A Framework for Context-Specific Theorizing on Trust and Reliance in Collaborative Human-AI Decision-Making Environments

Completed Research Paper

Niko Spatscheck

Julius-Maximilians-Universität, Chair of Management and Information Systems, Würzburg, Germany
niko.spatscheck@uni-wuerzburg.de

Abstract. Collaborative human-AI decision-making is an emerging concept, that recognizes the complementary qualities of humans and AI systems for decision-making. However, to achieve a successful collaboration between both, individuals must appropriately trust and rely on the AI systems. Despite extensive research on trust and reliance within collaborative decision-making, empirical research indicates inconclusive findings. We aim to clarify these findings by incorporating context-specific theorizing as a crucial element of a broader, integrative perspective, emphasizing that the neglect of contextual factors provides a key explanation for the observed inconclusive results. We review 59 empirical studies, culminating in a synthesized framework that highlights the often unrecognized contextual factors shaping decision-making environments and influencing trust and reliance. We inductively conceptualize human-related, AI-related, and decision-related contextual factors and their interaction effects into our framework. Our work contributes to the emerging literature on collaborative decision-making and highlights the importance of context-specificity.

Keywords: AI Systems, Trust, Reliance, Collaborative Decision-Making

1 Introduction

The widespread adoption of AI systems is fundamentally transforming human life, with AI becoming pervasive across personal and professional domains (Maedche et al. 2019). While AI shares foundational principles with earlier algorithms, it is distinguished by its greater complexity, opaqueness, and capacity to augment even intellectual and creative tasks (Vössing et al. 2022, Jussupow et al. 2021). Many researchers see AI as a complement to human decision-making, combining the strengths of both to achieve better outcomes (Harper 2019, Seeber et al. 2020, Vössing et al. 2022). The main advantage of human-AI collaboration is its ability to enhance decision-making accuracy beyond what either could achieve alone (Dellermann et al. 2019). This is particularly valuable in decision-making scenarios, where humans are often limited by cognitive biases and inconsistent reasoning (Dellermann et al. 2019).

However, achieving optimal collaboration performance is challenging due to humans' tendency to miscalibrate trust in AI, which can impair appropriate reliance and thus

20th International Conference on Wirtschaftsinformatik
September 2025, Münster, Germany

reduce collaborative decision accuracy (Schemmer et al. 2023). Both overtrust and undertrust have serious consequences: overtrust can lead to overestimating AI's capabilities and relying on incorrect advice (Lee & See 2004, Hoff & Bashir 2015), while undertrust can result in ignoring correct AI advice (Lee & See 2004, Hoff & Bashir 2015).

Despite significant research on trust and reliance on AI systems within collaborative decision-making (Hoff & Bashir 2015), two main limitations persist: First, trust is often ambiguously conceptualized as either an attitude or a behavior, blurring the distinction with reliance, which creates theoretical confusion and misinterpretations of empirical findings (Schemmer et al. 2023, Schmitt et al. 2021, Burton et al. 2020). Second, conflicting empirical findings have emerged, ranging from overtrust to undertrust in AI systems (Leichtmann et al. 2023, Gomez et al. 2023) and from overreliance to underreliance (Vasconcelos et al. 2023, He et al. 2023), as well as from AI aversion to appreciation (Jussupow et al. 2024).

Building on Hong et al. (2014), we argue that the under-theorization of context has contributed to these equivocal findings and limits the generalizability of existing research in behavioral IS. A deeper focus on context-specific factors would provide both scholars and practitioners with a more actionable understanding of trust and reliance in human-AI decision-making environments, leading to improved AI system design, better usage, and ultimately, enhanced collaboration performance in terms of decision accuracy. In response to these challenges and previous calls for research (Venkatesh 2025), our study aims to address the following question: *What are the contextual factors that influence (appropriate) trust and reliance in collaborative human-AI decision-making environments?*

To answer this, we build on Lee & See (2004) and Hoff & Bashir (2015) and conduct a systematic literature review, synthesizing recent empirical studies to develop a conceptual framework that integrates previously overlooked contextual factors that affect the formation of trust and reliance. In doing so, our framework offers a theoretical lens for future research, providing greater specificity and applicability across diverse decision-making contexts.

2 Theoretical Background: Trust and Reliance on AI

Human-AI collaboration involves the synergy between humans and AI systems working together to achieve shared objectives. This collaboration can range from fully automated systems, where AI systems operate independently, to augmented collaboration, where humans remain actively involved and AI assists in task execution (Raisch & Krakowski 2021). A prominent domain within augmented collaboration is decision-making, where AI systems provide data-driven insights that individuals can incorporate into their judgment, ideally resulting in more informed and accurate decisions (Lai et al. 2023, Hemmer et al. 2024). This approach harnesses the complementary strengths of humans such as emotional intelligence and ethical reasoning, with AI's capacity to process large datasets and detect patterns (Hemmer et al. 2024). The focus of our study is on augmented human-AI collaboration in decision-making contexts, e.g. in situations where fully automated systems are either undesirable or legally infeasible.

The success of human-AI collaboration is contingent upon the level of trust individuals place in AI systems and the extent to which they rely on AI-generated advice. Trust plays a foundational role in determining reliance behavior, however, the relationship between trust and reliance is complex. As highlighted by Lee & See (2004), trust can positively shape how individuals interact with AI systems, but it can also lead to both misuse and disuse of AI systems. Misuse occurs when individuals overtrust AI, excessively relying on its advice even when it is inaccurate (overreliance) (Parasuraman & Riley 1997, Lee & See 2004). Conversely, disuse arises when individuals undertrust AI, failing to utilize its advice even when it is accurate (underreliance) (Parasuraman & Riley 1997, Lee & See 2004). In this context, appropriate trust and reliance refer to the optimal balance, where individuals accurately assess the reliability of AI advice and integrate it effectively into their decision-making. Achieving this appropriateness requires individuals to trust AI enough to rely on accurate advice while retaining the judgment to override flawed advice, ultimately maximizing collaborative performance and ensuring optimal decision accuracy. However, calibrating trust and reliance appropriately is a complex endeavor, shaped by various contextual factors such as cognitive biases and situational influences, which we explore in this paper.

For conceptual clarity, we briefly define trust and reliance, recognizing that both are multifaceted and complex constructs, particularly within the interdisciplinary field of human-AI collaboration (Montealegre-López 2025), where the comparison and generalization of research findings can be challenging. In this study, we adopt a widely accepted and used definition of trust and reliance within the information IS literature. Accordingly, trust is defined as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See 2004, 51). Integrating this trust model into the typology of Fishbein & Ajzen (1977) and previous IS research on attitudinal trust further elucidates the causal chain of trust formation and the subsequent reliance behavior (McKnight et al. 2020, You et al. 2022, Qiu & Benbasat 2009, Lee & See 2004). Accordingly, subjective and attitudinal trust influences observable reliance behavior, guiding but not fully determining an individual's reliance behavior on AI systems. Consistent with the prevailing views in IS literature (Lee & See 2004, Hoff & Bashir 2015, You et al. 2022, Westphal et al. 2023, Castelo et al. 2019), we thus conceptualize individuals' trust in AI as a mediating factor that influences, but does not unilaterally dictate, their reliance hehavior on AI systems.

