August 2024

Trust in Digital Human-AI Team Collaboration: A Systematic Review

Essi Janhunen
LUT University, essi.janhunen@lut.fi

Tuuli Toivikko
LUT University, tuuli.toivikko@lut.fi

Kirsimarja Blomqvist
Lappeenranta University of Technology, kirsimarja.blomqvist@lut.fi

Dominik Siemon
LUT University, dominik.siemon@lut.fi

Trust in Digital Human-AI Team Collaboration: A Systematic Review

Completed Research Full Paper

Essi Janhunen
LUT University
essi.janhunen@lut.fi

Tuuli Toivikko
LUT University
tuuli.toivikko@lut.fi

Kirsimarja Blomqvist
LUT University
kirsimarja.blomqvist@lut.fi

Dominik Siemon
LUT University
dominik.siemon@lut.fi

Abstract

Artificial intelligence (AI) is increasingly used to support team collaboration on digital platforms. Trust has been identified as critical for team collaboration yet adopting AI changes team dynamics and challenges team trust. Based on a systematic literature review this study shows how different digital human-AI team role and task configurations affect team processes and related trust dynamics between AI and human collaborators. It also identifies cross-level effects as team-level trust is dynamically linked to both individual perceptions of AI trustworthiness and organizational level trust and psychological safety in AI adoption. This study contributes to research on trust in digital human-AI-team collaboration by identifying relevant literature streams and providing a conceptual framework for the multi-referent and multi-level trust in digital human-AI team collaboration. The paper concludes with implications for research and practice as well as future research ideas stemming from the review.

Keywords

Artificial intelligence, trust, digital team collaboration, systematic literature review.

Introduction

AI is shaping collaboration (Jarrahi, 2018), especially when teamwork is increasingly organized in AI-supported digital collaboration platforms (Razmerita, Brun and Nabeth, 2022). The integration of AI into teams creates new collaborative structures between humans and AI (Anthony, Bechky and Fayard, 2023), and AI is increasingly perceived as more of a collaborator than just a tool (Seeber et al., 2020). AI affects how people collaborate, coordinate, communicate, and manage work (Razmerita et al., 2022), requiring a focus on the socio-technical nature of human-AI configurations (Makarius et al., 2020). Existing literature has emphasized more individuals than teams in human-AI collaboration (Zercher, Jussupow and Heintzl, 2023), although knowledge work is increasingly organized in teams (Siemon and Strohmann, 2021). Therefore, research on how AI reshapes team collaboration and its implications for trust is needed (Benbya, Pachidi and Jarvenpaa, 2021; Gkinko and Elbanna, 2023).

Trust is critical for team performance (Costa, Fulmer, and Anderson, 2018), and for technology adoption (Söllner, Hoffman and Leimeister, 2016; Gkinko and Elbanna, 2023). Moreover, the shift to digital collaboration technologies itself has been identified to affect team trust (Razmerita et al., 2022; Robert, Denis and Hung 2009). Despite AI supporting teams, if AI is used in a controlling way (Kellogg, Valentine, and Christin, 2020), it can deteriorate team trust, psychological safety, and well-being (Fulmer and Gelfand, 2009). The multi-referent (or target) and multi-level nature of trust create a need to better understand the relationship between evolving AI and trust in digital team collaboration.

Past reviews on AI and trust show that research has been focused on human trust towards AI and AI trustworthiness (Glikson and Woolley, 2020; Kaur et al., 2022). Researchers interested in AI are moving towards team-level human-AI configurations (Rix, 2022; Zercher et al., 2023), making trust dynamics more complex. Such human-AI team research is focused on co-located teams, although teamwork is increasingly digitally mediated. To understand human-AI team configurations as socio-technical collaboration systems, an interdisciplinary approach is needed (Østerlund et al., 2021). Therefore, this systematic literature review (SLR) analyzes research from the domains of Information Systems (IS) and Organization and Management to answer the question of *How does AI affect team trust in digital human-AI team collaboration?* This study contributes by identifying different literature streams and synthesizing the findings as a conceptual framework.

Conceptual Background

AI in Digital Team Collaboration

AI is a group of technologies enabling pattern recognition, reasoning, sensing, and learning (Jarrahi, 2018; Russell and Norvig, 2021) by applying for example neural networks, machine learning, natural language processing, and computer vision. AI capabilities and appearance vary across applications. In a digital context, AI does not have a physical form. Instead, it is virtual or embedded with varying anthropomorphic features, i.e., human-like appearance and behavior (Glikson and Woolley, 2020). AI capabilities and autonomy vary from low-level assistance and human control to autonomous AI acting independently (Shneiderman, 2020). Between these are automating AI and augmenting AI. Its capabilities, autonomy, and anthropomorphic features are determined by the specific AI application and its purpose in the team, e.g., conversational agents or interactive dashboards.

