11

Challenges of Generative AI on Human–AI Interaction and Collaboration

Shaobo Liang and Yiting Cai

11.1 User-level Challenges

11.1.1 Trust Building

Trust and comprehension of AI systems are essential components of effective human–AI collaboration and interaction. The degree of trust that users place in AI significantly influences their perceptions of AI-generated content and serves as a critical psychological factor that impacts their adherence to system recommendations. The decision-making processes employed by AI are inherently complex and present certain technical challenges, which can hinder users' understanding of how AI-generated content is created. Furthermore, users often lack the capacity to modify system recommendations, which complicates their ability to lower their defenses against generative AI. Currently, strategies such as enhancing users' decision-making control and elucidating the decision-making mechanisms of AI systems are frequently implemented to foster trust. Nevertheless, these approaches also present limitations, as they raise questions regarding the extent to which AI systems should offer explanations.

Research indicates that excessive elaboration on the reasoning processes of AI may heighten users' perceptions of task complexity, resulting in cognitive overload and a subsequent decline in trust toward generative AI systems (Westphal et al., 2023). This phenomenon may adversely affect users' favorable assessments of human–AI interactions. According to cognitive load theory, human working memory has inherent limitations, and effectively managing cognitive load intensity is essential for facilitating successful learning outcomes. For individuals with lower cognitive abilities who encounter difficulties with complex tasks, an excess of task-related information can overwhelm their cognitive processing capacities, thereby hindering their ability to understand and adhere to the intended applications of generative AI.

268

Furthermore, trust in AI systems is a dynamic construct, with users' comprehension of AI performance and reliability evolving gradually as familiarity and individual cognitive capabilities improve. Given that human–AI collaboration and interaction represent emerging technologies, it is reasonable to expect that the development of trust will occur at a gradual pace.

In the context of fostering trust between users and generative AI, it is essential to recognize that factors extending beyond users' subjective influences are also significant. These include advancements in emotional contagion pathways and the humanization of AI. Generative AI systems are required to process various types of data while simultaneously providing feedback and facilitating interaction (Lukyanenko et al., 2022). Products such as Character AI, Janitor.AI, and Pi have garnered substantial user engagement, underscoring the critical role of emotional companionship that AI models can offer in the realm of human–AI collaboration and interaction. It is imperative for AI systems to effectively manage the degree of emotional transmission to engage users and cultivate their trust. Nevertheless, the extent of humanization must be judiciously calibrated. While enhancing the human-like qualities of AI systems can improve user-friendliness, excessive humanization may provoke skepticism and potentially erode users' trust in the professionalism of the AI.

11.1.2 Algorithm Aversion

Some people in society demonstrate a pronounced aversion to content generated by AI, which is often manifested through behaviors such as reluctance to utilize AI products and dismissal of AI-generated outputs. This resistance impedes the potential for collaboration between generative AI systems and their users (Cheng et al., 2022). The psychological factors underlying this algorithmic aversion can be classified into three distinct categories: a perceived competition with AI, a desire for transparency and control in decision-making processes, and a bias against the creative capabilities of AI.

Generative AI represents a culmination of significant technological advancements, and the collaboration and interaction between humans and AI have substantially transformed both daily life and production processes. In light of these profound changes, users have expressed concerns regarding potential job displacement and have experienced anxiety related to technology, resulting in the emergence of algorithmic aversion.

Even when AI models can provide more accurate and higher-quality responses than their human counterparts, users frequently exhibit a preference for human references. This phenomenon is referred to as "algorithmic aversion, a tendency to favor human input over AI" (Mariadassou et al., 2024, p. 2).

Research indicates that artworks produced through the collaboration of human creators and AI are often perceived as more aesthetically pleasing. Furthermore, the involvement of AI can enhance the creativity of human-generated works (Hitsuwari et al., 2023). The interaction and collaboration between humans and AI present significant opportunities in the realm of artistic creation; however, algorithmic aversion remains a substantial obstacle to its advancement. There is a prevailing belief that works created by humans possess greater beauty and are infused with a sense of humanistic care. The increasing prevalence of AI-generated artworks, coupled with the challenges associated with distinguishing between human and AI creations, has, in certain respects, undermined the dominance of human creators, thereby exacerbating negative perceptions and evaluations of AI-generated content.