3 Methodology

We conducted a systematic literature review in accordance with the methodological guidelines of Webster & Watson (2002) to ensure rigor, transparency, and reproducibility in synthesizing research on trust and reliance in human-AI decision-making. We queried ACM Digital Library, AIS eLibrary, ScienceDirect, and Web of Science, using a standardized search string to the title and abstract fields of each database: ("Artificial Intelligence" OR "AI" OR "Algorithm" OR "Machine learning" OR "Deep learning") AND ("Trust" OR "Reliance") AND ("Decision*"). We constrained our database queries to retrieve only completed, peer-reviewed research articles, systematically excluding records classified as books, review articles, short papers, dissertations, keynotes, or panel discussions.

Moreover, to ensure coverage of recent advancements and to minimize overlap with prior reviews (notably Hoff & Bashir (2015)) we limited our search to publications from 2015 onwards. Applying these initial inclusion and exclusion criteria yielded 1193 papers. After removing duplicates, inaccessible full texts, and non-English publications, 876 studies remained. We then screened titles and abstracts to include only empirical studies that explicitly addressed trust and reliance in human-AI decision-making, e.g. excluding those that focused solely on either the human or AI perspective. This process reduced our set to 125 studies. Subsequently, we conducted a comprehensive full-text analysis, applying the same inclusion criteria as previously outlined. Additionally, we excluded all non-empirical studies, as well as empirical studies that did not align with our theoretical conceptualizations of trust and reliance as defined in our theoretical background; for instance, omitting studies that conceptualized trust exclusively as a behavioral construct. This process reduced our corpus to 49 studies. Finally, we employed a snowballing approach by conducting backward and forward citation searches of highly cited primary studies, which resulted in 10 additional empirical studies (seven through backward and three through forward citation), resulting in a final corpus of 59 studies.

The majority of studies in our final corpus employed incentivized, factorial online experiments, with 41 utilizing a between-subjects design and 4 a within-subjects design. Field experiments (n = 3), survey-based studies (n = 7), and mixed-method approaches (n = 4) were less common. Most experiments utilized vignette-based scenarios, predominantly using decision-making environments related to healthcare (n = 11), finance (n = 13), and human resources (n = 14); studies examining leisure, gaming, or abstract problem-solving contexts were comparatively rare. Furthermore, participant recruitment was primarily conducted via international crowdsourcing platforms such as Prolific and Amazon Mechanical Turk. Sample sizes ranged from 50 to over 1,000 participants, though most studies involved 200–300 individuals. Recruitment strategies were generally aligned with the requirements of the specific decision-making context, with participant expertise and task familiarity ranging from novice students to employed domain experts with several years of relevant experience.

To address our research question, we developed a comprehensive conceptual framework based on the 59 studies that incorporates "contextual factors as antecedents of core constructs or dependent variables" and as "moderators of relationships" (Hong et al. 2014, 115). To this end, we employed a rigorous hybrid process, combining deductive structuring with inductive conceptual development. We began by delineating three overarching categories: *Human*, *AI*, and *Decision* - drawing deductively on Maedche et al. (2019) and Hemmer et al. (2021)'s conceptualization, itself grounded in Task-Technology Fit Theory (Goodhue & Thompson 1995). Within each of these categories, we employed an inductive approach to identify and elaborate relevant contextual factors, guided by the eight-phase conceptual framework development procedure outlined by Jabareen (2009). This iterative process involved comprehensive mapping, systematic coding, and constant comparative analysis of the 59 empirical studies. Across four iterative cycles of in-depth analysis, conceptual mapping, and coding, as well as integration and synthesis, we inductively aggregated the initially identified 28 contextual factors and systematically consolidated these into four to five salient factors per overarching category (in total n=14 factors). This approach ensured both parsimony and conceptual rigor in the resulting framework (Jabareen 2009).

4 Literature Review and Framework Synthesis

Our framework (cf. Figure 1) contextualizes the collaborative human-AI decision-making environment via human-related, AI-related, and decision-related contextual factors. As illustrated in our framework, these factors and their interaction effects shape the decision-making environment which has a subsequent impact on the formation of trust as a core mediating role of reliance. Although trust plays a key role, it does not fully determine reliance (Lee & See 2004, Hoff & Bashir 2015), making it essential to consider the combined impact of all contextual factors on reliance directly, rather than relying solely on trust as a mediator. Consistent with Schemmer et al. (2023), we consider reliance as a case-by-case decision, with the aggregate of all reliance cases collectively contributing to complementary collaboration performance, which is associated with achieving the best or highest decision accuracy. Additionally, trust and reliance are dynamic, a process that is "prone to changes based on the behavior of the trusted agent" (Glikson & Woolley 2020, 630). We expand on this perspective by asserting that these dynamics are shaped not only by the agent's behavior but also by the interplay of contextual factors, which evolve over time, thus requiring a feedback mechanism.

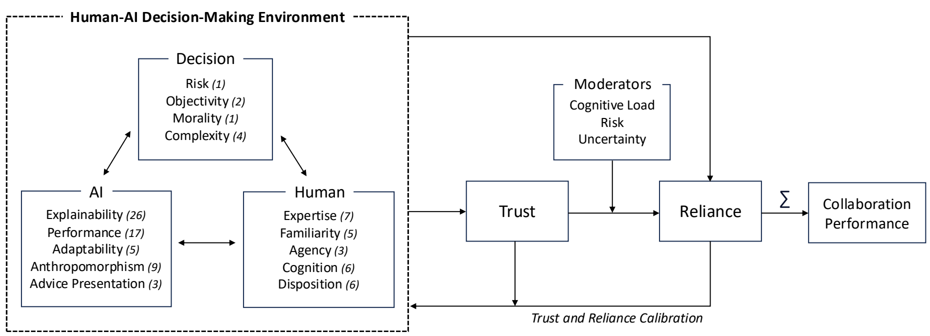

Figure 1. Conceptual Framework for Collaborative Human-AI Decision-Making Environments *(Note: The numbers in parentheses indicate how many papers in our literature corpus examined the respective factor*)

4.1 AI-related Contextual Factors

Explainability: Research on Explainable Artificial Intelligence (XAI) shows mixed effects on trust and reliance on AI systems. Studies by Lai & Tan (2019), Mehrotra et al. (2023), and Buçinca et al. (2021) suggest that providing explanations increases trust, but the impact depends on the type of explanation. Comprehensive explanations boost trust but can lead to over-reliance, while simpler explanations may reduce trust and reliance (Bussone et al. 2015). Visual explanations tend to lower trust but improve performance by mitigating overtrust (Leichtmann et al. 2023). Moreoever, XAI can reduce overtrust and reliance by making AI errors more visible, which lowers the cognitive effort needed to evaluate advice (Vasconcelos et al. 2023).

