ARTICLE IN PRESS

xxx (xxxx) xxx

Contents lists available at ScienceDirect

Data and Information Management

journal homepage: www.journals.elsevier.com/data-and-information-management

Human-AI interaction research agenda: A user-centered perspective

Tingting Jiang [a,b,*], Zhumo Sun [a], Shiting Fu [a,**], Yan Lv [a]

[a] School of Information Management, Wuhan University, Wuhan, Hubei, China
[b] Center for Studies of Information Resources, Wuhan University, Wuhan, Hubei, China

A R T I C L E I N F O

Keywords:
Human-AI interaction
Human-AI collaboration
Human-AI competition
Human-AI conflict
Human-AI symbiosis

A B S T R A C T

The rapid growth of artificial intelligence (AI) has given rise to the field of Human-AI Interaction (HAII). This study meticulously reviewed the research themes, theoretical foundations, and methodological frameworks of the HAII field, aiming to construct a comprehensive overview of this field and provide robust support for future investigations. HAII research themes include human-AI collaboration, competition, conflict, and symbiosis. Theories drawn from communication, psychology, and sociology support these studies, while the employed methods include both self-reporting and observational approaches commonly utilized in user studies. It is suggested that future research should broaden its focus to encompass diverse user groups, AI roles, and tasks. Moreover, it is necessary to develop multi-disciplinary theories and integrate multi-level research methods to support the sustained development of the field. This study not only furnishes indispensable theoretical and practical insights for forthcoming research endeavors but also catalyzes the realization of a future distinguished by seamless interaction between humans and AI.

1. Introduction

Since Alan Turing posed the famous question, “Can machines think?” in 1950, a new technology, Artificial Intelligence (AI), has emerged to simulate and expand human intelligence. In the subsequent decades, AI technology has experienced rapid advancement, exerting a profound influence on diverse industries and reshaping societal structures. Due to its autonomy and anthropomorphic attributes, the interaction between human and AI is markedly distinct from traditional human-computer interaction. AI has evolved beyond being a mere tool and is gradually becoming a companion, partner, friend, and even an opponent. Scenarios such as competition and conflict, originally confined to interpersonal interactions, also manifest in Human-AI interaction. Meanwhile, the extraordinary capabilities of AI have also sparked public concerns regarding issues such as privacy breaches, algorithmic discrimination, misinformation, and digital divides (UNESCO, 2022).

Driven by AI technology, the focus of HCI work is transitioning from human interaction with non-AI computing systems to interaction with AI systems, which has given rise to an emerging field known as Human-AI Interaction (HAII) (Sun et al., 2023). Currently experiencing rapid growth, the HAII field has attracted scholars from information science, computer science, psychology, and other disciplines, leading to a proliferation of related studies. However, this field grapples with unclear concepts and inconsistent terminology, hindering the establishment of a cohesive global viewpoint. Moreover, current research is scattered across different disciplines, leading to isolated investigations within each field. This overlooks the essential need for interdisciplinary collaboration to tackle complex and long-term issues. To address these challenges, this study undertakes a comprehensive review of existing HAII research, aiming to establish a holistic view of the field. It begins by revisiting and categorizing common AI-infused systems. Subsequently, it explores cutting-edge research themes in HAII, extracting theoretical foundations and research methods from diverse disciplinary domains. Finally, insights into future research trends are presented. This study contributes to laying a solid foundation for the theoretical development and practical advancement of the field of HAII.

2. Understanding AI systems

2.1. Basic AI technologies

Machine learning (ML) is the driving force behind the development of AI. The ML process involves selecting and applying appropriate algorithms to train models to learn patterns and relationships from data,

* Corresponding author. School of Information Management, Wuhan University, Wuhan, Hubei, China.
** Corresponding author.
E-mail addresses: tij@whu.edu.cn (T. Jiang), sunzhumo@whu.edu.cn (Z. Sun), fushiting@whu.edu.cn (S. Fu), lyyyan@whu.edu.cn (Y. Lv).

https://doi.org/10.1016/j.dim.2024.100078
Received 5 February 2024; Received in revised form 2 July 2024; Accepted 14 July 2024
Available online 15 July 2024

enabling the models to make predictions on new data (Tyagi, 2019). ML algorithms can be broadly classified into supervised learning (e.g., Linear Regression, Decision Trees, and Support Vector Machines), unsupervised learning (e.g., K-Means Clustering and Hierarchical Clustering), and reinforcement learning (e.g., Q-Learning and Deep Q Networks). Deep learning (DL) is a subset of ML that utilizes artificial neural networks to learn hierarchical and intricate patterns automatically from raw data. These networks, inspired by the structure of the human brain, consist of interconnected neurons organized in layers, including multiple hidden layers. Various DL models (e.g., Convolutional Neural Networks, Recurrent Neural Networks, and Generative Adversarial Networks) have been designed for processing different types of data and made powerful tools for such complex tasks as language translation, image classification, and speech recognition (Schultz et al., 2021).