Digital collaboration is defined as two or more entities working for a common goal via digital platforms sharing their knowledge, independent of location and time (Faraj, Pachidi and Sayegh, 2018; Razmerita et al., 2022). Traditionally at least two individuals, now AI can replace or complement humans, changing team collaboration (Razmerita et al., 2022). Algorithms can imitate a knowledge worker in acquiring knowledge (Faraj et al., 2018), and help experts to leverage knowledge rather than replace them (Pettersen, 2019). AI is changing from a tool to a machine as a teammate (Seeber et al., 2020; Rix, 2022), and when human and machine intelligence are used collectively (hybrid intelligence), AI will change team processes and organization including communication, creativity, facilitation, and management (Khakurel and Blomqvist, 2022; Zercher et al., 2023). For example, AI can support teamwork as a teammate (Seeber et al., 2020), a communication mediator (Anthony et al., 2023), or a team leader (Rix, 2022). Despite the envisioned advantages for team processes, loss of human agency and responsibility issues can lead to decreased trust (Shneiderman, 2020; Siemon and Strohmann, 2021).

Trust in Digital Team Collaboration

Trust is the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other party (Mayer, Davis and Schoorman, 1995, 712). Organizational studies examine social trust at the individual, team, and organizational levels (Fulmer and Gelfand, 2012), while IS research has mainly focused on human trust in technology (Söllner et al., 2016). Despite the different approaches, definitions of trust consistently in both domains include positive expectations and vulnerability (Rousseau et al., 1998; Söllner et al., 2016). Human-AI trust considers AI technical properties affecting the trustworthiness of AI (Kaur et al., 2022), including reliability, functionality, and helpfulness (Lankton, Harrison Mcknight and Tripp, 2015). This corresponds closely to human-like trust dimensions of ability, integrity, and benevolence in interpersonal relationships (Mayer et al., 1995), and is also applicable to trust in AI (van der Werff, Blomqvist and Koskinen 2021).

Human-like trust constructs may become important in evaluating AI trustworthiness when it has human-like features (Lankton et al., 2015). Technical characteristics affect cognitive trust, while anthropomorphism counts toward emotional and affect-based trust (Glikson and Woolley, 2020). In human-AI teams, trust is required towards humans and AI, increasing the complexity of who or what is trusted (Zercher et al., 2023). Therefore, this review focuses on digital team collaboration considering both human-like trust in the management literature and human-AI trust in the IS literature (Benbya et

al., 2021). Digital context further challenges trust compared to face-to-face due to digital materiality and dispersion (Robert et al., 2009), affecting the efficiency of team processes and trust.

Methodology

A SLR was conducted to explore trust dynamics in digitally collaborating human-AI teams. Planning, implementation, and reporting followed IS field standards (vom Brocke et al., 2015) and Kitchenham and Charters' (2007) framework. Based on the initial exploration, terms related to AI, trust, team collaboration, and organizing activities were defined. The search string used was: (AI OR "artificial intelligence" OR "hybrid intelligence" OR "human-machine") AND trust AND (collaborat* OR team* OR group OR community) AND (organis* OR coordinat* OR cooperat* OR communicat* OR interact* OR manag* OR work* OR facilitat* OR innovat*). Scopus database which indexes various research fields was chosen. Also, Web of Science was considered, but not utilized due to highly duplicated results. The AIS Senior Scholars list and suitable FT50 management journals (e.g., Academy of Management Journal, Academy of Management Review, Journal of Applied Psychology) were also screened. The search terms were required to appear in the title, abstract, or keywords. English-language journals and conferences between 2013 and 2023 were considered. The search yielded a total 1234 of papers.

Inclusion criteria required the paper to address 1) AI technology, 2) trust, 3) digital team collaboration, 4) team processes and organizing, and 5) collaborative unit always two or more (human team vs human-AI team). A paper was excluded if 1) AI technology was not discussed, 2) trust was not examined, 3) the team was physically co-located, 4) team processes were not discussed, or 5) less than two entities were considered as unit (individual worker vs human-AI team). Due to the novelty of the topic, both empirical and conceptual papers were included. Initially, titles and abstracts were screened, followed by a criteria-based assessment of relevant records, resulting in 14 papers. A backward and forward search then added a further eight papers, giving a total of 22 papers (Figure 1). This final set was collectively approved by the authors. The quality of the sources was ensured by selecting studies from reputable, peer-reviewed articles and reputable publishers. Data on AI applications, trust focus, type of collaboration, and team organizing processes were documented on a data extraction sheet.