As illustrated in our framework, interaction effects exist between AI-related, human-related, and decision-related factors. The impact of XAI on trust and reliance is contingent not only on the system's design but also on user knowledge and cognitive ability to process complex decisions (Westphal et al. 2023). For example, Bayer et al. (2022) found that explanations increased trust in individuals with high domain expertise, while those with limited knowledge struggled to understand them. This finding aligns with prior research suggesting that insufficient domain knowledge can diminish the effectiveness of XAI (Wang & Yin 2021, Bayer et al. 2022). However, well-crafted explanations can help less knowledgeable users by reducing automation bias (Wysocki et al. 2023). Moreover, Zhang et al. (2024) found that explanations are effective in reducing overreliance in simple decision tasks, but this effect diminishes as decision complexity increases; for highly complex decisions, explanations may even exacerbate overreliance. Similar patterns were observed by de Jong et al. (2025), who emphasized the need to account for decision difficulty when evaluating the effectiveness of explanations. These nuanced interactions may help explain the mixed, and sometimes paradoxical, findings reported in the XAI literature (Kahr et al. 2023, Ueno et al. 2022, Schemmer et al. 2023, Goel et al. 2022).

Performance: Even though it may seem evident at first glance, it is not always empirically supported that individuals trust high-accuracy AI systems more than low-accuracy AI systems and consequently rely more on the advice of high-accuracy AI systems (Kahr et al. 2023). While the majority of empirical studies have demonstrated a positive relationship between increased AI performance and trust, reliance, and human-AI collaboration performance (Cao & Huang 2022, Papenmeier et al. 2022, Zhang et al. 2022), some studies found no effects (Schmitt et al. 2021). In fact, Juravle et al. (2020) found that individuals do not increase their trust in AI systems with increasing AI performance. Moreover, even in the absence of explicit information about an AI system's performance, individuals still assess its accuracy and adjust their trust and reliance accordingly (Cao & Huang 2022), with the AI's response time serving as an important performance cue (Efendić et al. 2020). Thus, both stated and observed accuracy are important factors, whereby the effects of stated accuracy can be overridden by individuals' observed accuracy (Yin et al. 2019).

As shown in our framework, trust and reliance formation is a dynamic process that can undergo multiple feedback cycles. Therefore, individuals' trust in AI systems can increase over time when collaborating with high-accuracy AI systems, as individuals increasingly perceive the AI system's competence and consequently rely more on its advice (Kahr et al. 2023). Furthermore, research demonstrated that trust and reliance are at their lowest when the output of an AI system overlaps with an individual's knowledge and at their highest when there is perfect complementarity, suggesting a nuanced interaction effect between AI performance and human knowledge (Zhang et al. 2022, Inkpen et al. 2023).

Adaptability: Our review indicates that AI systems' adaptability has a significant impact on individuals' trust and reliance. Specifically, when individuals perceive AI systems as adaptable or capable of learning from their environment, they tend to have considerably higher levels of trust, which leads to increased reliance on the system (Lohoff & Rühr 2021, Kim & Song 2023*b*). Demonstrating or framing an AI system's adaptability in terms of its dynamic learning capabilities is also a promising strategy to counteract algorithm aversion (Berger et al. 2021). However, Xu & Xu (2021) discovered that showcasing an AI system's adaptability, particularly through feedback from users, can create the impression that the system needs continuous improvement or supervision, which can ultimately reduce reliance on the AI. Another key factor is the AI system's ability to adapt to an individual's preferences and provide personalized advice based on this understanding, which further strengthens trust in the system (Shi et al. 2021).

Anthropomorphism: Extensive empirical research indicates that anthropomorphizing AI systems increases individuals' trust in these systems (Li & Hahn 2022, Shi et al. 2021). Furthermore, anthropomorphism not only elicits increased trust in AI systems but also reduces the perceived psychological distance to the AI (Park et al. 2023, Zhang et al. 2022) and increases social presence (Morana et al. 2020). Anthropomorphic cues such as a human voice (Schreuter et al. 2021), human images (Park et al. 2023), self-references (Zhang et al. 2022), or a combination of human identity, verbal and nonverbal cues (Morana et al. 2020) also increase individuals' reliance on the AI system's advice. In addition, an individual's disposition to anthropomorphize and the nature of the decision task (whether it is computer-like or human-like) also significantly affects anthropomorphism, which in turn affects trust and reliance (Seeger et al. 2021). However, anthropomorphizing highly human-like AI systems can evoke an uncanny sense of familiarity, which can also reduce users' trust in these systems (Spatscheck et al. 2024).

Advice Presentation: The presentation of AI advice covers both quantitative and qualitative aspects. Regarding the quantitative aspects, Newman et al. (2022) found that the number of advice options is crucial, as individuals are more likely to rely on AI advice when presented with multiple advice alternatives. Their study found that individuals preferred multiple advice options to a single advice, even when the quality of the single advice was better than the combined multiple advice options, suggesting that decision autonomy is more important to individuals than the quality of advice.
While research on the qualitative aspects of AI-generated advice remains limited, recent advances in AI have enabled the generation of nuanced, text-based advice with considerable qualitative variation. In this emerging line of inquiry, Rubin & Benbasat (2023) examined various forms of textual justification accompanying AI advice and found that such justifications play a crucial role in fostering individuals' trust in AI systems. Likewise, Zhang et al. (2022) explored the influence of different belief markers within AI-generated advice, revealing that embracing belief markers (e.g., "know" and "realize") led to significantly greater reliance on AI advice compared to distancing markers (e.g., "think" and "believe"), though at the cost of increased overreliance. In a similar vein, Kim et al. (2024) demonstrated that the use of natural language expressions of uncertainty can mitigate overreliance on large language models. Furthermore, including disclaimers such as "Remember to verify this information" within qualitative advice has been shown to reduce both overreliance and under-reliance in certain tasks, suggesting an interaction effect with cognitive forcing (Bo et al. 2025). Finally, Kim et al. (2025)

found that the presence of inconsistencies in qualitative AI advice leads to decreased overreliance.