ML plays a crucial role in advancing both natural language processing (NLP) and computer vision (CV), the two essential subfields of AI. NLP techniques allow AI systems to comprehend, decipher, and generate human language, thus bridging the gap in human-AI communication. NLP often involve the analysis and extraction of meaning from text data, performing sentiment analysis and language translation, as well as producing responses that resemble those of humans (Hirschberg & Manning, 2015). Automatic speech recognition (ASR), an important component of NLP, is responsible for convert spoken language into text. This process usually comprises analyzing audio signals, recognizing phonemes and words, and generating a textual representation of the speech (Aldarmaki et al., 2022). On the other hand, computer vision techniques, such as image classification, segmentation, generation, and captioning, etc., focus on processing images, videos, and other visual data and extracting meaningful insights based on visual input. They enable AI systems to interpret visual scenes like humans' visual system does, such as detecting objects, tracking motions, recognizing human facial features, and determining human poses, which is indispensable to human-AI interaction in the physical world (Voulodimos et al., 2018).

In addition to the above-mentioned technologies, robotics, knowledge representation and reasoning, cognitive computing, and other building blocks of AI are combined and applied in various ways to create intelligent systems that exhibit human-like cognitive abilities and behaviors. The overall goal is to empower AI systems to solve complex problems, make informed decisions, and interact with humans and the world in a more natural and sophisticated manner.

AI-infused systems are built from the ground up with AI in mind, aiming to optimize and enhance the system's functionalities by leveraging the above basic AI technologies. AI infusion implies that AI is a fundamental and intrinsic component of the system (Ueno et al., 2022). There are also AI-enabled systems that incorporate AI capabilities as an augmentation to their existing functionalities.

2.2. Classification of AI systems

Two basic dimensions, presence and embodiment, can be taken into consideration in the classification of AI systems (Fig. 1). While presence refers to whether AI is presented in physical or electronic proximity to the user, embodiment refers to whether AI appears in an anthropomorphic morphology or not (Li, 2015). Chatbots, voice assistants, personalized recommenders, and virtual humans are all telepresent AI as users interact with them through desktop or mobile devices. Autonomous vehicles and service robots are typical copresent AI for being touchable in the real world. Under the embodied category, virtual humans usually have highly realistic human appearance, whereas service robots can be in different human-like forms. The rest of the major AI-infused systems are unembodied.

Chatbots are conversational agents created to simulate natural language interactions with users mainly via text. There are scripted chatbots programmed to respond to specific user inputs with predetermined responses by following a set of rules. A more advanced form, intelligent chatbots are powered mainly by ML and NLP and able to understand user intent, generate more natural, personalized, and sophisticated responses, and adapt to changing user needs. AI-powered chatbots are gaining popularity in a variety of domains, such as customer service,