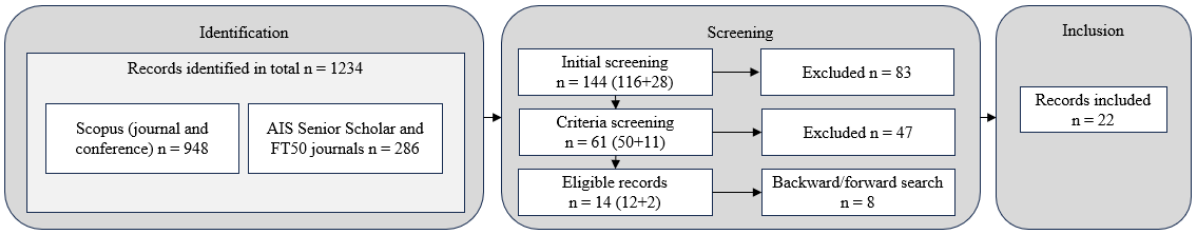

Figure 1. Inclusion Process

Findings

Findings indicate recent and rapidly increasing interest in trust in digital human-AI teams. Five distinct research streams were identified based on publishing outlets, keywords, co-authoring, and citations. **Stream 1** combines research on IS with management and organization research, recognizing AI teammates for collaboration activities, like co-creation and innovation. Trust towards various referents, i.e., AI, team, and organization is discussed. Based on a high number of citations, two subgroups were identified. **Stream 1.1.** from IS management focuses on AI conversational agents that support team social processes and build team trust. **Stream 1.2.** in the technology management field focuses on human-AI collaboration performance, and trust in AI teammates and organization. **Stream 2** in the computer-mediated communication field focuses on AI-mediated communication and sees AI more as a tool. Here interpersonal and AI trust is considered, while organizational trust and psychological safety are important as well. **Stream 3** on human-computer interaction and computer-supported collaborative work literature considers AI as an autonomous, humanlike teammate and focuses primarily on trust towards AI and human-AI teaming processes. Streams are interconnected through citations, Stream 1 being the most influential. All research streams indicate trust as a multi-referent and multi-level concept

in digital human-AI teams. A table with more detailed information about each stream is available in the scientific data sharing application Figshare: 10.6084/m9.figshare.25324939.

Human Collaborator's Perception of AI Trustworthiness

Technological Factors

AI trustworthiness concerns technological design and beliefs about AI ability, integrity, and benevolence (Benke, 2020; Dennis, Lakhiwal and Sachdeva, 2023). AI ability is crucial for trust (Fleischmann, Cardon and Aritz, 2021; Chowdhury et al., 2022). Performance is expected to be even better than humans (Schelble et al., 2022) and it can have a stronger effect on trust than teammate identity (G. Zhang et al., 2023). However, algorithms are a black box, leading to decreased trust due to a lack of transparency (R. Zhang et al., 2021; Chowdhury et al., 2022). Explainable AI (R. Zhang et al., 2021), involving members (Chhibber, Goh and Law, 2022) and increasing understanding of AI can reduce negative attitudes toward AI (Jain, Garg and Khera, 2022; Bezrukova et al., 2023). Also, ethical issues are of concern especially if personal and collaboration data is involved (Hancock, Naaman, Levy, 2020; Cardon, Ma and Fleischmann 2021; Shin et al., 2023).

Anthropomorphic design is relevant for AI trustworthiness (Hauptman, Duan, and McNeese, 2022; G. Zhang et al., 2023), including communication (Correia et al., 2021; Schelble et al., 2022), politeness (Flathmann et al., 2023), and emotional intelligence (Getchell et al., 2022; Hofeditz et al., 2022). Human-like appearance creates a sense of presence (Hofeditz et al., 2022; Flathmann et al., 2023) and a "baseline trust" comparable to human teammates showing up for work (Hauptman et al., 2022). Because AI lacks real human emotions (Fleischmann et al., 2021; Getchell et al., 2022) its ability for relationships is limited (Dennis, et al., 2023), which can decrease trust (Flathmann et al., 2023). Instead of being able to provide relational trust its technical performance is more important (Flathmann et al., 2023). Furthermore, concerns are raised about disguising AI as human (Seeber et al., 2020) as the lack of physical presence can lead to potential manipulation of trust in digital team collaboration (Hancock et al., 2020; Hohenstein and Jung, 2020).

Human Factors

As a response to AI technical design, trust formation towards AI is dependent on individual perceptions of AI (Fleischmann et al., 2021; Chhibber et al., 2022; Hauptman et al., 2022), such as attitudes, beliefs, and expectations (R. Zhang et al., 2021; Chowdhury et al., 2022; Bezrukova et al., 2023; Flathmann et al., 2023). While cognition-based trust is the user's perceptions of the AI's characteristics and trustworthiness, affect-based trust is the user's emotional responses to the system, based on social interaction (Bao et al., 2021; Chhibber et al., 2022). Without previous AI experience affecting trust (R. Zhang et al., 2021), this stage is mainly about perceived interaction complexity compared to interaction comfort (Bao et al., 2021). In this stage trust can be easily violated if AI does not perform as expected (G. Zhang et al., 2023). The level of human expertise influences how AI is trusted (Correia et al., 2021); for workers with more expertise, AI capabilities are relatively more important than whether a teammate is human or AI (G. Zhang et al., 2023). Jain et al. (2022) found that higher decision-making position increases algorithmic aversion, which can indicate that people trust their own capabilities more than AI. Accordingly, AI can also be seen as a threat if a person feels incapable (Flathmann et al., 2023). Sometimes, a lack of understanding leads to ignorance of AI (Schelble et al., 2022). Also, demographics, such as age and language predict AI aversion and trust (Cardon et al., 2021; Fleischmann et al., 2021). Training diverse employees on AI is important since AI skills have a positive effect on human-AI trust (Chowdhury et al., 2022), and therefore, can promote the whole team's trust towards AI.