Although people may dislike the label of "algorithm" or "AI," they often express appreciation for the actual outputs generated by generative AI. For example, when the origin of content is ambiguous, users often exhibit a more favorable response to jokes suggested by algorithms than to those proposed by humans. In the context of emotional support, content generated by ChatGPT is perceived as more attentive and is rated higher in terms of emotional value. This suggests that users are not inherently opposed to the content generated by algorithms; rather, it is their reluctance toward the labels "AI" and "algorithm" that hinders the wider acceptance of generative AI technologies (Elyoseph et al., 2023; Yeomans et al., 2019).

11.1.3 Acceptability

Individual differences among users significantly contribute to the varying levels of acceptance of generative AI. Factors such as cultural context and patterns of emotional expression play a crucial role in shaping these differences. To effectively meet the needs of a diverse user base, generative AI must integrate appropriate frameworks that account for emotional expression and cultural contexts. This necessity poses challenges for the localization strategies and cultural sensitivity of AI systems.

Firstly, linguistic and behavioral preferences are critical factors that influence the adoption of generative AI. AI systems must conduct extensive research on local dialects, semantics, and accents within the fields of speech recognition and natural language processing to achieve objectives related to language comprehension, language generation, and multilingual prediction. Moreover, in the context of product design, it is imperative to consider the usage habits and requirements of users from various regions to enhance the overall user experience (Khurana et al., 2023).

Secondly, users' perceptions of fairness regarding generative AI play a crucial role in its acceptance. The challenges associated with racial and cultural biases present in existing generative AI models remain inadequately addressed (Gilliard, 2022). When these biases are applied in critical domains such as healthcare, employment, and security governance, they have the potential to exacerbate discrimination and reinforce power imbalances, thereby diminishing users' trust in generative AI.

11.2 Algorithm Optimization Challenges

11.2.1 Theoretical Gaps

Generative AI is a significant advancement in technological development; however, there exist notable gaps in the emerging theories surrounding human–AI interaction. Specifically, there is a lack of clarity regarding the advantages of AI and the application of these theories to enhance the usability and acceptability of AI systems. A discernible disconnect persists between theoretical research on generative AI and the rapid pace of practical advancements in the field. In recent years, generative AI and AI-driven self-service systems have experienced swift progress and have been implemented in critical domains such as autonomous driving, healthcare, and legal services. On one hand, generative AI has emerged as a reliable tool for improving human efficiency and accuracy, facilitating human–AI interaction and collaboration, and alleviating individuals from repetitive tasks. Conversely, there remains a lack of consensus on the allocation of responsibility in tasks that involve generative AI and automated systems. The inherent complexity of AI models complicates the ability of human agents to assume full responsibility for the outcomes generated by these systems (Königs, 2022). This situation has led critics to highlight the issue of a "responsibility gap" in the context of generative AI and human–AI collaboration, a concern that current theoretical frameworks are insufficient to address comprehensively.

Beyond the theoretical divide between generative AI and its practical applications, there exists a notable deficiency in empirical evidence that substantiates the advantages of human–AI interaction and collaboration within specific domains. In the field of education, for instance, intelligent robots and adaptive learning systems are increasingly utilized by both educators and learners (Chen et al., 2020). AI technologies facilitate personalized learning environments by customizing educational plans to align with individual learning styles and enhancing learner motivation. Furthermore, these technologies

assist educators in alleviating the burden of repetitive tasks, thereby allowing for more meaningful and individualized instruction. Despite the burgeoning interest in AI within educational research, there has been a lack of concerted efforts to integrate AI technologies with established educational theories. Consequently, this gap has hindered the ability to fully articulate the essential value of AI in the educational sector.