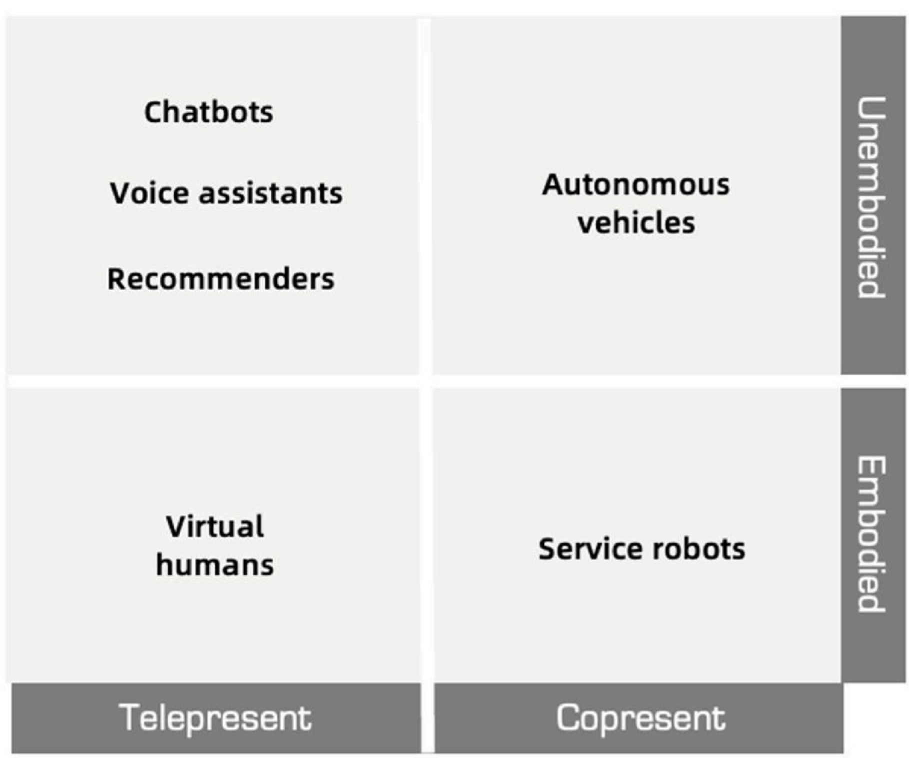

Fig. 1. The classification of AI systems.

healthcare, and education, for their effectiveness in handling user inquiries, automating routine tasks, and providing tailored recommendations, etc. Due to their limited abilities to interpret social cues or subtle nuances of human language, however, chatbots can give users an impression of being robotic and impersonal. It is important to increase social presence, i.e., the "sense of being with another", in human-chatbot interaction (Jin & Youn, 2023).

Voice assistants are virtual assistants that are able to understand and respond to voice commands and queries. They leverage on ASR to convert spoken words into text and then ML and NLP to identify the user's intent and determine what action should be taken. Amazon Alexa, Apple Siri, Microsoft Cortana, and Google Assistant are among the popular voice assistants that have already been widely embraced by consumers and businesses. They are typically integrated into mobile or wearable devices, smart speakers, in-car systems, or home automation systems, etc. and used to execute commands (e.g., controlling lights, playing music, and ordering products) or provide information (e.g., finding directions, checking the news, and searching the Web). A wake-up word is often needed to activate a voice assistant, such as "Alexa" and "Hey Siri". The future development of voice assistants will continue to focus on improving voice recognition accuracy and the understanding of context as well as reducing response time and privacy risks (Zwakman et al., 2021).

Personalized recommenders are information filters that make data-driven predictions about individual users' preferences and recommend relevant items they may like based on ML, NLP, and data mining (Isinkaye et al., 2015). They are built upon various ML algorithms, including content-based filtering and collaborative filtering. The former focuses on the similarity between items and recommends items with similar attributes to the items that a user has liked, while the latter focuses on the similarity between users and recommends items that similar users have liked to the target user. There are also hybrid recommenders that combine multiple algorithms to increase the accuracy of recommendation. It is believed that personalized recommenders may give rise to filter bubbles in which individuals are exposed to homogeneous information (Pariser, 2011). This would aggravate the negative social impact of information cocoons and echo chambers.

Virtual humans are virtual avatars that are created with the combination of ML, NLP, ASR, CV, 3D modelling, animation, and motion capture etc., to imitate human appearance and act and interact with users in a lifelike manner (Gratch et al., 2002). They are different from digital doubles, i.e., replicas of real-life people in the digital form (Domingos & Veve, 2018). Service-oriented virtual humans have emerged as virtual instructors or trainers, health consultants, tour guides, banking representatives, and shopping assistants, etc., with a superiority in engendering engaging and immersive user experience. In recent years, virtual idols, i.e., virtual characters appearing as singers or performers with distinct appearances and personalities, are becoming increasingly popular in East Asian, such as Luo Tianyi (China), Hatsune Miku (Japan), and K/DA (South Korea). With great efforts devoted to addressing the uncanny valley, a phenomenon in which people something that is almost but not fully human-like causes feelings of unease or revulsion in the observer (Mori et al., 2012), steady progress has been made towards generating realistic facial features and expressions, voices, gestures, and movements for virtual humans.