4.2 Decision-Related Contextual Factors

Risk: When making high-risk decisions, individuals tend to place less trust in AI systems than when making low-risk decisions (Juravle et al. 2020). The decrease in trust is especially evident when AI systems are involved in high-stakes decision-making environments, in which potential decision errors are associated with high prospective loss outcomes and typically cannot be reversed once a decision has been made (Kunreuther et al. 2002). Furthermore, individuals expect a higher level of accuracy from AI systems in high-risk decision-making situations but are still less likely to rely on AI advice for such decisions (Juravle et al. 2020). Instead, individuals tend to prefer advice from other humans in these decision-making environments, indicating algorithm aversion (Jussupow et al. 2020, Longoni et al. 2019). Such a decision strategy can be explained by the tendency of individuals to rely on human judgment when they believe that AI systems are unlikely to produce near-perfect results and therefore prefer even less reliable and error-prone human advice when making risky decisions (Dietvorst & Bharti 2020).

Objectivity: The perceived objectivity of a decision further affects trust and reliance. According to Lee (2018), in machine-like decision tasks, both AI and human advice were perceived as similar in terms of fairness and trustworthiness. However, in tasks that require more human-like decision-making, AI systems' advice was perceived as less fair and trustworthy due to perceived deficits in intuition and subjective judgment. Castelo et al. (2019)'s research confirmed that individuals trust and rely less on AI advice for subjective tasks. Conversely, they found that trust in an AI system can be increased by increasing the perceived objectivity of a decision task.

Morality: Similar to the perceived risk of a decision, the morality of the decision also influences the extent to which individuals trust AI systems in decision-making environments and rely on their advice. Although research on moral decision-making is still limited, the studies that have been conducted show consistent results. Based on nine conducted studies, Bigman & Gray (2018) concluded that individuals tend to rely less on AI systems when making morally sensitive decisions in domains such as driving, legal matters, or the military. This reluctance arises because "it seems that we want another human - with a fully human mind - to make the call" in highly moral decision-making environments (Bigman & Gray 2018). Furthermore, the study revealed that this aversion remains even when the AI advice leads to a positive outcome and may only be alleviated by limiting the AI system to an advisory role, increasing the perceived expertise of the AI system, or increasing the perceived experience of the AI system (Bigman & Gray 2018). Although Gogoll & Uhl (2018)'s study did not directly involve AI systems but rather machine agents, the results also suggest that individuals are reluctant to delegate decisions to machines in moral domains.

Complexity: The inherent complexity of a decision also affects trust and reliance on AI systems. Individuals selectively rely more on AI advice when making difficult, and therefore more complex, decisions as compared to easier ones (Kaufmann 2021, Parkes 2017). Vasconcelos et al. (2023) corroborated this finding by showing that individuals' reliance on AI systems increases with the complexity of the decision. These findings were attributed to the increased cognitive effort required to either confirm the correctness of an AI system's advice or perform the decision task independently. Surprisingly, de Jong et al. (2025) observe that participants' collaborative performance is significantly lower for medium-difficulty tasks compared to both easy and hard tasks in both studies, highlighting the importance of differentiating decision support strategies according to task difficulty (de Jong et al. 2025).

4.3 Human-related Factors

Expertise: Individuals' expertise for example related to their domain knowledge, can significantly impact trust and reliance on AI systems. Individuals who possess a high level of expertise in a particular domain are typically more capable of discerning when to rely on AI advice and when to use their own judgment, promoting appropriate reliance on AI systems (Parkes 2017, Inkpen et al. 2023). Similarly, Wysocki et al. (2023) found that less experienced healthcare professionals compared to experienced ones had more trust in the AI system and relied more on the AI advice. Additionally, individuals' trusting beliefs and trusting intentions in the AI system decreased with increasing domain-specific expertise (Bayer et al. 2022). Although Lacroux & Martin-Lacroux (2022) found paradoxical results in this regard, they were attributed to a flawed measure of expertise, namely in the context of a traditional rather than an algorithmic recruitment process.

Familiarity: Our review shows that individuals' familiarity with the decision task or with the AI system plays another central role in shaping individuals' trust and reliance. Individuals who are initially more familiar with the decision task tend to exhibit greater trust in the AI system (Schaffer et al. 2019), which also influences reliance behavior (Cao et al. 2024). Besides initial, pre-existing familiarity, research has also investigated whether individuals who are initially unfamiliar with the decision task or an AI system can be partially compensated by tutorials and educational interventions. Chiang & Yin (2022), He et al. (2023), and Chiang & Yin (2021) found that providing user tutorials and brief education sessions on AI systems can help individuals to rely on AI systems more appropriately, e.g. by reducing overreliance. Educating users about the AI systems' behavior, such as highlighting consistent output patterns or AI failures, further helps individuals to build better mental models of the AI system, increasing their trust and significantly improving collaborative decision-making (Cabrera et al. 2023). However, the results of Chiang & Yin (2022) suggest that the positive effects of tutorials and educational interventions may only be effective for users who had a high prior ability to solve the decision task by themselves, suggesting an interaction with individuals' expertise. Finally, it is important to recognize that increasing individuals' familiarity with an AI system does not always lead to favorable outcomes. In fact, it can sometimes result in decreased AI reliance, especially in cases where individuals have observed AI errors (Berger et al. 2021, Schaffer et al. 2019, Dietvorst et al. 2015).

Agency: In addition to the factors mentioned above, the degree of human agency - defined as the extent to which individuals are able to make and act upon their own decisions - is another critical factor that influences trust and reliance on AI systems. Kim & Song (2023*a*) concluded that individuals tend to place greater trust in AI systems when they do not perceive their decision-making ownership to be compromised. These findings support Westphal et al. (2023)'s results that trust in AI systems increases when users retain control over their decisions, which subsequently increases both intended and actual compliance with AI advice. However, there is a significant difference in the agency preferences of individuals based on their performance levels; high performers tend to seek more control over AI advice, whereas low performers are more inclined to receive less controllable AI advice (Kawaguchi 2021).

Cognition: Our review highlights that systematic cognitive errors, exemplified by cognitive biases, significantly influence trust and reliance in human-AI decision-making. For example, confirmation bias leads individuals to favor information that aligns with pre-existing beliefs, resulting in selectively biased reliance on AI systems that reinforce stereotypes (Alon-Barkat & Busuioc 2023). Similarly, Selten et al. (2023) found that police officers were more likely to trust AI recommendations when these confirmed their prior assumptions, thereby limiting AI's potential to mitigate negative stereotypes. Moreover, the Dunning-Kruger effect further exacerbates cognitive biases by causing individuals with low competence to overestimate their abilities, leading to underreliance on AI due to overconfidence in their own judgment (Lu & Yin 2021, He et al. 2023). Conversely, automation bias fosters overreliance on AI, as individuals substitute automated outputs for comprehensive information processing (Schemmer et al. 2022). The severity of automation bias is influenced by AI accuracy and task complexity - individuals are more susceptible to errors when AI systems demonstrate consistently high accuracy or when cognitive load is elevated (Lyell & Coiera 2017).
A promising strategy to mitigate cognitive biases in human-AI decision-making is the implementation of cognitive forcing strategies, which interrupt heuristic-driven reasoning and encourage analytical deliberation (Lambe et al. 2016, Wason & Evans 1974). Buçinca et al. (2021), for example, examined cognitive interventions such as a 30-second delay before receiving AI advice, demonstrating that such cognitive forcing functions can reduce overreliance. However, despite these interventions, collaborative human-AI performance remained inferior to that of AI systems alone (Buçinca et al. 2021).