11.2.2 Inequality and Lack of Accessibility

Fairness and accessibility are critical directions in the development of generative AI, as these technologies should serve users from diverse educational backgrounds and economic conditions equally. Currently, the human–AI interaction services provided by AI systems pose certain technical barriers, challenging the knowledge levels of users and limiting the application market for generative AI. AI systems need to reduce the difficulty for non-expert users to engage with and benefit from generative AI, expanding the user base while leveraging technology to serve society. The accessibility for users with varying health conditions is another key point that generative AI must address. At present, generative AI has not fully met the expectations of its target audience. Fairness and accessibility are essential considerations in the advancement of generative AI, as these technologies must equitably serve users from diverse educational backgrounds and socioeconomic conditions. Currently, the human–AI interaction services offered by AI systems present certain technical barriers that challenge users' knowledge levels and restrict the market applicability of generative AI. It is imperative for AI systems to mitigate the complexities that non-expert users face in engaging with and benefiting from generative AI, thereby broadening the user base and utilizing technology to serve societal needs. Additionally, addressing accessibility for users with varying health conditions is a critical aspect that generative AI must prioritize. At present, generative AI has not fully incorporated technologies to support individuals with disabilities. For example, when visually impaired users utilize screen readers to access generative AI outputs, platforms such as ChatGPT fail to provide clear indications of where the output begins and ends, nor do they label the locations of buttons for copying, editing, or rating the content.

The integration of technology to assist individuals with disabilities is crucial. For example, visually impaired users who depend on screen readers to interact with generative AI outputs may encounter challenges, as ChatGPT fails to offer clear demarcations indicating the beginning and end of the output. Additionally, it does not adequately label the locations of buttons for copying, editing, or rating the content.

The unfairness of generative AI is reflected in aspects such as race, gender, and age. Research has found that generative AI associates terms like "Africa" with "poverty" and depicts all flight attendants as female. These outputs from generative AI are not objective representations of the real world but rather amplify unfair biases (Ananya, 2024). The discriminatory inclinations of generative AI models are frequently nuanced and challenging to identify. Mitigating this issue by eliminating biased content from training datasets is both resource-intensive and difficult to execute. Consequently, it is imperative to engage in a collaborative effort among governments, researchers, and users to oversee the content produced by AI models and to swiftly rectify any inappropriate biases present in the models' reasoning.

11.2.3 Emotional Design

The emotional expressions exhibited by AI systems can significantly impact user responses to the information presented. Positive emotions conveyed by AI may influence users through two distinct pathways: the affective pathway, which can enhance trust via emotional contagion, and the cognitive pathway of expectation–disconfirmation, which may diminish trust. Users' expectations regarding the emotional expression patterns of generative AI vary based on the context and the intended purpose of the interaction.

In addition to methods of emotional expression, users' emotional requirements for AI systems differ across various application contexts and needs. Research has investigated how the type of AI and the nature of human–AI collaboration influence consumer acceptance (Peng et al., 2022). The findings indicate that when AI functions as a supportive entity for humans, there is an increase in consumer acceptance of AI services for tasks that necessitate a high degree of warmth. However, this effect is not evident when AI operates under human supervision. Consequently, it is essential to enhance the emotional design framework when developing AI systems to accommodate diverse collaboration scenarios and modes, thereby more effectively addressing user needs.

11.2.4 Cultural and Linguistic Adaptation

Generative AI training datasets are frequently derived from extensive global datasets, which typically provide superior representation of mainstream languages and cultures. However, users from underdeveloped regions who utilize minority languages or possess distinct cultural backgrounds may be marginalized by generative AI systems. Consequently, their specific needs may not be

adequately recognized, resulting in challenges for these users in obtaining equivalent levels of service compared to their counterparts.