Trust in AI Adoption

AI serves both the team and organization (Dennis et al., 2023). AI acceptance and adoption are heavily dependent on trust (Chhibber et al., 2022; Hofeditz et al., 2022) and perceived psychological safety (Cardon et al., 2021; Fleischmann et al., 2021). Psychological safety reduces technology anxiety and motivates teams to use AI (Fleischmann et al., 2021). This becomes important when AI becomes an equal teammate (Seeber et al., 2020; Hauptman et al., 2022; Schelble et al., 2022). Command order (Jain et al.,

2022), role clarity, task allocation (Chhibber et al., 2022) as well as AI task fit (Bao et al., 2021) are important for trust in AI. Jain et al. (2022) suggest that AI aversion could instead be human-AI work configuration aversion if humans cannot express their expertise. Also, whether the use of AI is mandatory or voluntary can affect psychological safety, perceived autonomy, and control in the team (Bezrukova et al., 2023). Still, AI can make workers feel vulnerable (Hancock et al., 2020), since managers can use AI for surveillance, authority, control, and evaluation (Benke, 2020; Correia et al., 2021; Dennis et al., 2023). However, AI can also enhance human relationships, e.g., AI-based worker evaluation apprehension can be lower than with a human, thus increasing psychological safety and trust (Gozzo, Woldendorp and de Rooij, 2022). To maintain trust, for example, social contracts (Cardon et al., 2021) could be used to manage data and privacy appropriately (Fleischmann et al., 2021).

Trust in Collaboration with AI

Team Composition and Dynamics

Increasingly, scholars focus on trust in AI as a teammate (Hauptman et al., 2022; Schelble et al., 2022; Dennis et al., 2023). The changing role of AI makes team trust dynamics more complex (Chhibber et al., 2022). Comparisons of human-AI and human-only dyads reveal how AI teammates compare to human collaborators (Gozzo et al., 2022; Jain et al., 2022). For example, humans self-reported higher trust when they were told that they were working with a human, but their behavioral trust indicated higher trust for AI (G. Zhang et al., 2023). Most of the teams include two humans and one AI (for example, Bao et al., 2021; Hofeditz et al., 2022; Flathmann et al., 2023; Shin et al., 2023). A team composition of a human and several AIs was not found beneficial for trust (Schelble et al., 2022). In collectives such as crowdwork, trust-related issues reach the whole community even if AI is supporting individuals (Correia et al., 2021; Chhibber et al., 2022). This shows how not only the individual's trust towards AI but also the social trust is affected (Hancock et al., 2020; Bao et al., 2021). Thus, if AI can give cues that human collaborators trust it, it can build further interpersonal trust (Hofeditz et al., 2022). Vice versa, a loss of AI trust can lead to a loss of interpersonal trust in the human-AI team (Schelble et al., 2022). Sometimes it becomes unclear whether AI is just acting on behalf of the participant or being a separate entity. If the presence of AI is not evident for all collaborators, it decreases transparency and can alter trust between humans (Hohenstein and Jung, 2020).

AI Role, Autonomy and Agency

AI's role in a team varies from a tool to a teammate and from assisting to fully autonomous (Chowdhury et al., 2022; Bezrukova et al., 2023). Team trust dynamics change when AI takes task-related roles (Jain et al., 2022; G. Zhang et al., 2023) and adopts social roles (Dennis et al., 2023), such as a facilitator (Bao et al., 2021; Shin et al., 2023) and emotional support (Benke, 2020; Hofeditz et al., 2022) or communication mediator (Hancock et al., 2020; Hohenstein and Jung, 2020; Fleischmann et al., 2021). Such social roles have been used to support whole team trust (Hofeditz et al., 2022; Shin et al., 2023).

In addition to assisting (Chhibber et al., 2022), automating (Fleischmann et al., 2021), and augmenting humans (Hohenstein and Jung, 2020; Cardon et al., 2021), AI tools can also have autonomy (Hancock et al, 2020). Usually, AI teammates possess autonomy (Seeber et al., 2020; Hauptman et al., 2022; Schelble et al., 2022; Flathmann et al., 2023), which is argued to differentiate AI tools from teammates (Bezrukova et al., 2023). Yet AI teammates can also automate (Bao et al., 2021; Jain et al., 2022; Hofeditz et al., 2022) or assist (Bao et al., 2021; Correia et al., 2021; R. Zhang et al., 2021) to a varying degree. Leaving autonomy to humans is considered better for trust and well-being (Benke, 2020) even if Schelble et al. (2022) found that some people trusted their autonomous AI teammates more than other humans including themselves.

When AI shifts from a passive tool to an autonomous teammate, it creates a perception of collective agency between humans and AI (Dennis et al., 2023). Agency is an attribution given for AI based on the perception of its independent behavior (Hohenstein and Jung, 2020; Gozzo et al., 2022; Schelble et al., 2022). Its social agency is however limited as humans do not usually attribute AI social abilities (Gozzo et al., 2022). In AI-mediated communication, the agency generally stays with the communicator (Hancock et al., 2020). However, when AI fails in its role as a mediator, AI is associated with agency and responsibility for the unsuccessful communication that could decrease trust between human collaborators (Hohenstein

and Jung, 2020). The fear of humans losing agency (Getchell et al., 2022) becomes especially pertinent if AI does not follow human norms (Benke et al., 2020), requiring human agency and control mechanisms for trust in human-AI collaboration (Benke et al., 2020; Shin et al., 2023).

