INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE (IMJRISE)
Online ISSN: 3028-032X | Print ISSN: 3028-0370
PUBLISHED BY: RISEJOURNALS.ORG
Volume 2, Issue no. 6 | **169**

Exploring Human-AI Interaction: A Systematic Review of Trust, Ethics, and Adoption in Artificial Intelligence Systems

DOI: 10.5281/zenodo.15628541

Jason Isaac III A. Rabi
Cebu Technological University, M.J. Cuenco Ave., Corner R. Palma St., Cebu City, Philippines
Luray II National High School, Department of Education, Toledo City Division
https://orcid.org/0000-0003-0912-2058

Abstract:

Artificial Intelligence (AI) is rapidly reshaping the landscape of human communication through advances in natural language processing (NLP), speech recognition, and conversational interfaces. This paper critically examines the dual impact of AI on both interpersonal and human-machine communication, emphasizing the balance between innovation and its ethical, psychological, and social implications. Drawing from an extensive review of scholarly literature, the study identifies key themes such as human-AI collaboration, the role of trust and transparency, ethical responsibility, cognitive dissonance, and the evolving dynamics of social interaction in AI-mediated contexts. The paper explores how AI systems influence user perception, emotional engagement, and identity formation, while also raising concerns about manipulation, overreliance, and authenticity in communication. It highlights the importance of integrating ethical design principles—such as fairness, accountability, and inclusivity—into the development of AI communication tools. The study concludes by proposing guidelines for creating AI systems that support and enrich human discourse while safeguarding psychological well-being and social cohesion. This work contributes to a growing body of interdisciplinary research that seeks to understand and guide the responsible evolution of AI in communication.

Keywords: Artificial Intelligence, human-AI communication, natural language processing, ethics

Introduction:

Artificial intelligence (AI) has become increasingly embedded in modern communication systems, with technologies like chatbots, virtual assistants, and automated speech recognition now commonplace in everyday life. These tools are transforming how humans engage with digital platforms, offering efficiency and scalability in various domains such as customer service, healthcare, education, and entertainment. Unsupervised automatic speech recognition, for example, has allowed machines to interact using natural language without extensive manual labeling or training data (Aldarmaki et al., 2022), making AI more accessible and responsive to human communication patterns.

Beyond technical performance, AI's influence on human interaction raises deeper questions about social cognition and perceived authenticity. As AI systems begin to simulate conversational behaviors and decision-making processes, people may anthropomorphize these tools, forming social bonds or expectations that are typically reserved for human partners (Čaić et al., 2019). The emotional and cognitive effects of interacting with intelligent agents—particularly those designed to mimic empathy or assertiveness—are still being understood. Studies have shown that social robots and assertive AI systems can evoke trust or discomfort depending on how conflict or goal alignment is managed (Babel et al., 2021; Babel et al., 2022).

The adoption of AI-powered communicative tools is also shaped by external factors such as user trust, technological design, and institutional transparency. Amershi et al. (2019) proposed a foundational set of guidelines to support trustworthy and effective human-AI interaction, emphasizing predictability, control, and user feedback. These principles are crucial as AI systems increasingly influence decision-making and dialogue. In academic and professional contexts, for instance, tools like PEERAssist use AI to analyze peer-review discourse and predict outcomes, potentially transforming scholarly communication processes (Bharti et al., 2021).

Despite these advancements, there are growing concerns about fairness, bias, and ethics in AI language systems. Canaan et al. (2019) emphasize the importance of ensuring fairness in AI-human interactions, particularly in competitive or evaluative settings such as games, hiring, or academic peer review. When AI systems lack transparency or fail to account for diverse user contexts, they risk reinforcing existing inequalities or producing unpredictable outcomes. The broader public adoption of these tools depends not only on technical reliability but also on whether users feel they are treated equitably and respectfully.

Literature Review:

The evolution of artificial intelligence (AI) has rapidly transformed the nature of human communication and collaboration. In recent years, AI-driven technologies such as chatbots, virtual assistants, and autonomous systems have increasingly mediated interpersonal and institutional interactions. Researchers have explored various

dimensions of these technologies, including trust, ethics, psychological implications, and socio-technical design. The literature offers a complex view of human-AI interaction, revealing both the profound benefits and potential risks of integrating AI into everyday life.

