles in AI auditing	Systematic	✓	✓	✓	✓	93	Up to March 2022	Fairness, transparency, non-maleficence, responsibility, privacy, trust, beneficence, and freedom/autonomy

Note: This study only considered the literature reviews on ethics in AI in general, published in the social sciences domain.
Keys: T = transparency; F&B = fairness and bias; P&S = privacy and security; X = explainability.

Source: Authors' own work

Table 2: Comparison of previous related review papers

Key Studies	Current Study	Hunkenschroer and Luetge (2022)	Laato *et al.* (2022)	Birkstedt *et al.* (2023)	Haque *et al.* (2023)	Laine *et al.* (2024)	Brasse *et al.* (2023)
Disciplines Covered	Business	Law, management, organizational psychology, robotics, computer science	Computer science	Computer science, social sciences, humanities, management	Healthcare, media and entertainment, education, transportation, finance, e-commerce, human resource management, digital assistant, e-governance, social networking	Computer science, information systems, electrical and electronics engineering	Information systems
Knowledge Synthesis Method	Bibliometric study	Systematic review					Structured review

Coverage of AI Transparency	As the main and exclusive topic	As one of the five key ethical principles	As a component of AI explainability	As algorithmic transparency, along with explainability and inscrutability of algorithms	In the context of AI explainability	As a related concept of AI accountability	In the context of AI explainability
Definition of AI Transparency	"the clarity and openness to increase stakeholders' ability to understand how AI systems work"	n.a.	"the degree of information that is disclosed about the AI system"	n.a.	"the concept of revealing the opaque procedure of decision making, allowing the whole procedure to be scrutinized by non-technical/average users if needed"	"a way to make information accessible and an entitlement of a counterpart outside the accountable organization to obtain that information"	"the willingness to disclose (parts of) the AI system by the owners and is thus considered a strategic management issue"
Intellectual Foundations of AI Transparency	Trust, explaining AI, bias and power, undesirable usage, user acceptance/aversion, user heuristics	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Theoretical propositions	Signaling theory to strategically improve AI transparency	n.a. (social contract theory for AI recruiting)	Link between transparency and trust	n.a. (social contract theory for AI governance)	n.a. (IS models for AI explainability)	n.a.	n.a.

Note: n.a. = not applicable

Source: Authors' own work

Table 3: Clusters summary

Cluster (Number of Articles)	**Cluster Focus and Link to Transparency**	**Representative Articles**	**Relevance to the Cluster Theme**	**Research Methods**	**Main Findings**
1. Trust (14)	A lack of transparency can impact trust in the system.	Dzindolet *et al.* (2003)	Users distrust automated aid following errors, but will trust it when understanding why errors occurred.	The authors carried out three experiments on the relationships between trust, automation reliability, and reliance.	Knowing the reasons for errors in automated aid increases trust in decisions and reliance, even when trust is unwarranted.
		Mayer *et al.* (1995)	Trust can be conceptualized as a willingness to be vulnerable to the actions of another party and to take risks.	The authors developed a conceptual model of dyadic trust, focusing on the critical role of risk in an organizational relationship.	It is possible to identify the determinants of trustworthiness and differentiate between causes and outcomes of trust.
		Mercado *et al.* (2016)	Agent transparency level affects trust in the context of human-agent teaming for multi-robot management.	The authors developed a within-subjects design on operator performance, trust, workload, and usability with three levels of agent transparency.	Transparency is increased by operator performance, trust, and usability, resulting in performance effectiveness without further costs.
2. AI explanation (20)	Explaining the way AI works is essential for transparency.	Guidotti *et al.* (2019)	There are different methods for explaining black box models.	The authors provided a review of methods for explaining decision systems based on machine-learning models.	There are four black box problems. Methods for explaining them can be classified across the four dimensions.
		Miller (2019)	Explanation in AI can be explored from philosophy, social psychology, and cognitive psychology angles.	The author provided a review of explanations in philosophy, social psychology, and cognitive psychology literatures.	Social sciences reveal how explanation is formed. Explanations are often contrastive, selective, social and targeting causal links rather than probabilities.
		Ribeiro *et al.* (2016)	There are different ways to explain the prediction of	The authors ran a simulated user experiment showing the	Explanations are useful for models in trust-related tasks across text and

			classifiers in AI models to users. Users can decide whether to trust the prediction based on the explanations.	utility of explanations in trust-related tasks. They introduced LIME and SP-LIME methods.	image, including deciding between models, evaluating trust, and improving untrustworthy models.
3. Bias and power (15)	Ethical issues relating to transparency emanate from biased decision-making and the power of corporations.	Burrell (2016)	Machine learning models have three levels of opacity, including corporate or state secrecy, technical illiteracy, and unavoidable complexity.	The author presented a review of existing computer science literature and industry practices, including code testing and manipulation for a lightweight audit.	There are three forms of opacity that are key to determining technical and non-technical solutions to prevent harm.
		O'Neil (2016)	Flawed mathematical models use big data to generate decisions that affect people's lives and increase inequality.	The author presented a critical review of data usage in AI models using case studies.	Misuse of modeling techniques and process opacity can create negative outcomes. Solutions include positive feedback loops, algorithmic audits, and data-sharing.
		Pasquale (2015)	Regarding transparency, the possibility for firms to scrutinize others without being scrutinized is a form of power.	The author analyzed power imbalance issues caused by information asymmetry in various black-boxing practices.	Regulators should address algorithmic complexities and uncontrolled data collection, holding corporations accountable and encouraging more privacy protections.
4. Undesirable usage (18)	AI should not be used in inappropriate contexts. There are limits to conventional views of transparency.	Ananny and Crawford (2018)	Transparency has boundaries. It can work as an ideal but there are situations where it can fail.	The authors provided a theoretical analysis and case studies showing the ideal of transparency through different political theories, regulatory regimes, material systems, and epistemological models.	An alternative typology of algorithmic governance that identifies limitations of transparency can help develop an ethics of algorithmic accountability.
		Lee (2018)	Algorithmic decision-making is sometimes viewed less favorably than human decision-making.	The author used a scenario-based experiment to explore views on algorithmic	Task characteristics requiring more human or mechanical skills significantly influence perceptions

				management (trust, fairness, and emotional response).	of algorithmic decisions compared to human-made ones.
		Martin (2019)	The paper focuses on ethical issues of transparency and accountability for algorithmic decisions.	The author presented theoretical arguments on algorithmic design and justifications for firms' responsibility in the ethical use of algorithms.	Algorithms are value-laden, with moral consequences. Companies that design them must take responsibility for the ethical issues their use creates.
5. User acceptance/ aversion (13)	Transparency influences user reliance on AI systems.	Davis (1989)	Perceived usefulness and perceived ease of use are important factors for user acceptance of information technology.	The author carried out a field study to assess the reliability and construct validity of scales, then a lab study to show the relationship between usefulness, ease of use, and self-reported usage.	Usefulness and ease of use are identified as important determinants of user acceptance of information technology. Usefulness is more significantly linked to usage than ease of use.
		Dietvorst *et al.* (2015)	Algorithmic aversion can occur despite algorithms outperforming human forecasters.	The authors used five experiments to show people's likelihood of using an algorithm over a human forecaster when observing the algorithm's performance and errors.	Seeing errors in an algorithm makes people less confident and less likely to choose it over an inferior human forecaster.
		Kizilcec (2016)	It is important to find the right amount of information that increases transparency and user trust in algorithmic interfaces.	The author used an online field experiment to test three levels of system transparency in the high-stakes context of peer assessment.	Balanced interface transparency (not too little or too much) is significant in deciding user trust.
6. User heuristics (10)	Transparency can help users make quicker judgments on AI decisions.	Rai (2020)	Understanding how to achieve explainability for different types of AI models can assist users.	The author carried out a literature review to distinguish different AI models and to give an overview of explainable AI (XAI) approaches.	XAI helps fulfill prediction accuracy and interpretability objectives within AI applications, offering simpler and interpretable explanations of black-box models.