Disposition: The disposition to trust, a personality trait defined as a "consistent tendency to be willing to depend on others across a broad spectrum of situations" (McKnight et al. 1998, 477), is a key factor in trust and reliance on AI in decision-making. Research by He et al. (2023) shows a direct link between dispositional trust in AI and post-collaboration trust, which also significantly affects their willingness to delegate decisions to AI (Shi et al. 2021). Furthermore, Ochmann et al. (2020)'s research further discovered that an individual's trust disposition, conceptualized as general trust in AI, moderates the extent of AI reliance. Apart from individuals' disposition to trust, their dispositional differences in cognitive motivation, exemplified by their need for cognition Cacioppo et al. (1996), is another important factor influencing trust and reliance on AI systems. Individuals with higher need for cognition are more inclined to appreciate AI systems and thus rely more on AI advisors than on human advisors (You et al. 2022). Additionally, Buçinca et al. (2021) found that individuals with a high need for cognition overrelied on AI systems significantly less than individuals with a low need for cognition when using cognitive forcing strategies.

4.4 Trust-Reliance Moderators

Apart from including contextual factors as antecedents of core constructs and dependent variables in our conceptual framework, it is also important to consider contextual factors as moderators of relationships (Hong et al. 2014). Our review identifies three factors moderating the relationship between trust and reliance. First, cognitive load negatively moderates the relationship between trust and reliance on AI systems, with higher cognitive load reducing AI reliance (You et al. 2022). Second, the level of perceived uncertainty in decision-making contexts influences the dynamics of trust and reliance on AI systems, with greater uncertainty leading to more reliance on AI advice (Elson et al. 2021). Third, Shi et al. (2021)'s research highlights that risk perception, particularly social risk, acts as a moderator between trust and reliance, encouraging further research into other types of risk, such as financial risk, to fully understand the impact of risk on the relationship between trust and reliance.

5 Discussion and Conclusion

This study aims to clarify prior mixed findings in the literature on trust and reliance in collaborative human–AI decision-making environments by examining contextual factors that compose these environments. Our analysis identifies decision-related (risk, objectivity, morality, complexity), human-related (expertise, familiarity, agency, cognition, disposition), and AI-related (explainability, performance, adaptability, anthropomorphism, advice presentation) factors that shape trust and reliance on AI. Furthermore, we elucidate the interactive and moderating effects among these factors. By systematically unpacking these contextual factors and their interdependencies, our study contributes to the emerging literature on collaborative human-AI decision-making (e.g., Vasconcelos et al. 2023, Hemmer et al. 2021, Spatscheck et al. 2024), by offering a more nuanced understanding of the antecedents of trust and reliance in such settings.

Importantly, our findings provide an initial explanatory framework for the observed inconsistencies surrounding overtrust and undertrust, as well as overreliance and underreliance, thereby offering a foundation for future empirical and practical inquiry. In line with prior IS theorizing on contextualized system use (Hong et al. 2014), our approach emphasizes that the effects of trust and reliance must be understood as emergent properties of specific constellations of contextual factors. In particular, our framework, enables researchers and practitioners to construct and compare diverse decision-making environments, formulate testable hypotheses, and develop context-sensitive interventions, thereby increasing the applicability of these IS studies to real-world decision-making environments. For example, clinical decision-making environments are typically contextualized by clinicians' high expertise and familiarity with decision tasks, alongside subjective, high-risk, and moral decisions. Our synthesized framework shows that these factors contribute to an environment where individuals generally exhibit lower trust and reliance on AI and are more averse to its use, which provides an important explanatory factor for those findings in previous empirical studies (Longoni et al. 2019, Bhattacherjee & Hikmet 2007, Lapointe & Rivard 2005). In contrast, contexts such as lifestyle, leisure, or abstract problem-solving are characterized by low risk, objective outcomes, and minimal moral or ethical concerns. In these settings, individuals typically have limited domain expertise and low familiarity with both the AI system and the decision task. Our framework suggests that these environments foster increased trust and, consequently, overreliance on AI systems, in line with prior studies reporting higher trust and reliance in such environments (de Jong et al. 2025, Buçinca et al. 2021, Vasconcelos et al. 2023, You et al. 2022). Accordingly, analogous contextual profiles, such as those found in legal, military, financial, or public administration decision-making environments, can be conceptualized and empirically investigated to advance context-specific knowledge.

Our review identifies several critical gaps in the current literature. First, there is a notable underrepresentation of human- and decision-related factors relative to the extensive focus on AI-centric dimensions. While much of the literature emphasizes the technical dimensions of AI, few studies sufficiently address human cognition, behavior, and judgment. This imbalance limits our understanding of how individuals effectively collaborate with AI systems and hinders the development of robust theories of trust and reliance. Second, the literature lacks theoretical integration. Many studies investigate narrowly defined constructs in isolation, often without grounding their work in established theoretical frameworks. As such, there is a clear need for empirically validated, integrative models that capture the interplay of multiple antecedents and contextual factors. To advance theorizing in this area, future research should build upon and extend established theories such as Task-Technology Fit (Goodhue 1995) and the Technology-Organization-Environment framework (Tornatzky et al. 1990). Third, despite the growing ubiquity of AI-supported decision-making in teams and organizational settings (Zercher et al. 2025, Dennis et al. 2023), research adopting a team-level perspective remains limited. In particular, the dynamics of trust and reliance in multi-agent contexts, where multiple human actors interact with one or more AI systems, are still underexplored. Addressing this gap will require greater incorporation of socio-psychological and organizational theories that can capture both individual-level and collective-level processes in human-AI teaming. Finally, our proposed framework offers a conceptual foundation for future research on human collaboration with generative AI systems (e.g., Fang et al. 2025, Spatscheck et al. 2024). As these systems become increasingly embedded in co-creative and knowledge-intensive decisions, there is a need to reassess established antecedents of trust and reliance and, where necessary, develop new factors that account for the interactive nature of generative AI in collaborative decision-making environments.

While our study builds upon and expands on existing literature reviews - particularly Hemmer et al. (2021), which was limited to XAI, and Li & Hahn (2022) and Jussupow et al. (2020), which gave limited attention to interaction effects - several limitations of our study still warrant acknowledgment. First, our focus on individual-level collaborative decision-making may restrict the transferability of our findings to organizational or societal settings, where additional dynamics and contextual factors are present. Second, our framework is derived from a systematic literature review that integrates both inductive and deductive approaches; while this methodological pluralism enhances the robustness and depth of our analysis, it may also introduce bias in the identification and categorization of themes. Moreover, as with all literature reviews, the comprehensiveness of our findings is constrained by the chosen search strings, databases, and inclusion/exclusion criteria.