The acceptance and trust in AI technologies are foundational to their successful integration. Davis (1989) established the Technology Acceptance Model (TAM), emphasizing that perceived usefulness and ease of use drive user acceptance of new technologies. This framework remains central in contemporary studies assessing AI adoption.

For instance, Ciechanowski et al. (2019) explored how users interact with chatbots, revealing that responses to AI agents can be shaped by design factors, emotional resonance, and user expectations. Similarly, Chien et al. (2022) found that perceived intimacy and proactive behavior in humanoid robots significantly increased user trust. These findings suggest that emotional cues and social behavior embedded in AI agents play critical roles in human acceptance.

Beyond user acceptance, trust repair mechanisms have become a critical concern, especially when AI systems make mistakes or behave unexpectedly. Esterwood and Robert (2021) examined strategies for rebuilding trust in human-robot interaction, concluding that transparency, accountability, and timely response are key to restoring confidence. D'Alfonso (2020) also highlighted the importance of trust in AI systems within mental health contexts, where the stakes are particularly high. Moreover, Babel et al. (2021, 2022) developed psychological conflict resolution strategies for assertive and domestic service robots, demonstrating how AI systems can manage goal conflicts to sustain harmonious interactions.

Ethical considerations and social responsibilities have also garnered increasing attention in the literature. Cánas Delgado (2022) argued that AI systems involved in human collaboration must be designed with ethical frameworks that prioritize human values and well-being. Cheng et al. (2021) supported this by identifying critical issues in designing socially responsible AI algorithms, such as fairness, accountability, and transparency.

Domingos and Veve (2018) introduced the idea of "digital doubles," warning against the misuse of personal data by AI systems that simulate human behavior. These concerns align with warnings from scholars such as Hawking and Bostrom, who stress the existential risks of uncontrolled AI development (Hawking, 2014; Bostrom, 2014).

Social and psychological implications of AI interactions also form a key area of investigation. Dang and Liu (2022) explored how implicit theories about the human mind influence competitive or cooperative behavior toward AI robots. Their findings suggest that user psychology significantly moderates the nature of human-AI interaction. De Dreu and Weingart (2003) also pointed out that relationship conflict can affect team performance and satisfaction, a concept that can extend to AI-human teams. Ciechanowski et al. (2019) discussed the "uncanny valley" phenomenon, where overly human-like chatbots can evoke discomfort rather than trust, further complicating design decisions. This psychological tension is echoed in the work of Bharti et al. (2021), who examined how digital systems could influence subjective judgments such as peer review decisions.

AI's potential to contribute meaningfully to areas like healthcare and education further underscores the importance of ethical and responsible integration. D'Alfonso (2020) showcased AI's application in mental health, where chatbots and virtual therapists can provide scalable support. However, this raises ethical dilemmas around empathy, data privacy, and efficacy. Canaan et al. (2019) and Alawadhi et al. (2020) both discussed fairness and trust in AI systems, especially in competitive domains like gaming and autonomous driving. Furthermore, Aldarmaki et al. (2022) reviewed unsupervised speech recognition systems, revealing technical challenges that can impact user experience and trust. These studies illustrate the multidimensional nature of trust and performance in AI technologies.

Foundational texts further enrich the academic conversation around AI. Scholars like Russell and Norvig (2009), Nilsson (1980), and Kaplan and Haenlein (2019) have laid the groundwork for understanding AI principles and their societal implications. Jerry Kaplan (2016) and Roger Schank (1991) questioned the philosophical assumptions behind AI, while thinkers like Joseph Weizenbaum (1976) highlighted ethical boundaries.

More recent commentaries by Delcker (2018) and Beth Kindig (2020) scrutinized AI's bias and corporate agendas, emphasizing the importance of regulatory oversight and critical public discourse. Reports by Skills for Care and Nature News (2020) added practical perspectives on AI deployment in social care, highlighting the need for policies that ensure equity and safety.