		Shin (2021)	Explainability plays an important role in user heuristics.	The author used a lab-based experiment, followed by post-exposure surveys, to test user interactions with an AI recommendation system.	Causable explainable AI helps people understand the decision-making process of AI algorithms by introducing transparency and accountability into AI systems.
		Shin and Park (2019)	The authors discussed the heuristic role of fairness, accountability, and transparency (FAT).	The authors used a mixed method approach (interpretive methods and surveys) to address FAT in algorithms.	User perceptions of FAT influence users' cognition and adoption. The interaction between trust and algorithm features affects users' satisfaction with the algorithm.

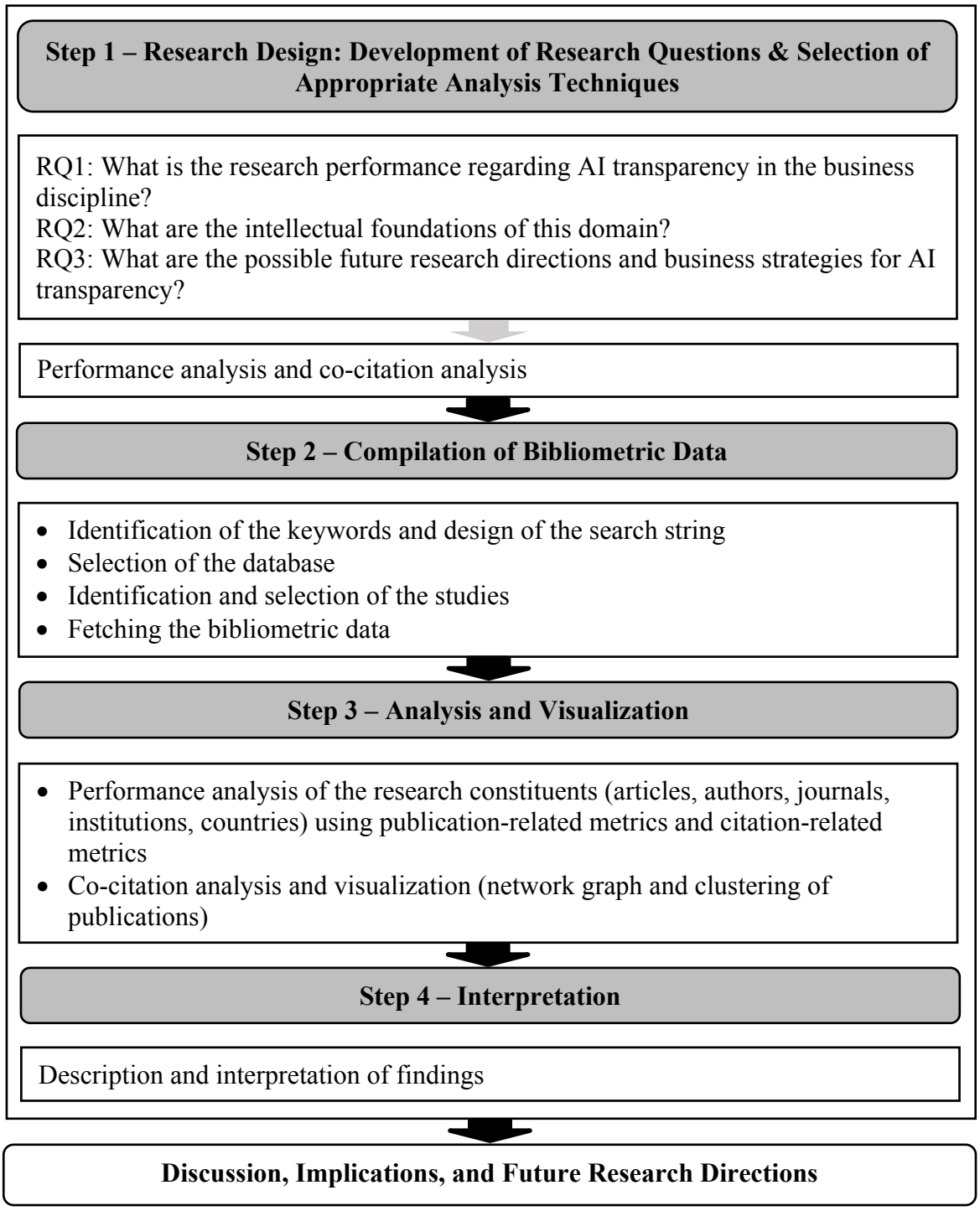
Source: Authors' own work

Table 4: Implementation of AI transparency signaling

Purpose	Low-cost Signaling	High-cost Signaling
Enhance user understanding	Using a standardized signaling approach through uniform messaging (e.g., producing infographics to explain how the system works)	• Using a customized and on-demand approach with variations in content and information quantity • Providing to the user information that is meaningful, requires low cognitive effort and is adaptable to various levels of user knowledge
Build user confidence	Dealing with transparency issues when they arise through public relations	• Conducting market research to determine barriers to trust and acceptance of AI systems • Developing strategies to build algorithmic literacy among users • Organizing co-creation opportunities
Maximize user value	Providing information to meet legal obligations, such as updating terms and conditions and governance documents	• Publishing procedures to allow for public contestability • Proactively acknowledging transparency issues and acting upon them with third-party oversight