Team Processes and Trust

Trust is relational and forms in social exchange, and therefore if AI can support communication, it can increase trust in both AI and humans (Hohenstein and Jung. 2020), Trust starts often lower and develops slower in digital team collaboration (Hohenstein and Jung, 2020; Chowdhury et al., 2022; Hofeditz et al., 2022). AI-facilitated familiarization process can also enhance team trust by lowering social distance (Shin et al., 2023). Also, how well AI fulfills the humans' expectations (Hofeditz et al., 2022) is critical for human-AI team formation and engagement (R. Zhang et al., 2021; Hauptman et al., 2022; Bezrukova et al., 2023).

AI modifies team communication and is therefore relevant for trust (Bao et al., 2021; R. Zhang et al., 2021; Getchell et al., 2022; G. Zhang et al., 2023). Multi-language support (Fleischmann et al., 2021), AI interaction etiquette (Flathmann et al., 2023) and neutral communication style (Hancock et al., 2020; Cardon et al., 2021) can increase trust in other humans. Moreover, AI-supported communication can promote team mental models, shared understanding, and cognition which are important for trust (R. Zhang et al., 2021; Schelble et al., 2022). AI can also be used to promote team interaction (Hofeditz et al., 2022; Shin et al., 2023) for higher social presence and team cohesion (Hauptman et al., 2022).

AI can also have a role in team decision-making (Correia et al., 2021; R. Zhang et al., 2021; Jain et al., 2022), problem-solving (Seeber et al., 2020), and innovation processes (Correia et al., 2021). According to Jain et al. (2022), decision-making configurations in roles and command order between humans and AI can lead to higher trust towards AI, except when AI takes the lead and humans are not given a chance to apply their expertise. Trust also influences whether a human is willing to delegate tasks to AI (Chhibber et al., 2022), which again affects perceived workload (G. Zhang et al., 2023). Correia et al. (2021) found that AI is trusted and accepted for supporting innovative work, although AI could not be as creative as a human collaborator (Gozzo et al., 2022).

AI can also be used for conflict resolution because it is perceived as a neutral party (Hohenstein and Jung, 2020; Dennis et al., 2023). Based on direct trust (positive/negative perception) and moderated trust (interpretation of past/future behavior) from interaction, humans tend to attribute blame to external factors if trust is high and to counterpart if trust is low. Lower social presence and limited information in digital collaboration may easily result in attributional errors. Here AI can act as a "moral crumple zone" reducing the trust loss between human communicators (Hohenstein and Jung, 2020).

Framework

The SLR revealed how the trust dynamics in digital human-AI teams were discussed from the perspectives of an individual trustor and AI trustworthiness, the adoption of AI in teams in relation to organizational trust, as emergent from the literature. These findings are synthesized into a framework, illustrated in Figure 2. Building on Söllner et al. (2016) and Fulmer and Gelfand (2012), it emphasizes the dynamic and interdependent nature of trust across levels (individual, organizational, and team) and multiple trust referents (human, AI, organization) (See Figure 2).

Human-AI trust is shaped by the individual perceptions of AI trustworthiness depending on its technical characteristics such as its capabilities, transparency, and anthropomorphism. In turn, human characteristics such as demographics, attitudes, expectations, and previous experience with AI impact perceived AI trustworthiness as well. Perceptions of AI trustworthiness affect organizational and team-level trust. In AI adoption, psychological safety, and work configurations as well as AI acceptance are critical to how trust is maintained in an organization. Individuals in teams need reassurance and knowledge sharing that their jobs are secure and that roles and tasks with AI are clear. To maintain organizational trust, privacy in platform data should be managed and not used for AI surveillance. Collaboration with AI influences team trust dynamics. The role of AI in a team, ranging from tool to teammate, affects the perception of autonomy and agency of AI that can further challenge trust. Yet AI can also build trust by supporting team processes, such as teaming, communication, cognition, and decision-making.