Culturally, research comparing the responses of five prominent language models – GPT-4, GPT-4 Turbo, GPT-3.5, and GPT-3 Turbo – against data from nationally representative surveys indicates the existence of cultural and linguistic biases (Tao et al., 2024). The findings suggest that the cultural values expressed by all examined language models are more closely aligned with those of English-speaking and Protestant European nations. From a linguistic perspective, many generative AI systems that are language-based depend on a restricted corpus of language data and frequently emphasize standardized language variants. This reliance can foster the perception that there exists a singular "correct" method of utilizing a specific language, thereby contributing to linguistic bias (Jenks, 2024).

While users typically perceive AI as more objective and rational than human decision-makers, AI algorithms frequently embody the subjective biases of their creators, including programmers, data scientists, and other human developers. This phenomenon hinders the advancement of generative AI and adversely affects the production of knowledge within human society.

11.2.5 Controlling the Degree of Humanization

Humanization represents a significant trend in the evolution of chatbots. Numerous studies have demonstrated that the incorporation of human-like attributes in AI systems, such as warmth and competence, can enhance user trust and positively influence user satisfaction during interactions (Han, 2021). In these contexts, generative AI affects users through various dimensions, including auditory and emotional responses, which contribute to an increase in overall interaction satisfaction. Generative AI is capable of performing tasks such as playing music, ordering products, and personalizing plans, with these human-like services directly contributing to improvements in users' life satisfaction and well-being. Nevertheless, fulfilling users' expectations for human-like services poses considerable challenges. A limited proportion of users believe that AI-based systems can deliver services that are more satisfying than those provided by humans (Zhu et al., 2023). Furthermore, frequent security and privacy breaches in human–AI interactions and collaborations adversely affect user evaluations.

As generative AI progresses, its behavior and communicative style increasingly exhibit characteristics reminiscent of human interaction. This evolution is reshaping the dynamics between humans and AI, fostering more romanticized emotional investments in these technologies. However, it is important to

recognize that the social roles assumed by AI in expressing emotions are not designed to fulfill user needs; rather, they reflect the intentions of the developers through pre-programmed behaviors and responses. In light of this reality, it is imperative to evaluate and address the moral risks associated with various forms of AI companionship. When users perceive AI companionship as a source of genuine emotional support, "emotional bubbles can impede personal emotional development and diminish users' capacity to cultivate diverse social relationships, thereby complicating their interactions with individuals who possess differing emotional perspectives (Mlonyeni, 2025). Conversely, emotional bubbles may create an illusion of external validation, which poses a significant threat to societal moral standards. Consequently, the ethical development of emotional companionship functionalities within generative AI in the context of human–AI interactions presents a complex challenge that necessitates thorough examination and consideration.

11.3 Psychological Game in Human–AI Interaction and Collaboration

11.3.1 AI as a Human Workforce Substitute

Technological advancements and innovations in automation technology have progressively supplanted repetitive and standardized tasks that were traditionally performed by humans. Furthermore, with the emergence of deep learning, big data analytics, and other digital technologies, the phenomenon of "machine replacement" has expanded beyond low-skilled labor, exerting a substantial impact across various industries.

According to the World Economic Forum's report titled *The Future of Jobs Report 2020*, economic downturn resulting from the COVID-19 pandemic, in conjunction with rapid advancements in automation technology, is accelerating changes in the job market at an unprecedented rate. It is projected that automation and the evolving labor dynamics between humans and machines will disrupt approximately 85 million jobs across fifteen global industries within the next five years. The demand for technical positions, including data entry, accounting, and management services, has been significantly affected. However, the ongoing wave of industrial upgrading and the robust growth of digitalization have led to an increase in hiring demand within the AI, big data, and manufacturing sectors, thereby creating additional job opportunities. Nevertheless, this growth also intensifies job insecurity. Research indicates that sectors characterized by high levels of automation – such as agriculture,

forestry, animal husbandry, fishing, mining, manufacturing, and construction – are particularly vulnerable to job displacement. Older workers with lower educational attainment are at an especially high risk of being replaced (Wang et al., 2022). Furthermore, the development and rapid proliferation of generative AI and AI models present unprecedented challenges for knowledge workers, with data analysts, product managers, and other high-level professionals potentially facing threats that may surpass those encountered by manual laborers (Dăniloaia & Turturean, 2024).