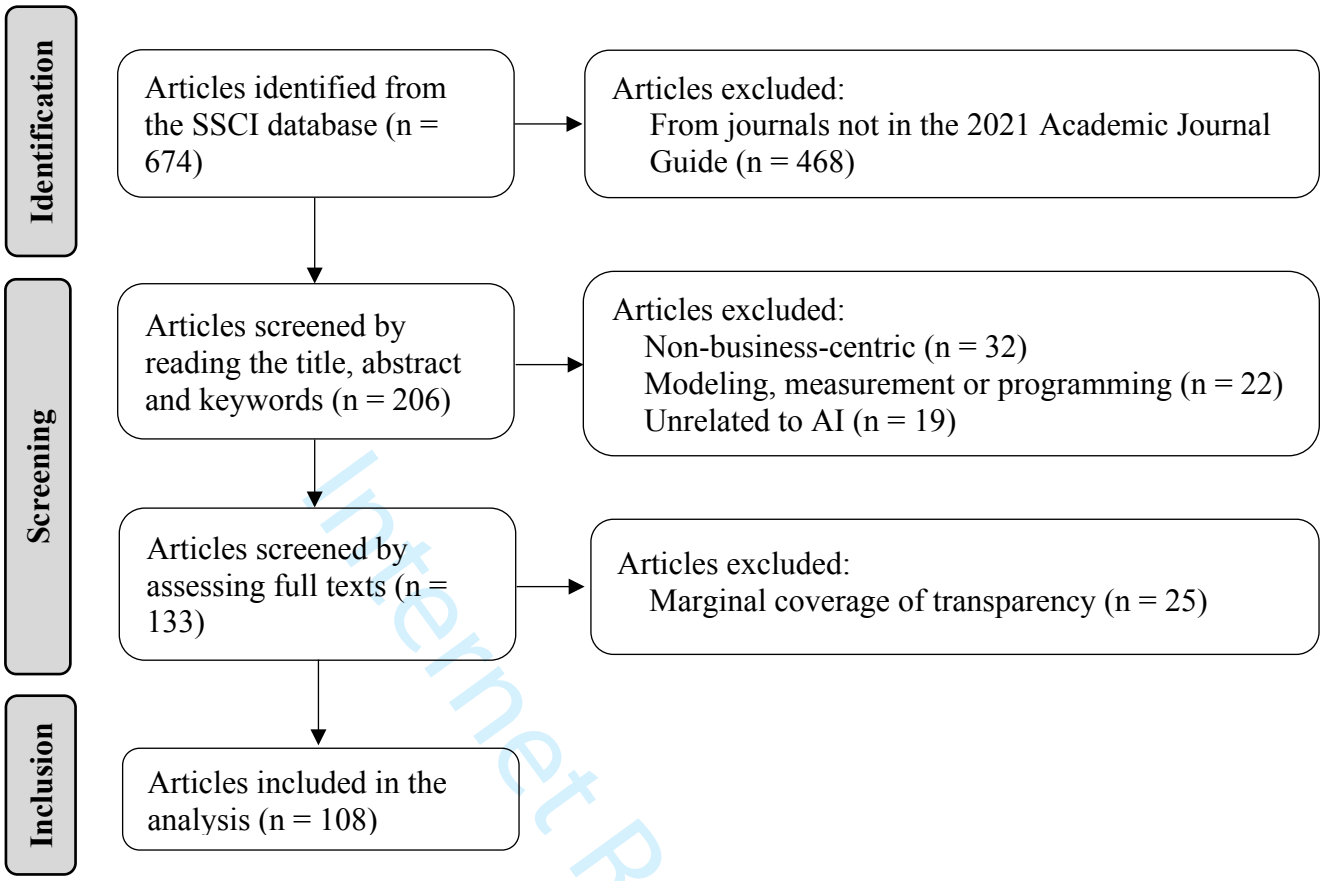
Source: Authors' own work

Figure 1. Research process



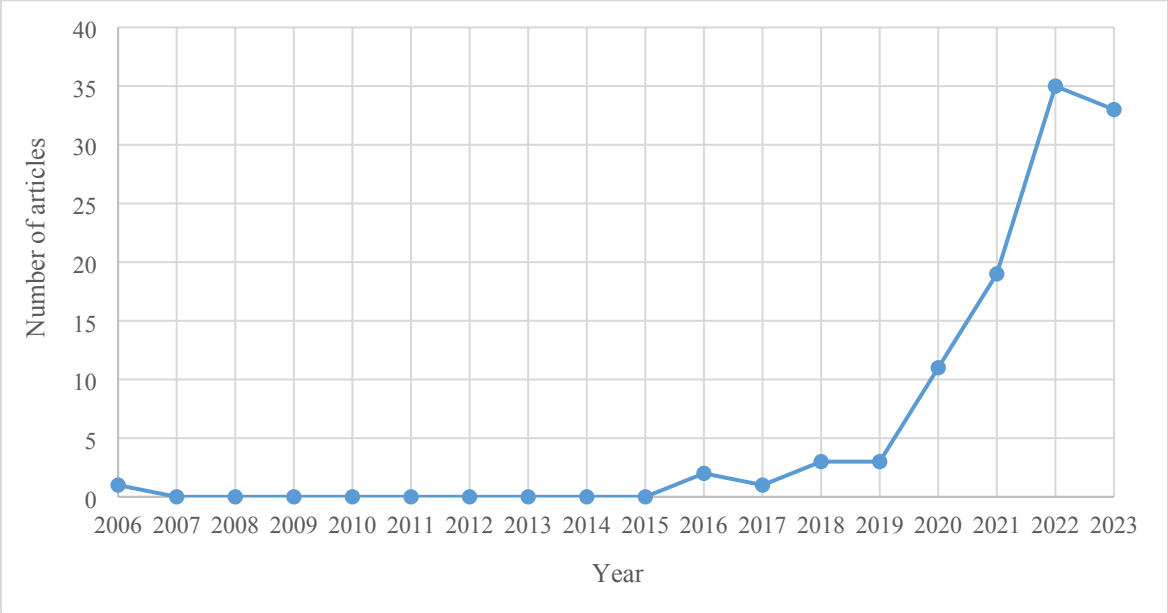
Source: Authors' own work

Figure 2. Studies identification and selection process



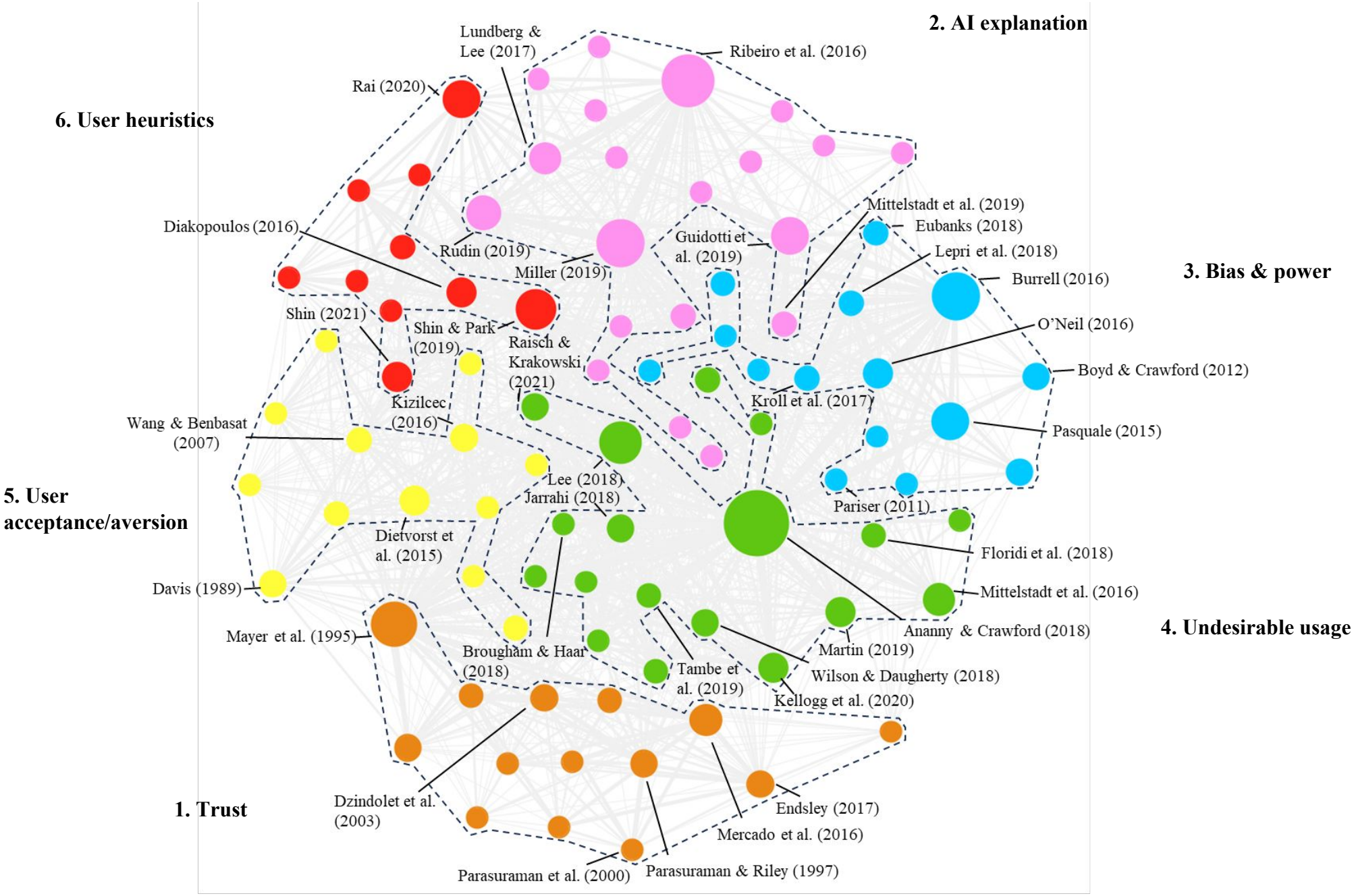
Source: Authors' own work

Figure 3. Number of yearly articles regarding AI transparency between 2006 and 2023