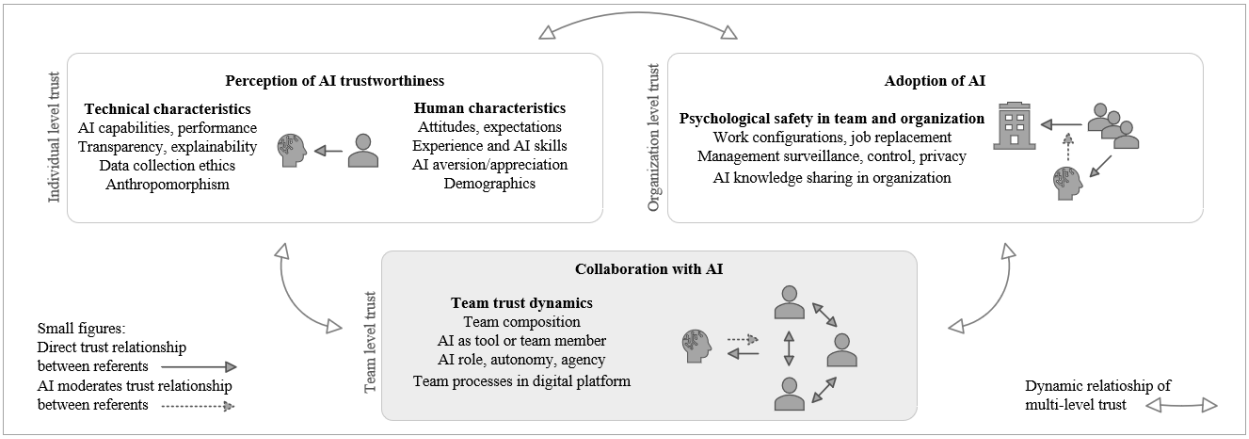

Figure 2. Multi-level trust referents and cross-level relationships in digital human-AI collaboration

Discussion

This study explored how AI affects trust in digital human-AI team collaboration. Synthesized in the framework, team-level trust dynamics build on individual perceived AI trustworthiness as well as psychological safety in the team and an organization. This review builds on earlier research (Glikson and Woolley, 2020; Kaur et al., 2022; Riz, 2022, Zecher et al., 2023) by emphasizing the socio-technical perspective in digital human-AI teams by integrating research from the domains of IS management and organization fields. This contributes to the understanding of the multi-referent and multi-level trust in human-AI teams in a digital context, which has been left for less attention in past research.

Research on AI in digital human-AI team collaboration has a mainly neutral or positive tone, but the related challenges for trust are also discussed (for example, Cardon et al., 2021; Getchell et al., 2022). Importantly, AI can support team tasks and social processes to build team trust and well-being (Bao et al., 2021; Hofeditz et al., 2022). AI is often trusted and sometimes even preferred over human collaborators (Jain et al., 2022; G. Zhang et al., 2023). However, the data and digital traces from digital platforms can lead to unethical and trust-decreasing AI control, and even surveillance (Cardon et al., 2021). As AI is usually digitally embedded or even invisible, deception of identity (Hancock et al., 2020; Seeber et al., 2020; G. Zhang et al., 2023) and human likeness (Getchell et al., 2022; Flathmann et al., 2023) are relevant challenges for social presence (Shin et al., 2023) and team cohesion (Hauptman et al., 2022; Hofeditz et al., 2022) that are vital for trust in the digital context.

Many scholars suggest shifting focus from trust towards AI as a tool to trust towards AI teammates (R. Zhang et al., 2021; Hauptman et al., 2022; Dennis et al., 2023). However, there are still barriers for AI to be considered a real colleague, as AI is missing a real social agency (Gozzo et al., 2022). Humans may not be able to distinguish it from a tool (Schelble et al., 2022), leading to further questions of what trust constructs are relevant.

Contribution to theory and practice

From a theoretical viewpoint, this review provides a framework of trust in digital human-AI team collaboration. The findings show that teams are mainly approached from individual or organizational viewpoints, calling for theoretical advancements for team-level processes. Also, the findings implicate a relatively general approach to trust and trustworthiness, while the human-AI team collaboration context would benefit from exploring more advanced human and AI-related trust referents and relationships. For practical implications, the framework is relevant for managers by highlighting the individual collaborator and organization viewpoint. Findings provide insight into how certain AI technical characteristics, often chosen by management, might be reflected in team members' trust and collaboration processes.

Limitations

This research does not come without limitations. First, the search was applied to the title, abstract, and keywords to get more precise results in the vast research, but this might have restricted the search. Moreover, this research is focused on the digital collaboration context, where literature is still scarce with varying terminologies, and therefore this review might have missed some relevant studies affecting the overall generalizability of the findings. Further, our framework requires validation with future research.

Further research

Since AI in digital teams is still non-routine (Dennis et al., 2023), current research focuses on human-controlled "AI" (Hofeditz et al., 2022). Therefore, field studies on various AI applications and research for example, on multimodal AI are suggested. Research on team-level attitudes, perceptions, and expectations would complement research on AI design, especially when current evidence comes from limited demographics (R. Zhang et al., 2021; Flathmann et al., 2023). Data is needed from more complex. and asynchronous teams (Hohenstein and Jung, 2020). New trust measures and methods could be explored, since using trust-related human constructs and measures for AI has been challenging (Schelble et al., 2022; Chowdhury et al., 2022). Current research on the evolution of trust is also scarce. Finally, there is a need for research on the dark side such as control and deception via AI, providing ground for research on evolving AI regulation and organizational policies.

Conclusion

This study reviewed the literature on the nexus of AI, trust, and digital team collaboration and synthesized the findings into a multi-level and multi-referent trust framework. The evolving AI technologies and related regulations provide researchers with new and important questions requiring an interdisciplinary understanding of the ethical and responsible AI in digital human-AI team collaboration.

REFERENCES

(Reviewed studies are indicated with *)

Anthony, C., Bechky, B. A., and Fayard, A.-L. 2023. "Collaborating with AI: Taking a System View to Explore the Future of Work," *Organization Science*, Advance online publication.