The literature provides a holistic view of the growing entanglement between humans and AI systems. It highlights not only the technical and psychological factors that influence trust and acceptance but also the ethical, social, and philosophical challenges that must be addressed.

As AI technologies continue to evolve and become more deeply embedded in daily life, it is crucial to develop interdisciplinary frameworks that ensure these tools are used in ways that support human dignity, well-being, and

autonomy. Future research should continue to explore these dynamics, focusing on co-design practices, inclusive AI policies, and adaptive trust models that evolve alongside technological advancements.

Methodology

This study adopts a systematic literature review (SLR) approach to analyze and synthesize existing research on artificial intelligence (AI), particularly focusing on human-AI communication, trust dynamics, and ethical considerations. The SLR method ensures a comprehensive and structured examination of current academic discourse by identifying, evaluating, and integrating findings from a wide range of scholarly sources.

The review includes peer-reviewed journal articles, conference proceedings, and authoritative academic texts published primarily between 2010 and 2024. Priority was given to high-impact studies that are frequently cited and widely recognized in the fields of computer science, psychology, and ethics. Among the core sources analyzed are Aldarmaki et al. (2022), which examines advancements in speech recognition technologies; Amershi et al. (2019), which outlines human-AI interaction guidelines; and Cañas Delgado (2022), which explores the ethical frameworks necessary for AI-human collaboration. These studies were selected based on their relevance to the themes of AI adoption, communication efficacy, trust-building mechanisms, and responsible AI design.

A thematic analysis was conducted to extract and categorize key insights from the literature. This involved coding and grouping findings into several core themes: (1) AI adoption and user engagement, (2) quality of AI-human communication, (3) the role of transparency and explainability in building trust, and (4) ethical and social implications of AI deployment. Each theme was critically examined to understand both the consensus and divergences across the studies reviewed.

This methodology allows for a nuanced and interdisciplinary understanding of how AI technologies are perceived, trusted, and ethically evaluated in human-centered contexts. The approach also highlights gaps in the existing literature and suggests avenues for future research in the field of human-AI interaction.

Findings and Discussion:

Opportunities in Human-AI Communication

Artificial Intelligence (AI) continues to open up significant opportunities in the realm of human-AI communication. One prominent advancement is in the field of speech recognition. Aldarmaki et al. (2022) present a comprehensive overview of unsupervised automatic speech recognition technologies, which allow for more natural, real-time verbal exchanges between humans and AI. Such technologies are instrumental in building voice-based interfaces and virtual assistants that emulate human conversational behavior.

Chatbots have also demonstrated the potential to enhance user engagement. Ciechanowski et al. (2019) explored the nuanced design elements that make chatbot conversations more human-like, finding that empathetic and context-aware responses can significantly improve interaction quality. Similarly, Čaić et al. (2019) argue that social robots offer emotional and cognitive support in customer-facing environments, such as eldercare and hospitality, thus blurring the line between service technology and human companion.

In healthcare, D'Alfonso (2020) highlights the increasing use of conversational agents in delivering initial mental health support. These agents can help mitigate barriers such as stigma and accessibility, providing a non-judgmental space for users to share concerns. Domingos and Veve (2018) further introduce the concept of "digital doubles," AI-driven models that mirror user behavior and preferences. These doubles can anticipate needs and tailor communications, suggesting a shift toward hyper-personalized AI interfaces.

The educational sector also benefits from AI-enhanced communication. Bharti et al. (2021) introduced PEERAssist, an AI tool that predicts peer-review decisions by analyzing linguistic patterns in manuscripts. This not only expedites the review process but also improves feedback quality and transparency.

AI systems are also making strides in trust-building. Amershi et al. (2019) proposed guidelines for human-AI interaction that promote transparency, predictability, and user control—factors critical to user trust. Chien et al. (2022) emphasize the role of intimacy and proactivity in establishing trust with humanoid robots. These elements create perceptions of agency and reliability, which are essential in settings that require cooperative tasks or emotional engagement.

Conflict resolution is another promising frontier. Babel et al. (2021) and Babel et al. (2022) explored psychological strategies and virtual reality-based evaluations of adaptive robots that can manage human-robot goal conflicts. Their work reveals that assertive yet empathetic robot behavior improves outcomes in collaborative environments, such as domestic service contexts.