Overall, traditional mechanistic perspectives frequently interpret the coexistence of machines and humans as a zero-sum game, neglecting to evaluate the broader opportunities afforded by generative AI and human–AI collaboration from a macroeconomic standpoint (Novella et al., 2023). This situation underscores the initial disparity between humans and artificial intelligence during the early phases of technological transformation.

11.3.2 AI and Team Collaboration

The swift advancement of generative AI is reshaping the parameters of human–AI interaction and collaboration, thereby influencing team dynamics across a range of task scenarios. Empirical research suggests that the capabilities of generative AI in executing various innovative tasks have exceeded those of 90–99 percent of human participants (Haase and Hanel, 2023). It is evident that AI is poised to become an essential collaborator in future work environments.

The integration of generative AI into workplace environments is poised to effect significant transformations. Conventional systems for assessing work capabilities are becoming increasingly obsolete, necessitating a redefinition of the value that individuals contribute within teams. Although attributes such as emotional intelligence and adaptability continue to be vital, proficiency in generative AI has emerged as an essential skill that stands apart from traditional competencies (Relyea et al., 2024). In the context of evaluating individual performance within a team, the capacity to enhance task quality through AI assistance is now a more critical determinant than traditional skill sets.

In addition to affecting individual performance evaluation criteria, human–AI interaction and collaboration may have negative impacts on the overall teamwork environment. On one hand, an overreliance on AI models can lead to the phenomenon known as "social loafing," focus and motivation compared to when they are working independently (Cymek et al., 2023). This is further exemplified by the concept of "automation complacency" in autonomous driving, which indicates that the presence of automation technology can lead

to distractions among workers (Li et al., 2024; Liu, 2023). Consequently, it is imperative to clearly define the boundaries of cognitive autonomy that is transferred from humans to AI within human–AI teams, and to establish appropriate interpersonal dynamics and interaction frameworks between humans and AI.

11.3.3 Security and Privacy

The swift advancement of generative AI has yielded substantial benefits across various domains, including healthcare, education, and the arts. Nevertheless, it has also engendered apprehensions regarding deepfake technology, breaches of privacy, data contamination, and the safeguarding of intellectual property. These concerns pose significant threats to user security and privacy rights. The origins of these issues can be attributed to both a deficiency in preventive awareness and the profit-driven motivations of capital, which highlight intrinsic shortcomings within the generative AI technology itself.

The advancement of generative AI and deepfake technology, which enables the production of images, audio, and video content derived from real-world materials, has raised significant concerns about misinformation and authenticity. These fabricated yet often difficult-to-detect forms of media present significant risks, including the manipulation of personal content, identity theft, and challenges associated with identity verification (Jones et al., 2018). In the digital age, personal information, such as images and life histories, can be accessed at minimal cost. Deepfake technology facilitates the generation of counterfeit content from personal photographs and audio recordings, thereby jeopardizing individual reputation and security while also enabling online fraud, extortion, and the dissemination of malicious software. Numerous identity verification methods depend on biometric data, including voice and facial characteristics, which underscores the importance of safeguarding sensitive personal information (Li et al., 2020). As detection technologies for deepfakes continue to evolve, users may experience a lack of trust in human–AI interactions, potentially hindering the acceptance and promotion of generative AI technology.

The security and privacy threats encountered by users during their interactions with generative AI are primarily attributable to the inherent limitations of the technology. Generative AI employs deep learning algorithms and sophisticated neural network models to learn from extensive datasets, with the objective of simulating human cognitive processes and providing conversational services. At its current developmental stage, generative AI lacks a genuine understanding of user needs; rather, it replicates patterns from existing

datasets (Sengar et al., 2024). This limitation can result in outputs that do not correspond with user input, exhibit biases, or contain inaccuracies. Furthermore, the challenge of accurately discerning a user's true intent complicates the regulation of generative AI outputs and exacerbates regulatory difficulties. In terms of data privacy, certain generative AI tools engage in excessive data collection, often without the user's explicit consent. Additionally, during interactions with generative AI, users may inadvertently disclose substantial amounts of personal information in pursuit of more precise and customized responses. This information may subsequently be incorporated into the AI's training data and potentially shared with other users (Kaswan et al., 2023). For instance, in March 2023, multiple Twitter users reported that ChatGPT generated content that included personal information such as names, phone numbers, and email addresses belonging to others. Although OpenAI promptly addressed this issue, it adversely affected user trust and raised significant concerns regarding data security, leading some countries and organizations to impose restrictions on the utilization of generative AI.