*Bao, Y., Cheng, X., De Vreede, T., and De Vreede, G.-J. 2021. "Investigating the relationship between AI and trust in human-AI collaboration," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, pp. 607-616.

Benbya, H., Pachidi, S., and Jarvenpaa, S. L. 2021. "Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research", *Journal of the Association for Information Systems* (22:2), pp. 281–303.

*Benke, I. 2020. "Towards Design Principles for Trustworthy Affective Chatbots in Virtual Teams", in *ECIS 2020 Research-in-Progress Papers*, p. 41.

*Bezrukova, K., Griffith, T. L., Spell, C., Rice, V., and Yang, H. E. 2023. "Artificial Intelligence and Groups: Effects of Attitudes and Discretion on Collaboration", *Group & Organization Management* (48:2), pp. 629–670.

*Cardon, P. W., Ma, H., and Fleischmann, C. 2021. "Recorded Business Meetings and AI Algorithmic Tools: Negotiating Privacy Concerns, Psychological Safety, and Control," *International Journal of Business Communication*, p. 232948842110370.

*Chhibber, N., Goh, J., and Law, E. 2022. "Teachable Conversational Agents for Crowdwork: Effects on Performance and Trust," in *Proceedings of the ACM on Human-Computer Interaction* 6(CSCW2), pp. 1–21.

*Chowdhury, S., Budhwar, P., Dey, P. K., Joel-Edgar, S., and Abadie, A. 2022. "AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework," *Journal of Business Research* (144), pp. 31–49.

*Correia, A., Fonseca, B., Paredes, H., Chaves, R., Schneider, D., and Jameel, S. 2021. "Determinants and Predictors of Intentionality and Perceived Reliability in Human-AI Interaction as a Means for Innovative Scientific Discovery," in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 3681–3684.

Costa, A. C., Fulmer, C. A., and Anderson, N. R. 2018. "Trust in work teams: An integrative review, multilevel model, and future directions," *Journal of Organizational Behavior* (39:2), pp. 169–184.

*Dennis, A. R., Lakhiwal, A., and 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), pp. 307–337.

Faraj, S., Pachidi, S., and Sayegh, K. 2018. "Working and organizing in the age of the learning algorithm," *Information and Organization* (28:1), pp. 62–70.

*Flathmann, C., McNeese, N. J., Schelble, B., Knijnenburg, B., and Freeman, G. 2023. "Understanding the impact and design of AI teammate etiquette," *Human–Computer Interaction*, Advance online publication.

*Fleischmann, C., Cardon, P., and Aritz, J. 2021. "Acceptance of Speech-to-Text Technology: Exploring Language Proficiency and Psychological Safety in Global Virtual Teams," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, pp. 411-420.

Fulmer, C., and Gelfand, M. J. 2012. "At What Level (and in Whom) We Trust: Trust Across Multiple Organizational Levels," *Journal of Management* (38:4), pp. 1167–1230.

Getchell, K. M., Carradini, S., Cardon, P. W., Fleischmann, C., Ma, H., Aritz, J., and Stapp, J. 2022. "Artificial Intelligence in Business Communication: The Changing Landscape of Research and Teaching," Business and Professional Communication Quarterly* (85:1), pp. 7–33.

Gkinko, L., and Elbanna, A. 2023. "Designing trust: The formation of employees' trust in conversational AI in the digital workplace," *Journal of Business Research* (158), p. 113707.

Glikson, E., and Woolley, A. W. 2020. "Human Trust in Artificial Intelligence: Review of Empirical Research," *Academy of Management Annals* (14:2), pp. 627–660.

*Gozzo, M., Woldendorp, M. K., and de Rooij, A. 2022. "Creative Collaboration with the 'Brain' of a Search Engine: Effects on Cognitive Stimulation and Evaluation Apprehension," in *ArtsIT 2021, LNICST 422,* M. Wölfel, J. Bernhardt, Thiel, S. (eds.), Springer, Cham, pp. 209–223.

*Hancock, J. T., Naaman, M., and Levy, K. 2020. "AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations," *Journal of Computer-Mediated Communication* (25:1), pp. 89–100.

*Hauptman, A. I., Duan, W., and McNeese, N. J. 2022. "The Components of Trust for Collaborating With AI Colleagues," *Companion Computer Supported Cooperative Work and Social Computing*, pp. 72–75.

*Hofeditz, L., Harbring, M., Mirbabaie, M., and Stieglitz, S. 2022. "Working with ELSA – How an Emotional Support Agent Builds Trust in Virtual Teams," in *Proceedings of the 55th Hawaii International Conference on System Sciences*, pp. 418-427.

*Hohenstein, J., and Jung, M. 2020. "AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust," *Computers in Human Behavior* (106), p. 106190.

*Jain, R., Garg, N., and Khera, S. N. 2022. "Effective human–AI work design for collaborative decision-making," *Kybernetes*, Advance online publication.

Jarrahi, M. H. 2018. "Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making," *Business Horizons* (61:4), pp. 577–586.