Moreover, Cheng et al. (2021) examine how socially responsible AI algorithms can be designed to align with ethical standards, promoting fairness and inclusivity in communication. This complements the ethical discourse raised by Cañas Delgado (2022), who urges developers to consider the shared responsibility when humans collaborate with AI agents, particularly regarding biases and decision accountability.

The issue of fairness in AI communication is echoed by Canaan et al. (2019), who discuss how AI performance benchmarks must ensure equitable comparisons with human capabilities. Dang and Liu (2022) add that individuals' implicit theories about human minds affect how they respond to AI agents, impacting cooperation and competition dynamics in joint tasks.

From a user adoption perspective, Davis (1989) noted that perceived usefulness and ease of use strongly influence technology acceptance—an idea that remains relevant in current AI communication tools. Alawadhi et al. (2020), in their study on autonomous systems, reinforce this by identifying trust, usability, and transparency as key adoption factors.

Historically, researchers such as Russell and Norvig (2009) and Nilsson (1980) envisioned AI as a transformative force in human reasoning. Contemporary discussions, including those by Bostrom (2014) and Hawking (2014), expand on these ideas by warning about the potential existential risks of misaligned AI goals. Meanwhile, Delcker (2018) and Gibney (2020) emphasize the ongoing struggle for ethical AI development amid increasing commercial and geopolitical stakes.

In the public discourse, authors like Beth Kindig (2020) and Dina Bass (2016) have pointed to real-world applications of AI, from cancer diagnostics to autonomous surgery, underscoring AI's communicative roles in high-stakes domains. These developments align with Kaplan and Haenlein's (2019) analysis of AI's interpretive roles and its ability to act as an interface between humans and complex information systems.

The opportunities in human-AI communication are vast and multidisciplinary, spanning emotional, ethical, and technical dimensions. These technologies are not merely functional; they are becoming increasingly human-centric—designed to understand, adapt, and ethically engage with users. The emerging literature supports a future where AI is not just a tool but a communicative partner, embedded in everyday interactions from mental health to education, from domestic settings to complex scientific endeavors. However, to harness these opportunities, a conscientious and collaborative approach is needed to align technological capabilities with human values and expectations.

Risks and Ethical Concerns

While the integration of artificial intelligence (AI) in human systems brings numerous advantages, it also raises a host of risks and ethical concerns that demand careful scrutiny. One prominent issue is the potential for conflict between human users and AI agents, especially when machines are perceived as undermining human authority or autonomy. Babel et al. (2021; 2022) highlight that such conflicts are not merely technical but psychological in nature. When AI appears too assertive or makes morally ambiguous decisions, users may feel disempowered or even threatened. This tension can erode cooperation and result in a breakdown of trust between humans and machines.

A major contributor to these tensions is the opacity of AI decision-making processes, often referred to as the "black box problem" (Delcker, 2018). Many advanced AI systems, especially those based on deep learning, offer little to no transparency in how decisions are made, making it difficult for users to understand, challenge, or trust the outputs. As AI systems are increasingly deployed in high-stakes areas such as healthcare, criminal justice, and autonomous transportation, the inability to audit or explain algorithmic outcomes poses significant ethical and operational risks (Domingos & Veve, 2018; Alawadhi et al., 2020).

Trust, a critical factor in human-AI collaboration, can be fragile. According to Davis's (1989) Technology Acceptance Model (TAM), the perceived usefulness and ease of use of technology influence users' willingness to adopt it. However, Esterwood and Robert (2021) emphasize that when trust is broken—due to unexpected behavior, ethical lapses, or a lack of transparency—rebuilding it becomes an uphill battle. This is particularly relevant in social or service settings, where user expectations for empathy and clarity are high. D'Alfonso (2020) warns that excessive reliance on AI in such contexts can lead to an erosion of interpersonal skills, empathy, and emotional intelligence among humans, especially in sectors like mental health and education.