11.4 Countermeasures

11.4.1 Providing More Transparent and Explainable Generative AI Services

To improve the explainability of generative AI, it is crucial to tackle the challenges associated with model complexity and data uncertainty. From an algorithmic standpoint, AI algorithms frequently exhibit high complexity and encompass numerous parameters, rendering their output mechanisms similar to a "black box," which diminishes their explainability. Simplifying model architectures and parameters, as well as employing more interpretable algorithms, can facilitate the development of decision models that are more comprehensible. Examples of commonly utilized algorithms that exemplify this approach include decision trees, logistic regression, linear regression, and random forests.

Numerous companies engaged in artificial intelligence development are presently concentrating on the research of generative AI systems that facilitate user comprehension of their decision-making processes. For example, OpenAI has been actively developing technologies and tools aimed at enhancing interpretability, thereby assisting users in understanding how AI models identify user requirements and generate decisions based on these attributes (Raiaan et al., 2024). By integrating visualization technologies and adopting more

interpretable model architectures, complex decision-making models can be represented through intuitive visualizations and animations. This approach not only improves the explainability of the overall decision model but also fosters increased user trust in the decisions made by generative AI systems.

Training datasets that contain noise or inaccuracies can make the decision output mechanisms difficult to understand, affecting the explainability of generative AI systems. Introducing uncertainty assessment and robustness analysis can provide measures of the trustworthiness of generative AI outputs. Additionally, improving the selection and cleaning of training data helps offer more transparent and explainable services during human–AI interaction and collaboration.

11.4.2 Addressing the "Hallucination" Problem in Large Generative AI Models

In recent years, generative AI models such as ChatGPT have gained widespread adoption, generating considerable interest across multiple domains. These models, which are underpinned by large language models (LLMs), possess the capability to discern user intentions and generate engaging and accurate interactive content. This functionality not only enhances work efficiency but also offers emotional support in human–AI interactions and collaborations. Nevertheless, despite the high level of precision and fluency exhibited in these interactions, a notable challenge remains: the phenomenon known as hallucination is inherent to generative AI models (Filippova, 2020; Ji et al., 2023).

Large generative AI models, characterized by extensive training datasets and diverse application contexts, are capable of generating content that may seem coherent and plausible, despite being a product of hallucination. Consequently, the evaluation and mitigation of hallucinations in these models are essential, as they significantly influence user satisfaction in interactions and collaborations between humans and AI (Tonmoy et al., 2024).

To mitigate the adverse effects of model hallucinations, it is essential to prioritize the quality of training datasets during the pre-training phase. In the instruction fine-tuning phase, the implementation of manual data cleaning can effectively prevent hallucinations that arise from behavior cloning phenomena. Furthermore, during the reinforcement learning phase, the application of varying degrees of penalties for incorrect responses, contingent upon different tones and attitudes, can incentivize generative AI to recognize its errors. This strategy aids in circumventing hallucinations that stem from the overconfidence of the AI model.

11.4.3 Solutions for Addressing Privacy Leakage
in Generative AI

To tackle privacy leakage and safeguard user security, two main approaches should be reinforced:

Enhancing Data Security Measures

Automated System Controls: Enhancing data security necessitates a proactive approach to preventing privacy breaches. The implementation of automated systems can substantially reduce the risks associated with data leakage. Cutting-edge technologies such as Robotic Process Automation (RPA), low-code development platforms, process mining, and Natural Language Processing (NLP) are leading the research efforts in this domain (Haleem et al., 2021; Ng et al., 2021). Prior to the extensive adoption of generative AI, intelligent automation had already facilitated the creation of conversational processes, wherein workflows and commands are initiated through keyword instructions. By integrating generative AI-driven automation within local or cloud-based systems, it is possible to establish an intermediary isolation layer between the user and the generative AI. This strategy enhances security in comparison to cloud-based content generation, thereby offering a superior level of data protection.