Source: Authors' own work

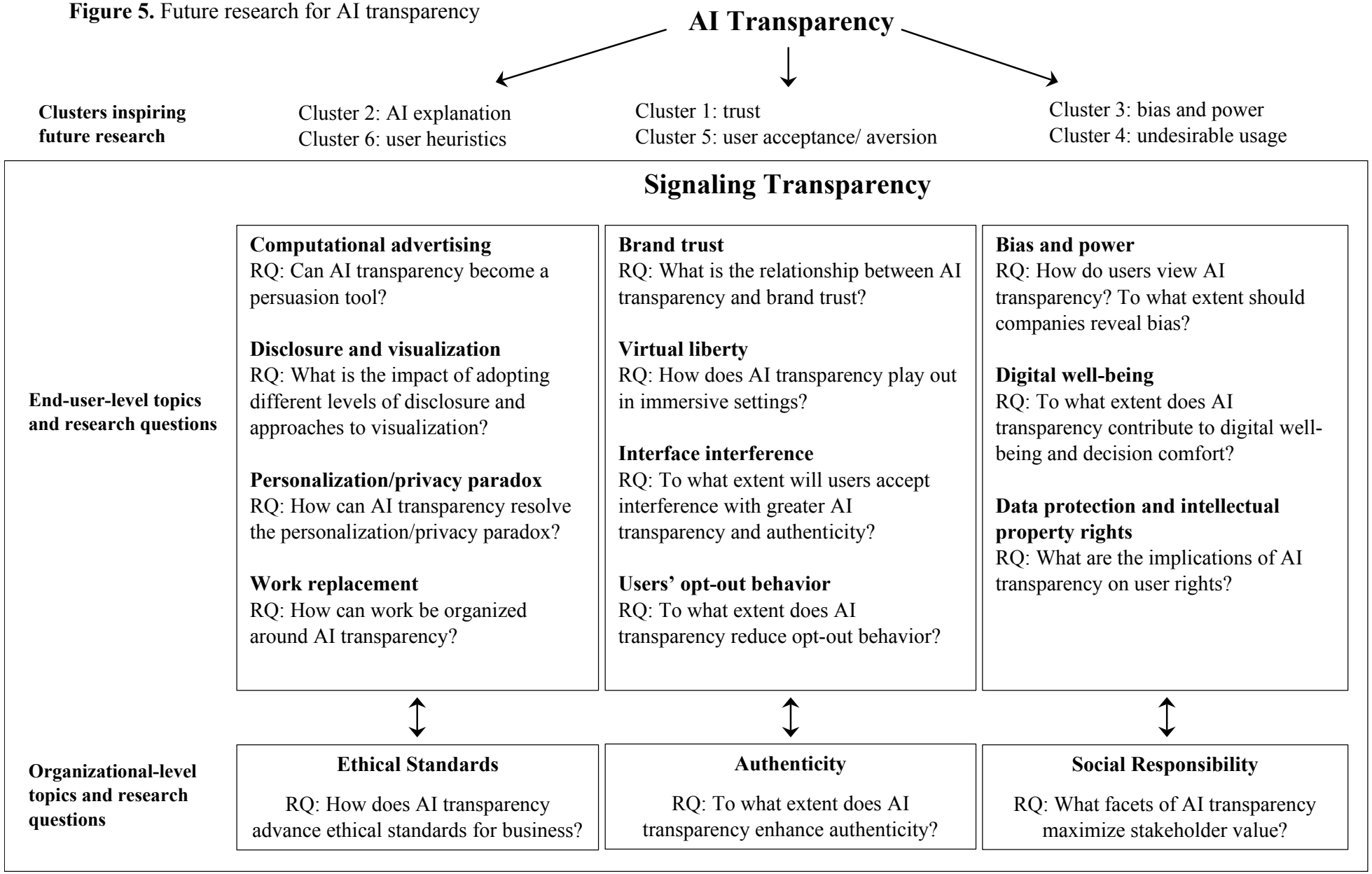
Figure 4. Dispersion of AI transparency ideas



Notes: The node (publication) size indicates the number of local citations. The strength of ties (links) between publications indicates the number of co-citations.

Source: Authors' own work

Figure 5. Future research for AI transparency



Source: Authors' own work

Web Appendix 1

Background of Bibliometric Analysis

Bibliometric analysis techniques play a crucial role in social science disciplines, facilitating the description and evaluation of existing research (Zupic and Čater, 2015). They enable researchers to determine influential scholars, papers, and themes, unveil the foundational intellectual structure of a research domain, and identify future research directions (Chabowski *et al.*, 2013; Donthu *et al.*, 2021; Ferreira, 2018). Traditional literature reviews often suffer from research biases and lack rigor (Podsakoff *et al.*, 2005; Tranfield *et al.*, 2003), whereas bibliometric analysis provides a more objective, systematic and transparent approach to literature reviews (Wilden *et al.*, 2017; Zupic and Čater, 2015).

Database Selection Criteria

WoS is known for its comprehensive coverage, encompassing over 22,000 peer-reviewed journals, 91 million records, and two billion cited references (Clarivate, 2024b). SSCI facilitates citation analysis in social sciences and has been widely employed in previous bibliometric studies (Alcaide-Muñoz and Rodríguez Bolívar, 2015; Khare and Jain, 2022). Similarly, AJG is recognized as the most authoritative business journal database and is used by all UK-based institutions and scholars for quality assessment purposes (Serenko and Bontis, 2024).

Search String

(("transparen*") AND ("artificial intelligence" OR "AI" OR “intelligent system*” OR “intelligent agent*” OR “intelligent assistant*” OR intelligent NEAR/2 assistant* OR “autonomous system*” OR “autonomous agent*” OR “virtual system*” OR “virtual agent*” OR “virtual assistant*” OR virtual NEAR/2 assistant* OR “voice assistant*” OR "robot*" OR "chatbot*" OR "social bot*" OR "socialbot*" OR "bot" OR "bots" OR "automated system*” OR “automated agent*”))

Rationale for the Louvain method and Ward's method

In recent years, network community detection has garnered increasing interest due to the availability of large networks. Unlike other algorithms, the Louvain grouping algorithm excels in speed, is independent of the size of the network, and is highly accurate (Wilden *et al.*, 2017; Fortunato, 2010; Liu *et al.*, 2012); thus, it performs exceptionally well on co-citation networks (Zupic and Čater, 2015). As a complementary grouping method, we utilized Ward's method in hierarchical cluster analysis, which is widely used in bibliometric analysis to determine subgroups (e.g., Chabowski *et al.*, 2011). Hierarchical cluster analysis is preferred over other techniques, such as multidimensional scaling, as the latter is suitable only for small datasets and fails to create explicit connections between items (Zupic and Čater, 2015). The Louvain modularity optimization method, implemented in Gephi, partitions "a network into communities of densely connected nodes, with the nodes belonging to different communities being only sparsely connected" (Blondel *et al.*, 2008, p. 2). This method is used to detect communities within large networks. The quality of the partitioning of a network into communities is measured by the modularity of the partition, which determines the density of links within rather than between communities (Blondel *et al.,* 2008). The algorithm optimizes the modularity of partitioning in a network.