Ethical concerns intensify when human-AI collaboration occurs in contexts that involve judgment, discretion, and emotional intelligence. Cañas Delgado (2022) argues that the moral accountability of decisions becomes blurry when both human and AI agents contribute to outcomes. This ambiguity raises critical questions: Who is to blame when an AI-assisted diagnosis is incorrect? Who should be held responsible when an autonomous vehicle causes an accident? These questions are especially pressing when AI systems are trained on biased datasets, potentially reinforcing or amplifying existing social inequalities.

Cheng et al. (2021) underline the issue of embedded biases in AI algorithms. These biases can stem from skewed training data, flawed assumptions in model design, or a lack of contextual awareness. For example, facial recognition systems have been shown to perform poorly on people of color, leading to disproportionate false positives or negatives (Canaan et al., 2019). In communication systems, such biases may manifest in linguistic inaccuracies or culturally insensitive outputs, affecting fairness and inclusivity. These algorithmic shortcomings can perpetuate discrimination and challenge the ethical justification for deploying AI at scale.

Another dimension of ethical risk involves emotional manipulation and anthropomorphism. Studies by Chien et al. (2022) and Ciechanowski et al. (2019) suggest that the more human-like an AI appears, the more likely users are to attribute emotional and cognitive capabilities to it. This phenomenon can create unrealistic expectations and blur the line between simulation and genuine interaction. For instance, users might over-rely on chatbots for emotional support, believing them to be more capable than they actually are. This false sense of companionship can lead to emotional dependency, which is ethically questionable when the AI has no true understanding or empathy.

The design and deployment of AI also raise broader philosophical concerns about human identity, autonomy, and the nature of decision-making. Dang and Liu (2022) show that people's implicit theories about the human mind influence their willingness to collaborate with AI. Those who believe that cognition is uniquely human are less likely to accept AI as a legitimate partner in creative or moral tasks. This resistance is not merely irrational; it stems from deeply held values about human dignity and agency.

Amershi et al. (2019) propose a set of guidelines for human-AI interaction that emphasize transparency, predictability, and user control. These design principles are crucial to mitigating risks, but their implementation is uneven across sectors. As Bharti et al. (2021) argue, efforts to make AI systems explainable and fair must go beyond technical solutions to include ethical reflection and stakeholder participation. Otherwise, the very systems designed to support human flourishing may end up undermining it.

Finally, the accelerating pace of AI development demands that we anticipate ethical dilemmas before they escalate. Scholars like Bostrom (2014) and Hawking (2014, as cited in BBC News) have warned that unchecked AI development could threaten not only societal norms but humanity itself. The battle for ethical AI, as noted by Gibney (2020), is not simply about programming fairness—it is about setting boundaries, enforcing accountability, and maintaining a clear-eyed view of what machines should and should not do.

The integration of AI into human life is a double-edged sword. While the potential benefits are considerable, the risks—ranging from bias and loss of empathy to eroded trust and moral ambiguity—are significant. A multidisciplinary, ethically informed approach is essential to ensure that AI serves humanity rather than subverts it. As the field advances, ethical vigilance must be seen not as a constraint on innovation but as its moral compass.

Social and Psychological Implications

AI language systems have profound psychological and social implications that shape how individuals interact with technology and with one another. As AI becomes increasingly integrated into human environments, understanding its impact on users' cognition, emotions, and social behaviors becomes essential.

One psychological factor influencing user interaction with AI is how individuals conceptualize intelligence and human cognition. Dang and Liu (2022) argue that users' implicit theories about the human mind—whether they see it as malleable or fixed—affect whether they respond to AI cooperatively or competitively. These mental models influence emotional engagement, openness to AI-generated suggestions, and trust in AI decisions.

Trust is a foundational element in human-AI interactions, particularly in emotionally charged or high-stakes contexts. Chien et al. (2022) examined how intimacy and proactive behavior in humanoid robots can foster user trust. When AI demonstrates emotional attunement or responsiveness, it can simulate interpersonal warmth, leading to stronger bonds and greater user compliance. However, this bond is vulnerable to disruption by the "uncanny valley" phenomenon—a state where AI appears almost human but not convincingly so—triggering discomfort and distrust in users (Ciechanowski et al., 2019). These eerie inconsistencies in behavior or appearance may lead users to perceive AI as deceptive or inauthentic, undermining social connection.