Using Synthetic Data

Synthetic Data Generation: Addressing data leakage in generative AI can be effectively accomplished through the utilization of synthetic data. Techniques such as Generative Adversarial Networks (GANs), sequence models, and data anonymization are capable of producing datasets that closely resemble real personal information while omitting actual identifiable details. Synthetic data is characterized by its high quality, efficiency, and cost-effectiveness, and its artificial generation inherently provides privacy protection (Guo & Chen, 2024). The evolution of privacy regulations is further promoting the adoption of synthetic data as a vital solution. Within the realm of generative AI, it is imperative to concentrate on enhancing the quality, authenticity, interpretability, and applicability of synthetic data to develop models that more effectively satisfy user requirements.

The implementation of these strategies can effectively mitigate the risks of privacy leakage in generative artificial intelligence, thereby ensuring the security and protection of user data.

Acknowledgements

References

Ananya. (2024). AI Image Generators Often Give Racist and Sexist Results: Can They Be Fixed? *Nature*, *627*, 722–725. https://doi.org/10.1038/d41586-024-00674-9

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and Theory Gaps during the Rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100002.

Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022). Human vs. AI: Understanding the Impact of Anthropomorphism on Consumer Response to Chatbots from the Perspective of Trust and Relationship Norms. *Information Processing & Management*, *59*(3), 102940.

Cymek, D. H., Truckenbrodt, A., & Onnasch, L. (2023). Lean Back or Lean in? Exploring Social Loafing in Human–Robot Teams. *Frontiers in Robotics and AI*, *10*, 1249252.

Dăniloaia, D. F., & Turturean, E. (2024). Knowledge Workers and the Rise of Artificial Intelligence: Navigating New Challenges. *SEA: Practical Application of Science*, *12*(35).

Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lvovsky, M. (2023). ChatGPT Outperforms Humans in Emotional Awareness Evaluations. *Frontiers in Psychology*, *14*, 1199058. https://doi.org/10.3389/fpsyg.2023.1199058

Filippova, K. (2020). Controlled Hallucinations: Learning to Generate Faithfully from Noisy Data. [arXiv preprint]. arXiv:2010.05873.

Gilliard, C. (2022, January 2). Crime Prediction Keeps Society Stuck in the Past. *WIRED*. www.wired.com/story/crime-prediction-racist-history/

Guo, X., & Chen, Y. (2024). Generative AI for Synthetic Data Generation: Methods, Challenges and the Future [arXiv preprint]. arXiv:2403.04190.

Haase, J., & Hanel, P. H. (2023). Artificial Muses: Generative Artificial Intelligence Chatbots have Risen to Human-level Creativity. *Journal of Creativity*, *33*(3), 100066.

Haleem, A., Javaid, M., Singh, R. P., Rab, S., & Suman, R. (2021). Hyperautomation for the Enhancement of Automation in Industries. *Sensors International*, *2*, 100124.

Han, M. C. (2021). The Impact of Anthropomorphism on Consumers' Purchase Decision in Chatbot Commerce. *Journal of Internet Commerce*, *20*(1), 46–65.

Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2023). Does Human–AI Collaboration Lead to More Creative Art? Aesthetic Evaluation of Human-made and AI-generated Haiku Poetry. *Computers in Human Behavior*, *139*, 107502.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., . . . & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, *55*(12), 1–38.

Jenks, C. J. (2024). Communicating the Cultural Other: Trust and Bias in Generative AI and Large Language Models. *Applied Linguistics Review*, *16*(2), 787–795.