References

Alon-Barkat, S. & Busuioc, M. (2023), 'Human–AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice', *Journal of Public Administration Research and Theory* **33**(1), 153–169.

Bayer, S., Gimpel, H. & Markgraf, M. (2022), 'The role of domain expertise in trusting and following explainable ai decision support systems', *Journal of Decision Systems* **32**(1), 110–138.

Berger, B., Adam, M., Rühr, A. & Benlian, A. (2021), 'Watch me improve—algorithm aversion and demonstrating the ability to learn', *Business & Information Systems Engineering* **63**(1), 55–68.

Bhattacherjee, A. & Hikmet, N. (2007), 'Physicians' resistance toward healthcare information technology: a theoretical model and empirical test', *European Journal of Information Systems* **16**(6), 725–737.

Bigman, Y. E. & Gray, K. (2018), 'People are averse to machines making moral decisions', *Cognition* **181**, 21–34.

Bo, J. Y., Wan, S. & Anderson, A. (2025), 'To rely or not to rely? evaluating interventions for appropriate reliance on large language models', *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* pp. 1–23.

Buçinca, Z., Malaya, M. B. & Gajos, K. Z. (2021), 'To trust or to think: cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making', *Proceedings of the ACM on Human-Computer Interaction* **5**, 1–21.

Burton, J. W., Stein, M.-K. & Jensen, T. B. (2020), 'A systematic review of algorithm aversion in augmented decision making', *Journal of behavioral decision making* **33**(2), 220–239.

Bussone, A., Stumpf, S. & O'Sullivan, D. (2015), 'The role of explanations on trust and reliance in clinical decision support systems', *2015 international conference on healthcare informatics* pp. 160–169.

Cabrera, Á. A., Perer, A. & Hong, J. I. (2023), 'Improving human-ai collaboration with descriptions of ai behavior', *ACM Human Computer Interaction* .

Cacioppo, J., Petty, R., Feinstein, J. & Jarvis, B. (1996), 'Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition.', *Psychological bulletin* **119**(2), 197.

Cao, S. & Huang, C.-M. (2022), 'Understanding user reliance on ai in assisted decision-making', *Proceedings of the ACM on Human-Computer Interaction* **6**, 1–23.

Cao, S., Liu, A. & Huang, C.-M. (2024), 'Designing for appropriate reliance: The roles of ai uncertainty presentation, initial user decision, and user demographics in ai-assisted decision-making', *Proceedings of the ACM on Human-Computer Interaction* **8**(CSCW1), 1–32.

Castelo, N., Bos, M. W. & Lehmann, D. R. (2019), 'Task-dependent algorithm aversion', *Journal of Marketing Research* **56**(5), 809–825.

Chiang, C.-W. & Yin, M. (2021), 'You'd better stop! understanding human reliance on machine learning models under covariate shift', *Proceedings of the 13th ACM Web Science Conference 2021* pp. 120–129.

Chiang, C.-W. & Yin, M. (2022), ‘Exploring the effects of machine learning literacy interventions on laypeople’s reliance on machine learning models’, *27th International Conference on Intelligent User Interfaces* pp. 148–161.

de Jong, S., Paananen, V., Tag, B. & van Berkel, N. (2025), ‘Cognitive forcing for better decision-making: Reducing overreliance on ai systems through partial explanations’, *Proceedings of the ACM on Human-Computer Interaction* **9**(2), 1–30.

Dellermann, D., Ebel, P., Söllner, M. & Leimeister, J. M. (2019), ‘Hybrid intelligence’, *Business & Information Systems Engineering* **61**, 637–643.

Dennis, A. R., Lakhiwal, A. & Sachdeva, A. (2023), ‘Ai agents as team members: Effects on satisfaction, conflict, trustworthiness, and willingness to work with’, *Journal of Management Information Systems* **40**(2), 307–337.

Dietvorst, B. & Bharti, S. (2020), ‘People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error’, *Psychological science* **31**(10), 1302–1314.

Dietvorst, B., Simmons, J. & Massey, C. (2015), ‘Algorithm aversion: People erroneously avoid algorithms after seeing them err’, *Journal of Experimental Psychology: General* **144**(1), 114–126.

Efendić, E., Van de Calseyde, P. P. & Evans, A. M. (2020), ‘Slow response times undermine trust in algorithmic (but not human) predictions’, *Organizational Behavior and Human Decision Processes* **157**, 103–114.

Elson, J. S., Derrick, D. C. & Merino, L. A. (2021), ‘An empirical study exploring difference in trust of perceived human and intelligent system partners’, *Proceedings of the Annual Hawaii International Conference on System Sciences* pp. 136–145.

Fang, C., Zhu, Y., Fang, L., Long, Y., Lin, H., Cong, Y. & Wang, S. J. (2025), ‘Generative ai-enhanced human-ai collaborative conceptual design: A systematic literature review’, *Design Studies* **97**, 101300.

Fishbein, M. & Ajzen, I. (1977), ‘Belief, attitude, intention, and behavior: An introduction to theory and research’, *Contemporary Sociology* .

Glikson, E. & Woolley, A. W. (2020), ‘Human trust in artificial intelligence: Review of empirical research’, *Academy of Management Annals* **14**(2), 627–660.

Goel, K., Sindhgatta, R., Kalra, S., Goel, R. & Mutreja, P. (2022), ‘The effect of machine learning explanations on user trust for automated diagnosis of covid-19’, *Computers in Biology and Medicine* **146**, 105587.

Gogoll, J. & Uhl, M. (2018), ‘Rage against the machine: Automation in the moral domain’, *Journal of Behavioral and Experimental Economics* **74**, 97–103.

Gomez, C., Unberath, M. & Huang, C.-M. (2023), ‘Mitigating knowledge imbalance in ai-advised decision-making through collaborative user involvement’, *International Journal of Human-Computer Studies* **172**, 102977.

Goodhue, D. L. (1995), ‘Understanding user evaluations of information systems’, *Management science* **41**(12), 1827–1844.

Goodhue, D. L. & Thompson, R. L. (1995), ‘Task-technology fit and individual performance’, *MIS quarterly* pp. 213–236.

Harper, R. H. (2019), ‘The role of hci in the age of ai’, *International Journal of Human–Computer Interaction* **35**(15), 1331–1344.

He, G., Kuiper, L. & Gadiraju, U. (2023), ‘Knowing about knowing: An illusion of human competence can hinder appropriate reliance on ai systems’, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* pp. 1–18.

Hemmer, P., Schemmer, M., Kühl, N., Vössing, M. & Satzger, G. (2024), 'Complementarity in human-ai collaboration: Concept, sources, and evidence', *arXiv preprint arXiv:2404.00029* .