Web Appendix 2

Article Performance

Among our primary documents concerning AI and transparency, eight papers have amassed over 100 citations in WoS, indicating their impact and relevance in the field (see table below). Topping the list is Glikson and Woolley's (2020) study, with 388 citations, which reviews human trust in AI literature, emphasizing AI transparency's role in cognitive trust. Next is Shin (2021), which has 269 citations. This study confirms the role of perceived transparency as a determinant of user trust in AI and as a consequence of explainability. Mercado *et al.*'s (2016) work ranks third with 166 citations, experimentally demonstrating agent transparency's positive influence on operator performance, trust, and perceived usability.

Highly cited articles based on global citations

Author(s) (Year)	Global Citations	Title	Journal
Glikson and Woolley (2020)	388	Human trust in artificial intelligence: Review of empirical research	Academy of Management Annals
Shin (2021)	269	The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI	International Journal of Human-Computer Studies
Mercado *et al.* (2016)	166	Intelligent agent transparency in human-agent teaming for multi-UxV management	Human Factors
Sundar (2020)	160	Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAII)	Journal of Computer-Mediated Communication
Rahwan (2018)	144	Society-in-the-loop: programming the algorithmic social contract	Ethics and Information Technology
Ma and Sun (2020)	132	Machine learning and AI in marketing - Connecting computing power to human insights	International Journal of Research in Marketing
De Visser *et al.* (2018)	130	From 'automation' to 'autonomy': the importance of trust repair in human-machine interaction	Ergonomics
O'Neill *et al.* (2022)	111	Human-Autonomy Teaming: A Review and Analysis of the Empirical Literature	Human Factors

Note: Global citations refer to the total number of citations a paper has received in the WoS

Source: Authors' own work

Author Performance

The creation of the primary documents engaged 328 authors. The table below displays the foremost authors based on productivity, listing those who have authored two or more papers within the dataset. Each author's total citations (global citations) are collated from WoS citations for their published articles on AI and transparency within the sample. Donghee Shin emerges as the top author, having contributed to four articles that garnered 328 citations. Following Donghee Shin are seven authors, each with two articles to their credit.

Number of articles and global citations by author

Author	Articles	Global Citations
Shin, D.	4	328
Sundar, S.S.	2	169
Chen, J.Y.C.	2	166
De Visser, E.J.	2	134
Shaw, T.H.	2	134
Kim, B.	2	53
Grimmelikhuijsen, S.	2	30
König, P.D.	2	26

Note: Global citations refer to the total number of citations a paper has received in the WoS

Source: Authors' own work

Journal Performance

The 108 primary documents were published across 55 journals (see table below). The top eleven journals, hosting three or more articles related to AI and transparency, account for nearly 52% of total publications. Following the 2021 Academic Journal Guide (Chartered Association of Business Schools, 2021) classification, these journals span various information management fields, covering ethical issues in information technology (*Ethics and Information Technology*), technology's role in government settings (*Government Information Quarterly*), the impact of technology on humans and society (*Computers in Human Behavior*, *International Journal of Human-Computer Studies*, *Internet Research,* and *Technological Forecasting and Social Change*), and the intersection of computing, information science, and information systems management (*Information Processing & Management, Journal of Strategic Information Systems*).

Number and percentage of articles by journal

Journal	Articles	%
Ethics and Information Technology	12	11.11
Government Information Quarterly	8	7.41
Computers in Human Behavior	6	5.56
International Journal of Human-Computer Studies	5	4.63
Internet Research	5	4.63
Electronic Markets	4	3.70
Journal of Computer-Mediated Communication	4	3.70
Human Factors	3	2.78
Information Processing & Management	3	2.78
Journal of Strategic Information Systems	3	2.78
Technological Forecasting and Social Change	3	2.78

Source: Authors' own work

Institution Performance

Lead authors from 112 institutions contributed to the primary documents. The table below displays the institutional productivity ranking, encompassing nine institutions involved in at least two articles. Zayed University in the United Arab Emirates stands out with four articles, consistent with the findings in the author performance section. The leading author, Donghee Shin, is affiliated with this institution, indicating its prominent role in the research area. Eight other institutions, predominantly in Western countries, share second place, each with two articles. Although their output is less than that of leading institutions, it still significantly contributes to the field.

Number of articles by institution

Affiliation	Country	Articles
Zayed University	UAE	4
Cornell University	USA	2
European Commission Joint Research Centre	Italy/Belgium	2
George Mason University	USA	2
Renmin University of China	China	2
TNO	The Netherlands	2
United States Air Force Academy	USA	2
University of Central Florida	USA	2
University of Zurich	Switzerland	2

Note: The affiliated institution was defined based on leading authorship. Fourteen leading authors had more than one affiliation.
Source: Authors' own work

Country Performance

Authors from 29 countries lead the 108 primary documents. The table below lists the most prolific countries by article count, including thirteen countries with three or more articles. The USA tops the productivity ranking with 28 articles, indicating its substantial contribution. Germany follows closely with 16 articles, representing its high level of engagement. The remaining countries contributed fewer than 10 articles, with the Netherlands and China standing out with 9 and 8 articles, respectively.

Number and percentage of articles by country

Country	Articles	%
USA	28	25.93
Germany	16	14.81
The Netherlands	9	8.33
China	8	7.41
Sweden	5	4.63
UAE	5	4.63
UK	5	4.63
Belgium	4	3.70
Italy	4	3.70
Austria	3	2.78
Canada	3	2.78
Israel	3	2.78
Switzerland	3	2.78

Note: The affiliated country was defined based on leading authorship. Seven leading authors had affiliations in two distinct countries.
Source: Authors' own work

References

Alcaide-Muñoz, L. and Rodríguez Bolívar, M.P. (2015), "Understanding e-government research", *Internet Research,* Vol. 25, No. 4, pp. 633-673.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. (2008), "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment,* Vol. 2008, No. 10, p. P10008.

Chabowski, B.R., Hult, G.T.M. and Mena, J.A. (2011), "The retailing literature as a basis for franchising research: using intellectual structure to advance theory", *Journal of Retailing,* Vol. 87, No. 3, pp. 269-284.

Chabowski, B.R., Samiee, S. and Hult, G.T.M. (2013), "A bibliometric analysis of the global branding literature and a research agenda", *Journal of International Business Studies,* Vol. 44, No. 6, pp. 622-634.

Chartered Association of Business Schools (2021), "Academic Journal Guide", available at: https://charteredabs.org/academic-journal-guide/academic-journal-guide-2021 (accessed 15 January 2024).

Clarivate (2024b), "Web of Science platform: summary of coverage", available at: https://clarivate.libguides.com/librarianresources/coverage (accessed 26 February 2024).

De Visser, E.J., Pak, R. and Shaw, T.H. (2018), "From 'automation'to 'autonomy': the importance of trust repair in human–machine interaction", *Ergonomics*, Vol. 61, No. 10, pp. 1409-1427.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim, W.M. (2021), "How to conduct a bibliometric analysis: an overview and guidelines", *Journal of Business Research,* Vol. 133, pp. 285-296.

Ferreira, F.A.F. (2018), "Mapping the field of arts-based management: bibliographic coupling and co-citation analyses", *Journal of Business Research,* Vol. 85, pp. 348-357.

Fortunato, S. (2010), "Community detection in graphs", *Physics Reports,* Vol. 486, No. 3, pp. 75-174.

Glikson, E. and Woolley, A.W. (2020), "Human trust in artificial intelligence: review of empirical research", *Academy of Management Annals,* Vol. 14, No. 2, pp. 627-660.

Khare, A. and Jain, R. (2022), "Mapping the conceptual and intellectual structure of the consumer vulnerability field: a bibliometric analysis", *Journal of Business Research,* Vol. 150, pp. 567-584.