Kaur, D., Uslu, S., Rittichier, K. J., and Durresi, A. 2022. "Trustworthy Artificial Intelligence: A Review," *ACM Computing Surveys (CSUR)* (55:2).

Kellogg, K. C., Valentine, M. A., and Christin, A. 2020. "Algorithms at Work: The New Contested Terrain of Control," *Academy of Management Annals* (14:1), pp. 366–410.

Khakurel, J., and Blomqvist, K. 2022. "Artificial Intelligence Augmenting Human Teams. A Systematic Literature Review on the Opportunities and Concerns," in *International Conference on Human-Computer Interaction*, pp. 51-68.

Kitchenham, B., and Charters, S. 2007. "Guidelines for performing Systematic Literature Reviews in Software Engineering," (EBSE Technical Report EBSE-2007-01).

Lankton, N. K., Harrison Mcknight, D., and Tripp, J. 2015. "Technology, Humanness, and Trust: Rethinking Trust in Technology," *Journal of the Association for Information Systems* (16:10), pp. 1.

Makarius, E., Mukherjee, D., Fox, J., and Fox, A. 2020. "Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization," *Journal of Business Research* (120), pp. 262-273.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. 1995. "An Integrative Model of Organizational Trust," *The Academy of Management Review* (20:3), pp. 709–734.

Østerlund, C., Jarrahi, M. H., Willis, M., Boyd, K., and T. Wolf, C. 2021. "Artificial intelligence and the world of work, a co-constitutive relationship," *Journal of the Association for Information Science and Technology* (72:1), pp. 128–135.

Pettersen, L. 2019. "Why Artificial Intelligence Will Not Outsmart Complex Knowledge Work," *Work, Employment and Society* (33:6), pp. 1058–1067.

Razmerita, L., Brun, A., and Nabeth, T. 2022. "Collaboration in the Machine Age: Trustworthy Human-AI Collaboration," in *Advances in Selected Artificial Intelligence Areas: World Outstanding Women in Artificial Intelligence*, M. Virvou, G. A. Tsihrintzis and L. C. Jain (eds.), Springer International Publishing, pp. 333–356.

Rix, J. 2022. "From Tools to Teammates: Conceptualizing Humans' Perception of Machines as Teammates with a Systematic Literature Review," in *Proceedings of the 55th Hawaii International Conference on System Sciences*, pp. 398-407.

Robert, L. P., Denis, A. R., and Hung, Y.-T. C. 2009. "Individual Swift Trust and Knowledge-Based Trust in Face-to-Face and Virtual Team Members," *Journal of Management Information Systems* (26:2), pp. 241–279.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. 1998. "Not so different after all: A cross-discipline view of trust," *Academy of Management Review* (23:3), pp. 393-404.

Russell, S., and Norvig, P. 2021. *Artificial Intelligence: A Modern Approach (4th Ed.)*, Pearson Education Limited.

*Schelble, B. G., Flathmann, C., McNeese, N. J., Freeman, G., and Mallick, R. 2022. "Let's Think Together! Assessing Shared Mental Models, Performance, and Trust in Human-Agent Teams," in *Proceedings of the ACM on Human-Computer Interaction 6(GROUP)*, pp. 1–29.

*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., and Söllner, M. 2020. "Machines as teammates: A research agenda on AI in team collaboration," *Information & Management* (57:2), p. 103174.

*Shin, D., Kim, S., Shang, R., Lee, J., and Hsieh, G. 2023. "IntroBot: Exploring the Use of Chatbot-assisted Familiarization in Online Collaborative Groups," *in Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–13.

Shneiderman, B. 2020. "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy," *International Journal of Human–Computer Interaction* (36:6), 495–504.

Siemon, D., and Strohmann, T. 2021. "Human-AI Collaboration: Introducing the Virtual Collaborator," in Collaborative Convergence and Virtual Teamwork for Organizational Transformation, J. Zhao (ed.), IGI Global, pp. 105-119.

Söllner, M., Hoffmann, A., and Leimeister, J. M. 2016. "Why different trust relationships matter for information systems users," *European Journal of Information Systems* (25:3), 274–287.

van der Werff, L., Blomqvist, K., and Koskinen, S. 2021. "Trust cues in artificial intelligence: A multilevel case study in a service organization," in *Understanding trust in organizations: A multilevel perspective*, N. Gillespie, C. A. Fulmer, & R. J. Lewicki (eds.), Routledge/Taylor & Francis Group, pp. 307–333.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," *Communications of the Association for Information Systems* (37), 9.

Zercher, D., Jussupow, E., and Heinzl, A. 2023. "When AI joins the Team: A Literature Review on Intragroup Processes and their Effect on Team Performance in Team-AI Collaboration," in *ECIS 2023 Research Papers*, 307.

*Zhang, G., Chong, L., Kotovsky, K., and Cagan, J. 2023. "Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation," *Computers in Human Behavior* (139), 107536.

*Zhang, R., McNeese, N. J., Freeman, G., and Musick, G. 2021. "'An Ideal Human': Expectations of AI Teammates in Human-AI Teaming," in *Proceedings of the ACM on Human-Computer Interaction, 4(CSCW3)*, 1–25.