Hemmer, P., Schemmer, M., Vössing, M. & Kühl, N. (2021), 'Human-ai complementarity in hybrid intelligence systems: A structured literature review.', *Pacific Asia Conference on Information Systems 2021* .

Hoff, K. A. & Bashir, M. (2015), 'Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust', *Human Factors* **57**(3), 407–434.

Hong, W., Chan, F. K., Thong, J. Y., Chasalow, L. C. & Dhillon, G. (2014), 'A framework and guidelines for context-specific theorizing in information systems research', *Information systems research* **25**(1), 111–136.

Inkpen, K., Chappidi, S., Mallari, K., Nushi, B., Ramesh, D., Michelucci, P., Mandava, V., VepŘek, L. H. & Quinn, G. (2023), 'Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making', *ACM Transactions on Computer-Human Interaction* .

Jabareen, Y. (2009), 'Building a conceptual framework: philosophy, definitions, and procedure', *International journal of qualitative methods* **8**(4), 49–62.

Juravle, G., Boudouraki, A., Terziyska, M. & Rezlescu, C. (2020), 'Trust in artificial intelligence for medical diagnoses', *Progress in brain research* **253**, 263–282.

Jussupow, E., Benbasat, I. & Heinzl, A. (2020), 'Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion', *European Conference on Information Systems 2020* .

Jussupow, E., Benbasat, I. & Heinzl, A. (2024), 'An integrative perspective on algorithm aversion and appreciation in decision-making.', *MIS Quarterly* **48**(4).

Jussupow, E., Spohrer, K., Heinzl, A. & Gawlitza, J. (2021), 'Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence', *Information Systems Research* **32**(3), 713–735.

Kahr, P. K., Rooks, G., Willemsen, M. C. & Snijders, C. C. (2023), 'It seems smart, but it acts stupid: Development of trust in ai advice in a repeated legal decision-making task', *Proceedings of the 28th International Conference on Intelligent User Interfaces* pp. 528–539.

Kaufmann, E. (2021), 'Algorithm appreciation or aversion? comparing in-service and pre-service teachers' acceptance of computerized expert models', *Computers and Education: Artificial Intelligence* **2**, 100028.

Kawaguchi, K. (2021), 'When will workers follow an algorithm? a field experiment with a retail business', *Management Science* **67**(3), 1670–1695.

Kim, S. S., Liao, Q. V., Vorvoreanu, M., Ballard, S. & Vaughan, J. W. (2024), '" i'm not sure, but...": Examining the impact of large language models' uncertainty expression on user reliance and trust', *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* pp. 822–835.

Kim, S. S., Vaughan, J. W., Liao, Q. V., Lombrozo, T. & Russakovsky, O. (2025), 'Fostering appropriate reliance on large language models: The role of explanations, sources, and inconsistencies', *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* pp. 1–19.

Kim, T. & Song, H. (2023*a*), 'Communicating the limitations of ai: the effect of message framing and ownership on trust in artificial intelligence', *International Journal of Human–Computer Interaction* **39**(4), 790–800.

Kim, T. & Song, H. (2023*b*), '"i believe ai can learn from the error. or can it not?": The effects of implicit theories on trust repair of the intelligent agent', *International Journal of Social Robotics* **15**(1), 115–128.

Kunreuther, H., Meyer, R., Zeckhauser, R., Slovic, P., Schwartz, B., Schade, C., Luce, M. F., Lippman, S., Krantz, D., Kahn, B. et al. (2002), 'High stakes decision making: Normative, descriptive and prescriptive considerations', *Marketing Letters* **13**, 259–268.

Lacroux, A. & Martin-Lacroux, C. (2022), 'Should i trust the artificial intelligence to recruit? recruiters' perceptions and behavior when faced with algorithm-based recommendation systems during resume screening', *Frontiers in Psychology* **13**, 895997.

Lai, V., Chen, C., Smith-Renner, A., Liao, Q. V. & Tan, C. (2023), 'Towards a science of human-ai decision making: An overview of design space in empirical human-subject studies', *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency* pp. 1369–1385.

Lai, V. & Tan, C. (2019), 'On human predictions with explanations and predictions of machine learning models: A case study on deception detection', *Proceedings of the conference on fairness, accountability, and transparency* pp. 29–38.

Lambe, K. A., O'Reilly, G., Kelly, B. D. & Curristan, S. (2016), 'Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review', *BMJ quality & safety* **25**(10), 808–820.

Lapointe, L. & Rivard, S. (2005), 'A multilevel model of resistance to information technology implementation', *MIS quarterly* pp. 461–491.

Lee, J. & See, K. (2004), 'Trust in automation: Designing for appropriate reliance', *Human factors* **46**(1), 50–80.

Lee, M. (2018), 'Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management', *Big Data & Society* **5**(1).

Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M. & Mara, M. (2023), 'Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task', *Computers in Human Behavior* **139**, 107539.

Li, Y. & Hahn, J. (2022), 'Review of Research on Human Trust in Artificial Intelligence', *International Conference on Information Systems 2022* pp. 0–17.

Lohoff, L. & Rühr, A. (2021), 'Introducing (Machine) Learning Ability as Antecedent of Trust in Intelligent Systems', *European Conference on Information Systems 2021* pp. 1–16.

Longoni, C., Bonezzi, A. & Morewedge, C. K. (2019), 'Resistance to medical artificial intelligence', *Journal of Consumer Research* **46**(4), 629–650.

Lu, Z. & Yin, M. (2021), 'Human reliance on machine learning models when performance feedback is limited: Heuristics and risks', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Lyell, D. & Coiera, E. (2017), 'Automation bias and verification complexity: a systematic review', *Journal of the American Medical Informatics Association* **24**(2), 423–431.

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S. & Söllner, M. (2019), 'Ai-based digital assistants: Opportunities, threats, and research perspectives', *Business & Information Systems Engineering* **61**, 535–544.

McKnight, D. H., Cummings, L. L. & Chervany, N. L. (1998), 'Initial trust formation in new organizational relationships', *Academy of Management review* **23**(3), 473–490.

McKnight, D. H., Liu, P. & Pentland, B. T. (2020), 'Trust change in information technology products', *Journal of Management Information Systems* **37**(4), 1015–1046.

Mehrotra, S., Jorge, C. C., Jonker, C. M. & Tielman, M. L. (2023), 'Integrity based explanations for fostering appropriate trust in ai agents', *ACM Transactions on Interactive Intelligent Systems* .

Montealegre-López, N. (2025), 'Exploring the role of trust in ai-driven decision-making: a systematic literature review', *Management Review Quarterly* pp. 1–51.

Morana, S., Gnewuch, U., Jung, D. & Granig, C. (2020), 'The effect of anthropomorphism on investment decision-making with robo-advisor chatbots.', *European Conference on Information Systems 2020* .