Jones, M. L., Kaufman, E., & Edenberg, E. (2018). AI and the Ethics of Automating Consent. *IEEE Security & Privacy*, *16*(3), 64–72.

Kaswan, K. S., Dhatterwal, J. S., Malik, K., & Baliyan, A. (2023, November). Generative AI: A Review on Models and Applications. In *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)* (pp. 699–704). IEEE.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural Language Processing: State of the Art, Current Trends and Challenges. *Multimed Tools Appl*, *82*, 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Königs, P. (2022). Artificial Intelligence and Responsibility Gaps: What Is the Problem? *Ethics and Information Technology*, *24*(3), 36.

Li, L., Mu, X., Li, S., & Peng, H. (2020). A Review of Face Recognition Technology. *IEEE Access*, *8*, 139110–139120.

Li, M., Guo, F., Li, Z., Ma, H., & Duffy, V. G. (2024). Interactive Effects of Users' Openness and Robot Reliability on Trust: Evidence from Psychological Intentions, Task Performance, Visual Behaviours, and Cerebral Activations. *Ergonomics*, *67*(11), 1612–1632.

Liu, P. (2023). Reflections on Automation Complacency. *International Journal of Human–Computer Interaction*, *40*(22), 7347–7363.

Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in Artificial Intelligence: From a Foundational Trust Framework to Emerging Research Opportunities. *Electronic Markets*, *32*(4), 1993–2020.

Mariadassou, S., Klesse, A. K., & Boegershausen, J. (2024). Averse to What: Consumer Aversion to Algorithmic Labels, but Not Their Outputs? *Current Opinion in Psychology*, *58*, 101839. https://doi.org/10.1016/j.copsyc.2024.101839

Mlonyeni, P. M. T. (2025). Personal AI, Deception, and the Problem of Emotional Bubbles. *AI & Society*, *40*, 1927–1938. https://doi.org/10.1007/s00146-024-01958-4

Ng, K. K., Chen, C. H., Lee, C. K., Jiao, J. R., & Yang, Z. X. (2021). A Systematic Literature Review on Intelligent Automation: Aligning Concepts from Theory, Practice, and Future Perspectives. *Advanced Engineering Informatics*, *47*, 101246.

Novella, R., Rosas-Shady, D., & Alvarado, A. (2023). Are We Nearly There Yet? New Technology Adoption and Labor Demand in Peru. *Science and Public Policy*, *50*(4), 565–578.

Peng, C., van Doorn, J., Eggers, F., & Wieringa, J. E. (2022). The Effect of Required Warmth on Consumer Acceptance of Artificial Intelligence in Service: The Moderating Role of AI–Human Collaboration. *International Journal of Information Management*, *66*, 102533.

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., ... & Azam, S. (2024). A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access*, *12*, 26839–26874.

Relyea, C., Maor, D., Durth, S., & Bouly, J. (2024, August 7). *Gen AI's Next Inflection Point: From Employee Experimentation to Organizational Transformation*. McKinsey & Company. www.mckinsey.com/capabilities/people-and-organiza

tional-performance/our-insights/gen-ais-next-inflection-point-from-employee-experimentation-to-organizational-transformation

Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2024). Generative Artificial Intelligence: A Systematic Review and Applications [arXiv preprint]. arXiv:2405.11029.

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural Bias and Cultural Alignment of Large Language Models. *PNAS Nexus*, *3*(9), 346.

Tonmoy, S. M., Zaman, S. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models [arXiv preprint]. arXiv:2401.01313.

Wang, X., Zhu, X., & Wang, Y. (2022). The Impact of Robot Application on Manufacturing Employment. *Journal of Quantitative Technology Economics*, *39*(4), 88–106.

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.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making Sense of Recommendations. *Journal of Behavioral Decision Making*, *32*(4), 403–414.

Zhu, Y., Shi, H., Hashmi, H. B. A., & Wu, Q. (2023). Bridging Artificial Intelligence-based Services and Online Impulse Buying in E-retailing Context. *Electronic Commerce Research and Applications*, *62*, 101333.