Furthermore, the introduction of AI into collaborative environments—such as workplaces or educational settings—can generate new forms of interpersonal and relational conflict. According to De Dreu and Weingart (2003), unresolved relationship conflicts in teams lower satisfaction and hinder performance. When AI participates as a "team member," especially in mixed human-AI groups, it may inadvertently provoke friction through misunderstood intent, rigid behavior, or perceived bias.

This challenge is echoed in Babel et al. (2021, 2022), who explore how assertive robots that are goal-driven may create tension in shared human-robot tasks. Their studies advocate for conflict resolution strategies tailored for AI, suggesting that robots must be equipped with adaptive, context-sensitive approaches to defuse disagreements. These strategies include programmed apologies, compromise behaviors, and context-aware withdrawal—similar to human social tactics.

Trust repair is another key area of concern. Esterwood and Robert (2021) emphasize the necessity of trust-repair mechanisms when AI systems fail or behave unexpectedly. Transparency, explainability, and feedback loops are essential tools to maintain or restore trust. These mechanisms allow users to understand not only what the AI is doing, but why—enabling more predictable and psychologically comfortable interactions.

Amershi et al. (2019) provide a comprehensive framework for human-AI interaction, offering 18 guidelines that stress transparency, user control, and timely feedback. Their work supports the idea that AI should behave in a way that respects human decision-making authority, reducing the likelihood of conflict or distrust.

Beyond interpersonal dynamics, the social perception of AI also impacts its adoption. Čaić et al. (2019) note that AI systems are evaluated not only based on technical performance but also on perceived social intelligence—how well they understand social cues, empathize, and behave appropriately. This is particularly critical in care settings, service robots, and educational environments where emotional sensitivity is essential.

Moreover, cultural and ethical perspectives shape how users perceive and accept AI systems. Cañas Delgado (2022) warns of ethical dilemmas arising when AI agents participate in decision-making processes traditionally reserved for humans, such as healthcare, criminal justice, or education. These dilemmas are not only moral but psychological, generating anxiety and resistance among users who fear AI overreach or misjudgment.

User adoption is also influenced by perceptions of usefulness and ease of use, as originally outlined by Davis (1989) and echoed in recent studies like Alawadhi et al. (2020). If AI systems are perceived as overly complex, intrusive, or emotionally insensitive, users may reject them, regardless of technical superiority.

Even in specialized domains like autonomous driving (Alawadhi et al., 2020) or unsupervised speech recognition (Aldarmaki et al., 2022), human factors—especially trust, safety perception, and emotional response—play a decisive role. This underlines that the success of AI integration is as much psychological as it is technical.

Finally, scholars like D'Alfonso (2020) highlight the emerging role of AI in mental health, where chatbots and digital companions provide therapeutic conversations. While promising, these applications raise concerns about over-reliance, misdiagnosis, or emotional manipulation, further stressing the importance of ethical and psychological safeguards.

AI language systems exert significant psychological and social influence on their users. Issues such as trust, conflict resolution, emotional perception, and ethical tension shape user experience and system acceptance. As AI continues to evolve, addressing these human-centered factors will be essential for sustainable, responsible integration into society.

Conclusion:

Artificial Intelligence (AI) represents a transformative force in human communication—one that offers both unprecedented opportunities and significant challenges. On the one hand, AI systems enhance communication efficiency, personalize user interactions, and enable access to information across linguistic and geographical barriers. They assist in decision-making, streamline collaboration, and support inclusivity by offering tools that bridge communication gaps. However, these benefits are accompanied by ethical, psychological, and social concerns.

As AI becomes more deeply embedded in everyday interactions, questions arise about trust, transparency, and the authenticity of human-AI relationships. Users may develop misguided perceptions of AI's capabilities, leading to overreliance or misplaced trust. Furthermore, phenomena such as the uncanny valley and emotionally manipulative design features can disrupt user comfort and social cognition. These psychological implications underscore the importance of designing AI systems that are not only intelligent but also ethically grounded and socially aware.