Autonomous vehicles, also known as self-driving cars, are capable of sensing their environment and navigating without human input. They depend mainly on computer vision and sensor fusion to perceive and understand the environment, and self-driving is made possible through the integration of the localization, path planning, and control modules (Mohamed et al., 2018). Human-vehicle interaction involve both in-vehicle and external interfaces. With an aim to ensure safe and comfortable driver/passenger experience, the intelligent cockpit provides an in-vehicle living space where multimodal interaction is enabled with such components as head-up displays, streaming rearview mirrors, in-vehicle voice assistants, and infotainment systems. Meanwhile, the external human-machine interfaces use visual (light-based or textual messages) or auditory cues (pure tones or spoken words) to communicate with pedestrians. For lack of trust, however, public acceptance of and consumer readiness for autonomous vehicles remain low at present (Alawadhi et al., 2020).

Unlike industrial robots that are programmed to perform simple repetitive tasks, e.g., welding and assembly, ***service robots*** (e.g., Plato, NAO, and Pepper) are embodied robots designed to interact with and provide personalized services to humans with a high degree of autonomy (Jörling et al., 2019). Computer vision and robotics engineering enable some service robots to provide labor-intensive services through particular physical capabilities, e.g., moving and carrying. For example, catering service robots are used in restaurants to take orders and serve food. In contrast, social interaction-oriented service robots are further enabled by NLP and ML to understand social cues and respond in a meaningful way, thus adept at such services as providing information, entertainment, companionship, and emotional support as well as assisting with activities in retail, education, healthcare, and other settings. With the wide adoption of service robots, there rise several major ethical concerns, including privacy, dehumanization, social deprivation, and disempowerment (Čaić et al., 2019).

3. Human-AI interaction research themes

The main research themes of human-AI interaction include human-AI collaboration, human-AI competition, human-AI conflict, and human-AI symbiosis, as shown in Fig. 2.

3.1. Human-AI collaboration

Human-AI collaboration is a joint effort of humans and AI in which a common goal is pursued. The aim is to create synergistic relationships where humans and AI collaboratively contribute to successful outcomes in various domains (Cañas Delgado, 2022). Such collaboration can leverage the strengths of both parties, combining humans' cognitive abilities, domain expertise, creative thinking, contextual understanding, and ethical judgement with AI's efficiency in identifying patterns and extracting insights from data and making data-driven predictions or recommendations, to tackle complex problems, make informed decisions, and drive innovation. Abundant evidence derived from empirical research and real-world applications has demonstrated that human-AI collaboration undeniably yields superior outcomes compared to scenarios involving only humans or AI (Kahn et al., 2020).

Human-AI collaboration has been revolutionizing medical and healthcare fields. AI can help with disease diagnosis by analyzing patient

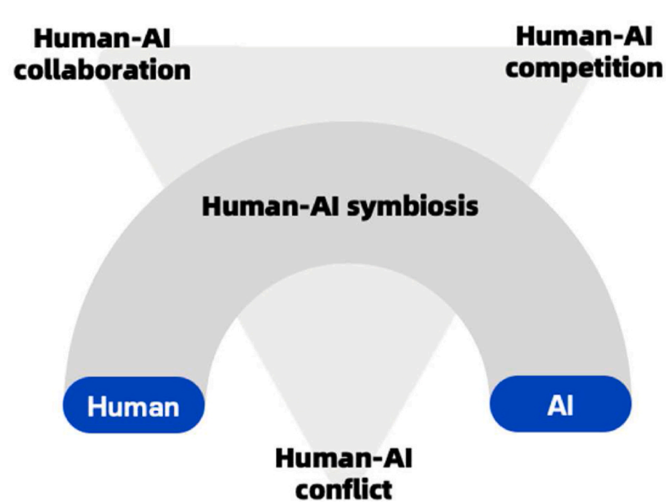

Fig. 2. Research themes of human-AI interaction.

data and support the development of personalized treatment plans. Medical imaging analysis is one the common applications in which AI algorithms can improve radiologists' accuracy and efficiency in identifying tumors, lesions, or fracturs from medical images (Rajpurkar et al., 2022). AI-enabled devices have been used in remote patient monitoring. They can collect and analyze real-time data on vital signs, symptoms, and disease progression to detect any abnormalities and trigger alerts, enabling proactive interventions (Lee et al., 2021). AI-powered robotic systems can assist surgeons during complex procedures, increasing precision and reducing errors (Lai et al., 2021). Virtual assistants and chatbots can offer patients basic symptom analysis and medical advice and direct them to appropriate healthcare resources (Roca et al., 2021). In addition, AI has the potential to enhance drug discovery and development, healthcare resource management, mental health support, and so on (D'Alfonso, 2020; Paul et al., 2021).