Liu, X., Glänzel, W. and De Moor, B. (2012), "Optimal and hierarchical clustering of large-scale hybrid networks for scientific mapping", *Scientometrics,* Vol. 91, No. 2, pp. 473-493.

Ma, L., and Sun, B. (2020), Machine learning and AI in marketing–Connecting computing power to human insights, *International Journal of Research in Marketing*, Vol 37, No. 3, pp. 481-504.

Mercado, J.E., Rupp, M.A., Chen, J.Y.C., Barnes, M.J., Barber, D. and Procci, K. (2016), "Intelligent agent transparency in human–agent teaming for multi-UxV management", *Human Factors,* Vol. 58, No. 3, pp. 401-415.

O'Neill, T., McNeese, N., Barron, A. and Schelble, B. (2022), "Human–autonomy teaming: a review and analysis of the empirical literature", *Human Factors*, Vol. 64, No. 5, pp. 904-938.

Podsakoff, P.M., MacKenzie, S.B., Bachrach, D.G. and Podsakoff, N.P. (2005), "The influence of management journals in the 1980s and 1990s", *Strategic Management Journal,* Vol. 26, No. 5, pp. 473-488.

Rahwan, I. (2018), "Society-in-the-loop: programming the algorithmic social contract", *Ethics and Information Technology*, Vol. 20, No. 1, 5-14.

Serenko, A. and Bontis, N. (2024), "Dancing with the devil: the use and perceptions of academic journal ranking lists in the management field", *Journal of Documentation,* Vol. 80, No. 4, pp. 773-792.

Shin, D. (2021), "The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI", *International Journal of Human-Computer Studies,* Vol. 146, No. 102551, pp.1-10.

Sundar, S.S. (2020), "Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII)", *Journal of Computer-Mediated Communication*, Vol. 25, No.1, pp.74-88.

Tranfield, D., Denyer, D. and Smart, P. (2003), "Towards a methodology for developing evidence-informed management knowledge by means of systematic review", *British Journal of Management,* Vol. 14, No. 3, pp. 207-222.

Wilden, R., Akaka, M.A., Karpen, I.O. and Hohberger, J. (2017), "The evolution and prospects of service-dominant logic: an investigation of past, present, and future research", *Journal of Service Research,* Vol. 20, No. 4, pp. 345-361.

Zupic, I. and Čater, T. (2015), "Bibliometric methods in management and organization", *Organizational Research Methods,* Vol. 18, No. 3, pp. 429-472.

"SIGNALING TRANSPARENCY IN THE ERA OF ARTIFICIAL INTELLIGENCE" (INTR-11-2023-1041.R5)

Our Responses to Reviewer 1

Thank you very much for the time you spent providing constructive comments. Below, we have explained how we addressed your concern.

Your comment:

Recommendation: Accept

Comments:
The authors have thoroughly addressed each of the suggestions made by the two reviewers and the AE in the previous round of reviews. I particularly appreciate the improvements made in the introduction and contributions sections. The introduction is now logically structured with a clear narrative thread. Additionally, the contributions and future research sections are much clearer and more concise, demonstrating the authors' considerable effort in revising the manuscript. Therefore, I recommend accepting the paper.

However, it should be noted that I observed there might exist some inconsistencies in the reference list, particularly regarding the citation format for conference papers. There seems to be a mix of different citation styles, such as the use of full details for some conference papers and abbreviations or missing elements for others. To ensure consistency and compliance with the journal's formatting guidelines, I recommend that the authors review and standardize the conference paper citations accordingly, paying particular attention to the proper inclusion of conference names, proceedings, and page numbers where necessary.

Our response:
Thank you for your positive feedback. We are delighted to know you recommend accepting the paper for publication.

Regarding the inconsistencies in the reference list, we have reviewed and standardized the conference paper citations.

Your comment:

Additional Questions:
1. Originality: Does the paper contain new and significant information adequate to justify publication?: YES.

2. Relationship to Literature: Does the paper demonstrate an adequate understanding of the relevant literature in the field and cite an appropriate range of literature sources? Is any significant work ignored?: YES.

3. Methodology: Is the paper's argument built on an appropriate base of theory, concepts, or other

ideas? Has the research or equivalent intellectual work on which the paper is based been well designed? Are the methods employed appropriate?: YES.

4. Results: Are results presented clearly and analysed appropriately? Do the conclusions adequately tie together the other elements of the paper?: YES.

5. Implications for research, practice and/or society: Does the paper identify clearly any implications for research, practice and/or society? Does the paper bridge the gap between theory and practice? How can the research be used in practice (economic and commercial impact), in teaching, to influence public policy, in research (contributing to the body of knowledge)? What is the impact upon society (influencing public attitudes, affecting quality of life)? Are these implications consistent with the findings and conclusions of the paper?: YES.

6. Quality of Communication: Does the paper clearly express its case, measured against the technical language of the field and the expected knowledge of the journal's readership? Has attention been paid to the clarity of expression and readability, such as sentence structure, jargon use, acronyms, etc.: YES.

Our response:

Thank you for your comments and suggestions that helped us improve our paper.

Our Responses to Reviewer 2

We sincerely appreciate your positive feedback and that you recommended acceptance. Please see our responses to your comments below.

Your comment:

Recommendation: Accept

Comments:
The authors spent more than a year revising the manuscript. I don't have any concerns about this manuscript. I now think it's enough to publish. Congrats!

Additional Questions:
1. Originality: Does the paper contain new and significant information adequate to justify publication?: Yes

2. Relationship to Literature: Does the paper demonstrate an adequate understanding of the relevant literature in the field and cite an appropriate range of literature sources? Is any significant work ignored?: Yes

3. Methodology: Is the paper's argument built on an appropriate base of theory, concepts, or other ideas? Has the research or equivalent intellectual work on which the paper is based been well designed? Are the methods employed appropriate?: Yes

4. Results: Are results presented clearly and analysed appropriately? Do the conclusions adequately tie together the other elements of the paper?: Yes

5. Implications for research, practice and/or society: Does the paper identify clearly any implications for research, practice and/or society? Does the paper bridge the gap between theory and practice? How can the research be used in practice (economic and commercial impact), in teaching, to influence public policy, in research (contributing to the body of knowledge)? What is the impact upon society (influencing public attitudes, affecting quality of life)? Are these implications consistent with the findings and conclusions of the paper?: Yes

6. Quality of Communication: Does the paper clearly express its case, measured against the technical language of the field and the expected knowledge of the journal's readership? Has attention been paid to the clarity of expression and readability, such as sentence structure, jargon use, acronyms, etc.: Yes

Our response:

Thank you for your feedback and comments throughout the process, which helped us improve our paper.

Our Responses to the Internet Research Editorial Office

Thank you very much for the opportunity to improve our manuscript. Below, we explain how we addressed your concerns.

Your comment:

The manuscript has room for improvement. Below are some examples. Improve the manuscript with the aid of AI tools if necessary.

1. Consider changing "Bibliometric Analysis" in the manuscript title to "systematic literature review" and adjust the contents accordingly to better reflect the nature of the manuscript. Bibliometric analysis focus on finding research trends in a field by analyzing authorship, publication sources and so on, while systematic literature review focuses on finding research gaps in the literature using thematic or content analysis. To avoid confusion, move detailed description of the bibliometric analysis results (e.g., "Author Performance", etc.) to Web Appendices so that the main text focuses on discussion of the gaps in the research landscape. Revise other sections such as "Limitations" accordingly.