Newman, L., Haran, U. & Fink, L. (2022), 'Let me decide: The importance of user autonomy in accepting online recommendations.', *European Conference on Information Systems 2022* .

Ochmann, J., Michels, L., Zilker, S., Tiefenbeck, V. & Laumer, S. (2020), 'The influence of algorithm aversion and anthropomorphic agent design on the acceptance of ai-based job recommendations.', *International Conference on Information Systems 2020* .

Papenmeier, A., Kern, D., Englebienne, G. & Seifert, C. (2022), 'It's complicated: The relationship between user trust, model accuracy and explanations in ai', *ACM Transactions on Computer-Human Interaction (TOCHI)* **29**(4), 1–33.

Parasuraman, R. & Riley, V. (1997), 'Humans and automation: Use, misuse, disuse, abuse', *Human factors* **39**(2), 230–253.

Park, G., Chung, J. & Lee, S. (2023), 'Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention', *Current Psychology* pp. 1–12.

Parkes, A. (2017), 'The effect of individual and task characteristics on decision aid reliance', *Behaviour & Information Technology* **36**(2), 165–177.

Qiu, L. & Benbasat, I. (2009), 'Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems', *Journal of management information systems* **25**(4), 145–182.

Raisch, S. & Krakowski, S. (2021), 'Artificial intelligence and management: The automation–augmentation paradox', *Academy of management review* **46**(1), 192–210.

Rubin, E. & Benbasat, I. (2023), 'Using toulmin's argumentation model to enhance trust in analytics-based advice giving systems', *ACM Transactions on Management Information Systems* **14**(3), 1–28.

Schaffer, J., O'Donovan, J., Michaelis, J., Raglin, A. & Höllerer, T. (2019), 'I can do better than your ai: expertise and explanations', *Proceedings of the 24th International Conference on Intelligent User Interfaces* pp. 240–251.

Schemmer, M., Kuehl, N., Benz, C., Bartos, A. & Satzger, G. (2023), 'Appropriate reliance on ai advice: Conceptualization and the effect of explanations', *Proceedings of the 28th International Conference on Intelligent User Interfaces* pp. 410–422.

Schemmer, M., Kühl, N., Benz, C. & Satzger, G. (2022), 'On the influence of explainable ai on automation bias', *arXiv preprint arXiv:2204.08859* .

Schmitt, A., Wambsganss, T., Söllner, M. & Janson, A. (2021), 'Towards a trust reliance paradox? exploring the gap between perceived trust in and reliance on algorithmic advice', *International Conference on Information Systems 2021* **1**, 1–17.

Schreuter, D., van der Putten, P. & Lamers, M. H. (2021), 'Trust me on this one: conforming to conversational assistants', *Minds and Machines* **31**, 535–562.

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G. J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G. & Söllner, M. (2020), 'Machines as teammates: A research agenda on AI in team collaboration', *Information & Management* **57**(2).

Seeger, A.-M., Pfeiffer, J. & Heinzl, A. (2021), 'Texting with humanlike conversational agents: Designing for anthropomorphism', *Journal of the Association for Information Systems* **22**(4), 8.

Selten, F., Robeer, M. & Grimmelikhuijsen, S. (2023), "just like i thought': Street-level bureaucrats trust ai recommendations if they confirm their professional judgment', *Public Administration Review* **83**(2), 263–278.

Shi, S., Gong, Y. & Gursoy, D. (2021), 'Antecedents of trust and adoption intention toward artificially intelligent recommendation systems in travel planning: a heuristic–systematic model', *Journal of Travel Research* **60**(8), 1714–1734.

Spatscheck, N., Schaschek, M. & Winkelmann, A. (2024), 'The effects of generative ai's human-like competencies on clinical decision-making', *Journal of Decision Systems* pp. 1–39.

Tornatzky, L. G., Fleischer, M. & Chakrabarti, A. K. (1990), *The processes of technological innovation*, Lexington Books.

Ueno, T., Sawa, Y., Kim, Y., Urakami, J., Oura, H. & Seaborn, K. (2022), 'Trust in human-ai interaction: Scoping out models, measures, and methods', *CHI Conference on Human Factors in Computing Systems Extended Abstracts* pp. 1–7.

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S. & Krishna, R. (2023), 'Explanations can reduce overreliance on ai systems during decision-making', *Proceedings of the ACM on Human-Computer Interaction* **7**, 1–38.

Venkatesh, V. (2025), 'Leveraging context: Re-thinking research processes to make "contributions to theory"', *Information Systems Research* pp. 230–253.

Vössing, M., Kühl, N., Lind, M. & Satzger, G. (2022), 'Designing transparency for effective human-ai collaboration', *Information Systems Frontiers* **24**(3), 877–895.

Wang, X. & Yin, M. (2021), 'Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making', *26th international conference on intelligent user interfaces* pp. 318–328.

Wason, P. C. & Evans, J. S. B. (1974), 'Dual processes in reasoning?', *Cognition* **3**(2), 141–154.

Webster, J. & Watson, R. T. (2002), 'Analyzing the past to prepare for the future: Writing a literature review', *MIS quarterly* pp. xiii–xxiii.

Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. B. & Rafaeli, A. (2023), 'Decision control and explanations in human-ai collaboration: Improving user perceptions and compliance', *Computers in Human Behavior* **144**, 107714.

Wysocki, O., Davies, J. K., Vigo, M., Armstrong, A. C., Landers, D., Lee, R. & Freitas, A. (2023), 'Assessing the communication gap between ai models and healthcare professionals: Explainability, utility and trust in ai-driven clinical decision-making', *Artificial Intelligence* **316**, 103839.

Xu, D. & Xu, D. J. (2021), 'AI for depression treatment: Addressing the paradox of privacy and trust with empathy, accountability, and explainability', *Internation Conference on Information Systems 2021* .

Yin, M., Vaughan, J. W. & Wallach, H. (2019), ‘Understanding the effect of accuracy on trust in machine learning models’, *Conference on Human Factors in Computing Systems - Proceedings* .

You, S., Yang, C. L. & Li, X. (2022), ‘Algorithmic versus Human Advice: Does Presenting Prediction Performance Matter for Algorithm Appreciation?’, *Journal of Management Information Systems* **39**(2), 336–365.

Zercher, D., Jussupow, E. & Heinzl, A. (2025), ‘Team climate in team-ai collaboration: Exploring the role of decisional ownership and perceived ai team membership’, *European Conference on Information Systems* .

Zhang, Q., Lee, M. L. & Carter, S. (2022), ‘You complete me: Human-ai teams and complementary expertise’, *Proceedings of the 2022 CHI conference on human factors in computing systems* pp. 1–28.

Zhang, Z. T., Buchner, F., Liu, Y. & Butz, A. (2024), ‘You can only verify when you know the answer: Feature-based explanations reduce overreliance on ai for easy decisions, but not for hard ones’, *Proceedings of Mensch und Computer 2024* pp. 156–170.