Modern battlefields often involve multi-domain operations and/or coalition operations, presenting unprecedented complexity and uncertainty. Military operations have become increasingly dependent on the strong computational capabilities of AI. With the collection and analysis of battlefield information, surveillance data, intelligent reports, and historical records, AI can promote situational awareness and assist commanders in strategic planning, risk assessment, operational decision-making, and resource allocation (Hung et al., 2021). Moreover, AI-powered autonomous military systems, e.g., unmanned aerial or ground vehicles, can perform tasks such as reconnaissance, surveillance, and logistics, reducing the threats to human personnel (Johnson, 2019).

Human-AI co-creation is believed to be a new strategy for creative processes with amplified creativity and increased productivity (Wu et al., 2021). AI has been introduced to a wide range of creative fields, such as painting (Oh, Bailenson, & Welch, 2018), storytelling (Zhang et al., 2021), music composition (Louie et al., 2020), fashion design (Zhao & Ma, 2018), and game design (Guzdial et al., 2019), etc. By analyzing vast amounts of existing creative works, AI can generate novel ideas and insights to inspire artists, writers, and designers or help them explore new frontiers and create original content.

In addition, it has been found that the routine work of peer reviewers (Bharti et al., 2021), teachers (Ng et al., 2020), truck drivers (Loske & Klumpp, 2021), and manufacturing workers (Mantravadi et al., 2020) can be enhanced through human-AI collaboration for either enabling humans to focus on more cognitively challenging or creative tasks or releasing them from dangerous, physically demanding, or monotone tasks.

The collaboration between humans and AI in the above scenarios varies in the level of human control and oversight. Four different modes of human-AI collaboration can be inferred, with each mode highlighting a specific range of research foci.

- ***Assisted intelligence***: humans use AI as an assistant to offer information or perform specific tasks. Existing related studies have focused on personalizing AI assistants and improving multi-modal interaction techniques to increase their effectiveness and efficiency in task automation (Islas-Cota et al., 2022; Varshan V et al., 2023).
- ***Augmented intelligence***: humans use AI as a supporter to amplify their own abilities. It has been widely investigated how to integrate AI into cognitive tasks and support decision-making and problem solving with insights, recommendations, and predictions derived from the analysis of medical, customer, or social media data (Sadiku & Musa, 2021). The recent outburst of generative AI services, e.g., ChatGPT and Midjourney, has resulted in a rapid increase in research exploring the ways to augment users in creative tasks with AI-generated content. Much attention has been attracted to prompt engineering that involves the intentional design and formulation of queries to elicit specific and desired responses from AI models (Liu & Chilton, 2022).
- ***Cooperative intelligence***: humans work with AI as a team to jointly create solutions to complex tasks. Researchers have devoted efforts to fostering alignment of mental models as well as cognitive and emotional styles, developing effective communication strategies in human-AI teams, enhancing humans' understanding of and trust in their AI teammates' decisions and actions, and seeking methods for role allocation and team performance assessment (Tabrez et al., 2020; Zhang et al., 2023).
- ***Autonomous intelligence***: AI operates independently and makes decisions without continuous human intervention, e.g., autonomous vehicles and robots. Key research topics specific to this scenario include safety and reliability, adaptivity to changing conditions in the environments, cybersecurity challenges, balance between autonomy and human control in critical situations, and liability and accountability standards, etc. (Huang et al., 2023).

3.2. Human-AI competition

Human-AI competition, in a narrow sense, refers to the contest between human and AI players in the context of game playing. IBM's Deep Blue and Google's AlphaGo are among the famous AI players that have defeated the world's top human experts in traditional tabletop games like chess, Go, and Texas hold'em. In more challenging online video games, such as Dota 2 and StarCraft II, AI players have also reached master levels with victories over most professional human players (Canaan et al., 2019).

Games have a long history of being used as AI testbeds and benchmarks. Human-AI competition in games enables AI to learn humans' strategies of thinking, deciding, and acting, which is an important approach to developing human-like AI. It also gives rise to more objectivized methods of measuring AI performance for involving a large number of referees in real decision-making situations (Świechowski, 2020). Meanwhile, humans can also benefit from playing games with AI. Stronger AI opponents can help humans improve mental capabilities and skills that have potential applicability to a variety of real-world games such as business negotiation, political campaigns, and medical treatment planning (Sandholm, 2017).