Our response:
We removed "bibliometric analysis" from the manuscript title. Following Donthu et al's (2021) study, our paper uses a series of bibliometric analysis techniques, which are quantitative in nature, for the evaluation and interpretation, with the qualitative part playing a role in interpretation only. As acknowledged by Mukherjee et al. (2022, p. 105), one of the theoretical contributions of bibliometric research is to "recognize crucial knowledge gaps to situate future research directions". This approach is adopted by several bibliometric studies (e.g., Batistič and van der Laken, 2019; Samiee and Chabowski, 2021; Wilden et al., 2017).

Following your recommendations, we moved all the descriptive analysis, including description and tables, to the Web Appendix. We also revised the limitations section.

References:

Batistič, S. and van der Laken, P. (2019), "History, evolution and future of big data and analytics: a bibliometric analysis of its relationship to performance in organizations", *British Journal of Management,* Vol. 30, No. 2, pp. 229-251.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim, W.M. (2021), "How to conduct a bibliometric analysis: an overview and guidelines", *Journal of Business Research,* Vol. 133, pp. 285-296.

Mukherjee, D., Lim, W.M., Kumar, S. and Donthu, N. (2022), "Guidelines for advancing theory and practice through bibliometric research", *Journal of Business Research*, Vol. 148, pp.101-115.

Samiee, S. and Chabowski, B.R. (2021), "Knowledge structure in product-and brand origin–related research", *Journal of the Academy of Marketing Science,* Vol. 49, No. 5, pp. 947-968.

Wilden, R., Akaka, M.A., Karpen, I.O. and Hohberger, J. (2017), "The evolution and prospects of service-dominant logic: an investigation of past, present, and future research", *Journal of Service Research,* Vol. 20, No. 4, pp. 345-361.

Your comment:

2. There are some typos and grammatical mistakes (e.g., a comma is missing from "Information Processing & Management Journal of Strategic Information Systems", add a comma before "viewing" in "Although other definitions of AI exist (Haenlein and Kaplan, 2019; Berente et al., 2021) viewing AI", "Glikson and Woolley (2020)'s study" should be "Glikson and Woolley's (2020) study", "Mercado et al. (2016) work" needs an apostrophe s, etc.). It seems that some articles (e.g., "the", "an", etc.) are missing or incorrect (e.g., "Netherlands" Web Appendix 5 should be "The Netherlands"). The spelling of some words should be changed because the manuscript uses American English (e.g., "towards" should be "toward"). Change "specific ethics construct" to "specific ethical constructs". "For example, the Alan Turing Institute's guide on AI ethics and safety (Leslie, 2019) and the Organization for Economic Co-operation and Development's (OECD) AI Principles" should be a complete sentence. Change "offering users with heuristics" to "offering users heuristics". "Accountability also simpler" should be a complete sentence. Add a comma before "but" if it is a clause (e.g., "... the credibility of sponsored advertising (Krouwer et al., 2020) but how it can also backfire"). Change "gain deeper insights into attitudes changes" to "... attitude changes". "Building user confidence is another crucial aspect in transparency signaling" should be "... aspect of transparency signaling", right? Remove the commas before "—" from "deepfakes, — digitally manipulated fake news, images, or videos, — has become".

Our response:
Thank you for your feedback and recommendations. We corrected all the typos and grammatical mistakes you identified.

Your comment:

3. The manuscript should use the required reference format and citation style (see the Author Guidelines at https://www.emeraldgrouppublishing.com/journal/intr). Volume, issue (if any) and page numbers should follow the words "Vol.", "No." and "pp." respectively (e.g., Abedin et al., 2013). A journal name, book title or conference proceedings should use the uppercase in the first letter of each content word (e.g., "Journal of informetrics" should be "Journal of Informetrics", etc.). The publication year of a reference should be within a pair of parentheses (e.g., "European Commission 2019" should be "European Commission (2019)", etc.). Remove the short form of journals (e.g., "TOIS", etc.). A reference from an electronic source must include both the URL after the words "available at:" and the access date after the word "accessed" within a pair of parentheses (e.g., House (2023), Kroll et al. (2017), Leslie (2019), etc.). Use the standard date format (e.g., "15th December 2023" should be "15 December 2023", etc.) Remove "(Ed.)^(Ed.s.)" from references without any editors (e.g., Heald, 2006; Martin, 2022). Replace the short forms of journals or conference proceedings with the full name (e.g., "Proc. ACM Hum.-Comput. Interact."). References published should be updated (e.g., Waseem et al., 2024). Different parts of a reference should be joined together with commas (e.g., after the publication year and after the paper title in European Commission (2019), etc.). Change ": p." in citations to ", p.". The web appendices should have their own reference list. The reference of some citation in Web Appendix 1 cannot be found (e.g., "Sundar (2020)", "Rahwan (2018)", "Ma and Sun (2020)", "De Visser, Pak and Shaw (2018)", "O'Neill et al. (2022)").

Our response:
Thank you for your feedback. We now use the required reference format and citation style. We added a reference list to the Web Appendices and included the missing reference of some citations.

Your comment:

4. The manuscript should be consistent. The main text should briefly describe which analysis answers RQ3 like other research questions. Sort studies in Table 2 by publication year in ascending order and author surname(s) in alphabetical order (after grouping them into categories if applicable). Studies in Table 3 in the same clusters should be sorted by alphabetical order (otherwise, include the number of local citations according to Figure 4). Cluster 5 "user acceptance/aversion" becomes "acceptance/aversion" in Figure 5. Sort items in alphabetical order (e.g., "ethics-related concepts (ethics, transparency, accountability, explainability, and fairness)"). Revise Figure 4 so that the isolated pink dot is not separated from others in the same cluster by blue and green dots.

Our response:
Further to your recommendations, we describe which analysis answers RQ3, sorted studies in Table 2 by publication year in ascending order and author surname(s) in alphabetical order within each category of knowledge synthesis method. Furthermore, studies in Table 3 have been sorted alphabetically, and we revised Figure 4 based on your suggestions. We also addressed the two additional comments related to Figure 5 and the order of the items.

Your comment:

5. The manuscript should be proofread to ensure accuracy. The main text says that "In the studies listed in Table 1, AI transparency is mentioned as a secondary element, either as one of the AI ethics principles (Birkstedt et al., 2023; Hunkenschroer and Luetge, 2022; Laine et al., 2024) or in relation to specific ethics construct, such as AI explainability (Brasse et al., 2023; Haque et al., 2023; Laato et al., 2022)", which does not mention "Ashok et al. (2022)", a study also mentioned in Table 1. Some citations in the chart in Figure 4 cannot be found in the reference list (e.g., "Endsley (2017)", "Tambe et al. (2019)", "Floridi et al. (2018)", "Jarrahi (2018)").

Our response:
Thank you for your suggestions. Ashok et al.'s (2022) study is now mentioned in the relevant section, and all citations in Figure 4 are now included in the reference list.