To maximize AI's benefits while mitigating its risks, developers and stakeholders must prioritize transparency, inclusivity, and responsible innovation. Ethical guidelines should be integrated into the design and deployment of AI tools to ensure fairness, accountability, and user well-being. Trust-building mechanisms—such as explainability, user control, and feedback systems—are essential for fostering healthy human-AI dynamics.

Looking ahead, future research should prioritize longitudinal and interdisciplinary studies that investigate how AI-mediated communication affects human behavior, relationships, and societal norms over time. Such studies can

uncover deeper insights into identity formation, digital empathy, and the evolving nature of social interaction in AI-integrated environments. By adopting a proactive and ethically conscious approach, society can harness the full potential of AI while safeguarding human values and psychological well-being in the digital age.

References:

Alawadhi, M., Almazrouie, J., Kamil, M., & Khalil, K. A. (2020). A systematic literature review of the factors influencing the adoption of autonomous driving. *International Journal of System Assurance Engineering and Management, 11*(6), 1065–1082. https://doi.org/10.1007/s13198-020-00961-4

Aldarmaki, H., Ullah, A., Ram, S., & Zaki, N. (2022). Unsupervised automatic speech recognition: A review. *Speech Communication, 139*, 76–91. https://doi.org/10.1016/j.specom.2022.02.005

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). https://doi.org/10.1145/3290605.3300233

Babel, F., Kraus, J. M., & Baumann, M. (2021). Development and testing of psychological conflict resolution strategies for assertive robots to resolve human–robot goal conflict. *Frontiers in Robotics and AI, 7*, Article 591448. https://doi.org/10.3389/frobt.2020.591448

Babel, F., Vogt, A., Hock, P., Kraus, J., Angerer, F., Seufert, T., & Baumann, M. (2022). Step aside! VR-based evaluation of adaptive robot conflict resolution strategies for domestic service robots. *International Journal of Social Robotics, 14*(5), 1239–1260. https://doi.org/10.1007/s12369-021-00858-7

Bharti, P. K., Ranjan, S., Ghosal, T., Agrawal, M., & Ekbal, A. (2021). PEERAssist: Leveraging on paper-review interactions to predict peer review decisions. In *Towards open and trustworthy digital societies* (pp. 421–435). https://doi.org/10.1007/978-3-030-84769-9_30

Čaić, M., Mahr, D., & Oderkerken-Schröder, G. (2019). Value of social robots in services: Social cognition perspective. *Journal of Services Marketing, 33*(4), 463–478. https://doi.org/10.1108/JSM-02-2018-0080

Canaan, R., Salge, C., Togelius, J., & Nealen, A. (2019). Leveling the playing field: Fairness in AI versus human game benchmarks. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (pp. 1–8).

Cañas Delgado, J. J. (2022). AI and ethics when human beings collaborate with AI agents. *Frontiers in Psychology, 13*, Article 836650. https://doi.org/10.3389/fpsyg.2022.836650

Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially responsible AI algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research, 71*, 1137–1181. https://doi.org/10.1613/jair.1.12986

Chien, S.-Y., Lin, Y.-L., & Chang, B.-F. (2022). The effects of intimacy and proactivity on trust in human-humanoid robot interaction. *Information Systems Frontiers, 26*, 75–90. https://doi.org/10.1007/s10796-022-10324-y

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems, 92*, 539–548. https://doi.org/10.1016/j.future.2018.01.055

D'Alfonso, S. (2020). AI in mental health. *Current Opinion in Psychology, 36*, 112–117. https://doi.org/10.1016/j.copsyc.2020.04.005

Dang, J., & Liu, L. (2022). Implicit theories of the human mind predict competitive and cooperative responses to AI robots. *Computers in Human Behavior, 134*, Article 107300. https://doi.org/10.1016/j.chb.2022.107300

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

De Dreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology, 88*(4), 741–749. https://doi.org/10.1037/0021-9010.88.4.741

Domingos, P., & Veve, A. (2018). Our digital doubles. *Scientific American, 319*(3), 88–93. https://www.jstor.org/stable/27173625

Esterwood, C., & Robert, L. P. (2021). Do you still trust me? Human-robot trust repair strategies. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)* (pp. 183–188). https://doi.org/10.1109/RO-MAN50785.2021.9515536