Your comment:

6. There are ways to improve the readability of the manuscript (e.g., contents in Table 1 to 4 and Figure 1 do not need the dot at the end if they are not complete sentences, remove footnote 2 "We

used the asterisk as a truncation symbol (e.g. transparen*) in the search string to include all possible ending variations (e.g. transparency, transparent). Our search string is available on request." and provide the search string in Figure 1 or Figure 2 or in a Web Appendix). Mention in the main text if the study complies with Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). Text in figures and tables should use the same font size as the main text (e.g., Figure 4, the chart in Web Appendix 6). Figures and tables should be self-explanatory. Figure 3 should include the x- and y-axis labels. The bar chart in Web Appendix 6 should include the y-axis label and include a key showing the meaning of different background colors of the values at the end of each bar (otherwise, make the background transparent). Change "Representative Papers Author(s) (Year)" in Table 3 to "Representative Articles". Many authors of serious English writing avoid presenting contents in point form (e.g., "the business discipline: 1. Trust...", "research gaps: (1) the elusive nature..."). If the point form must be used, change "1." and so on to "(1)" using the same format as other lists. Change "with its conceptualization" to "including its conceptualization" for clarity. Change "unrelated to AI or focused" to "unrelated to AI and focused" for clarity. Change "raise transparency effectively" to "effectively raise transparency" for clarity. Change "the neighboring cluster on Trust" to "... about Trust". Change "transparency information" in "transparency information can be provided in real-time or retrospectively" to "transparency". In "both positive and negative signals and companies can rehabilitate", add a comma before the second "and" for clarity. Change "papers clustering" to "clustering of papers" for clarity. Add a comma before "as" in "Co-citation analysis primarily looks backward (Chabowski et al., 2013) as accumulating many citations". In "For instance, they can analyze pre-purchase", change "they" to "researchers" for clarity. Sentences describing prior studies should use the past tense (e.g., "Diakopoulos (2016) emphasizes", "Rai (2020) advocates").

Our response:
Thank you for your suggestions. We addressed the above comments.

Your comment:

7. The manuscript should not be too much longer than the upper limit of 9500 words. Some contents can be changed to Web Appendices instead (e.g., description of the bibliometric analysis results). Remove the DOI from references. Repeated information can be removed.

Our response:
Thank you for your comments. We reduced the number of words as much as possible.

Your comment:

1. There are some typos (e.g., "Commununications" in the reference list). Some articles (e.g., "the", "an", etc.) are missing or incorrect (e.g., "due to lack of technical expertise", "The description and rationale of bibliometric approach", "e.g. Chabowski et al., 2011" should be "e.g., Chabowski et al., 2011"). The items after "The novelty of this study lies in filling the following research gaps:" should not include a dot at the end as they are not complete sentences. "The under exploration of AI transparency" should be "The underexplored AI...". Break "Article numbers remained low until

2019, but saw a steady increase, with a surge in 2021 and 2022, reflecting growing research interest, highlighted by the 33 articles published in 2023" into two sentences for clarity. Remove the repeated citation from "Kizilcec (2016) showed that excessive or insufficient explanations can engender distrust in the system when users are initially disadvantaged (Kizilcec, 2016)". Change "Following is Shin (2021), with 269 citations" to "Next is Shin (2021), which has 269 citations". Break "Zayed University in the United Arab Emirates stands out with four articles, consistent with the findings in the author performance section, as the leading author, Donghee Shin, is affiliated with this institution, indicating its prominent role in the research area" to two sentences.

Our response:
Thank you for your suggestions. We addressed the above comments as suggested.

2. The manuscript should use the required reference format and citation style (see the Author Guidelines at https://www.emeraldgrouppublishing.com/journal/intr). A reference should be complete and should not include the words "et al." among author names (e.g., "Floridi, L., Cowls, J., Beltrametti, M. et al. (2018)"). Different parts of a reference should be joined together with commas (e.g., '(2016). "'I always' should be '(2016), "'I always', '(2023). "Netflix' should be '(2023), "Netflix', 'interaction (HAII)". Journal' should be 'interaction (HAII)", Journal'). A paper title should use the lowercase except the first letter of the title or proper nouns (e.g., "Academic Journal Guide" should be "Academic journal guide", "EU Artificial Intelligence Act", "General Data Protection Regulation", "Artificial Intelligence in Human Resources Management: Challenges and a Path Forward"). Access dates should use the appropriate format (e.g., "26th February 2024" should be "26 February 2024"). "(Ed.s.)" should be "(Ed.s)". Give the surname of the first only if a citation has more than two authors (e.g., "De Visser, Pak and Shaw (2018)" should be "De Visser et al. (2018)").

Our response:
Thank you for pointing our mistakes. We really appreciate it.

3. The manuscript should be consistent. Make all findings in Table 3 complete sentences. Change "Articles removed before screening due to journals not in the 2021 Academic Journal Guide (n = 468)" in Figure 2 to "Articles excluded: From journals not in the 2021 Academic Journal Guide (n = 468)" following the style of other boxes. Change the stage "Included" in FIgure 2 to "Inclusion" following the style of the previous two. Consider changing "Cluster 2: explaining AI" to "Cluster 2: AI explanation" following the style of other clusters.

Our response:
Thank you for your suggestions. We addressed the above comments as suggested.

4. There are ways to improve the readability of the manuscript. "This study addresses the following research gaps: first, the elusive nature of AI transparency and its knowledge basis; second, the under-exploration of AI transparency in the business discipline, compared to information sciences and law; third, the ambiguity surrounding the implementation strategies for AI transparency, with companies often resorting to simplistic methods such as updating terms and conditions; and fourth, a lack of clear future research directions specifically for AI transparency, rather than within the broader context of AI ethics" should be broken into multiple sentences as "... research gaps. First,

the... basis. Second, the... law. Third, the... conditions. Fourth, a...". "This process led to the exclusion of non-business-centric articles (ergonomics, education, military science), unrelated to AI or focused on modeling, measurement, or programming" should be "... science), articles unrelated to AI, or articles focused on...". "A lack of clear future research directions specifically for AI transparency, rather than under the general context of AI ethics" should be "... transparency, as opposed to the broader context of AI ethics". "Second, this study provides a more stakeholder-oriented definition of AI transparency while past studies focused more on organizational information disclosure" may include a comma before the "while" clause. Consider changing "We used the references in the primary articles to identify six clusters spatially represented in Figure 4 to reveal the literature's intellectual structure" to "Using the references in the primary articles, we identified six clusters, spatially represented in Figure 4, to reveal the literature's intellectual structure" for clarity. Consider changing "Performance and co-citation" in Figure 1 to "Performance and co-citation Analysis" or "Performance and co-citation data" for clarity. "Fourteen leading authors used more than one affiliation" should be "Fourteen leading authors had more than one affiliation". "Seven leading authors used affiliations in two distinct countries" should be "... authors had affiliations...".

Our response:
Thank you for your suggestions. We addressed the above comments as suggested.

1. There are some typos (e.g., "ai" in the reference list). Change "comprising 108 primary articles and 7,459 secondary cited studies" to "... cited articles" for accuracy. "Thus, the more two documents are cited together, the more similar they are" should be "Thus, the more frequently two..."

Our response:
We have reviewed these sentences. As the cited studies are not limited to articles, we have changed the word 'studies' to documents.

2. The manuscript should use the required reference format and citation style (see the Author Guidelines at https://www.emeraldgrouppublishing.com/journal/intr). All ", &" between author names in references should be "and" without the comma (e.g., "Tambe, P., Cappelli, P., & Yakubovich, V." should be "Tambe, P., Cappelli, P. and Yakubovich, V.", "De Visser, E. J., Pak, R., & Shaw, T. H.", "O'Neill, T., McNeese, N., Barron, A., & Schelble, B."). Remove the spaces between the initials of each author or editor name in references (e.g., "De Visser, E. J." should be "De Visser, E.J.", etc.).

Our response:
Thank you for pointing these out. We addressed the above references as suggested.

3. There are ways to improve the readability of the manuscript. Consider changing "The underlying principle of co-citation is that if two secondary publications are cited together in the same primary document, they are similar thematically" to "... is that two secondary publications are considered as similar to each other thematically if they are cited together in the same primary document.

Our response:

Thank you for your suggestion. We changed the sentence to the following. "The underlying principle of co-citation is that two secondary publications are considered thematically similar if they are cited together in the same primary document."

If applicable, acknowledge the conference proceedings or journal paper upon which this manuscript is developed.

Please revise the manuscript thoroughly and submit the revised version. Thank you.

Our response:
Thank you for your feedback, which has helped us improve our paper.