Some basic issues need to be addressed to ensure benign competition between humans and AI in gaming (Canaan et al., 2019; Świechowski, 2020): (1) ***fairness*** – given that AI players are enabled by incomparable data and computing resources, what is the fair way to compare human and AI performance on a game? (2) ***transparency*** – how can we explain AI players' highly accurate decisions and actions to humans and help humans understand the roles and limitations of AI in the game? (3) ***challenge-skill balance*** – since both unbeatable and incompetent AI players are undesirable, what level of difficulty is appropriate for a game that offers both challenges and entertainment?

There is a growing concern about human-AI competition in general, especially with regards to job opportunities. Due to its superior productivity, accuracy, availability, cost-efficiency, and learnability, AI is increasingly replacing humans in relatively simple customer service tasks, and it is also extensively used to supplement the work of professionals such as medical practitioners, lawyers, software engineers, and financial advisors (Frey & Osborne, 2017).

More and more people have viewed AI as job competitors to humans and as a potential threat to human uniqueness and control over the world. The greater the autonomy of AI, the more pronounced the perception of its threat to humans. This may lead to negative attitudes towards AI, resistance to AI research, and a rejection of services rendered by AI agents (Złotowski et al., 2017). In particular, western cultures tend to treat AI agents as pragmatic assistants, showing more ambivalent attitudes towards AI than East Asian cultures (Dang & Liu, 2022).

However, the refusal to embrace AI development is not a feasible solution to job competition between humans and AI. Humans should make better use of their invaluable expertise in creative approaches, emotional intelligence, and complex problem-solving, while AI can be leveraged for its incredible computational capabilities for pattern recognition, reasoning, prediction, and decision making. It is important to promote a healthy job market in which the full potential of both humans and AI are unlocked through their collaboration rather than competition.

3.3. Human-AI conflict

Human-AI conflict is a state of incompatibility, disagreement, or opposition between humans and AI systems (Flemisch et al., 2020). Such tensions may occur during human-AI collaboration or competition. ***Task conflict*** and ***relationship conflict*** are the two major types of conflict. The former often involves concreate issues in which resources are limited or individuals have different goals, opinions, motivations, approaches, or decisions, etc. The latter can be attributed to negative feelings as well as differences in personality, value, expectation, and style, etc. (De Dreu & Weingart, 2003).

Prior studies of human-AI conflict are mainly interested in task conflict. For examples, a human and a robot need to pass a doorway or use an elevator at the same time (Thomas & Vaughan, 2018); self-driving or autopilot systems may sometimes operate unexpectedly and/or get out of the control of human drivers or pilots, such as “phantom braking” and “automation surprise” (Wen et al., 2022); and human participants and AI proposed different solutions to a collaborative task, e.g., human-agent cooperation in desert survival (Takayama et al., 2009). When AI, often regarded as a “machine”, is adapted to perform tasks that are “proper for humans”, e.g., babysitting and hairdressing, users would show a low level of trust and have a negative expectation for the outcomes. This has been investigated as the “human-machine trans roles conflict” (Modliński et al., 2023), a special kind of relationship conflict.

Human-AI conflict is a double-edged sword. The interference of AI can prevent humans from making mistakes, e.g., dangerous driving, and vice versa. The conflict resolution process can spark engagement and innovation (Jung & Yoon, 2018). However, it is also possible that human-AI conflict leads to lower task performance, undermines humans’ perception of AI’s trustworthiness, or even does harm to humans mentally or physically (Esterwood & Robert, 2021). Hence, there is an urging need to *avoid* and/or *resolve* human-AI conflict. By making AI’s decision processes more visible, understandable, and controllable to humans, human-centered design of AI systems is crucial to reducing the occurrences of human-AI conflict. Higher AI adaptiveness is also desirable so that the level of automation and authority can be modified when potential conflict is detected.

When it comes to conflict resolution, both *submissive* and *persuasive* strategies have been widely explored. AI’s submission to humans can take the form of apology, promise, and gratitude, etc. (Esterwood & Robert, 2021), which helps repair or recover humans’ trust in AI. Robots can also use non-verbal gestures or take actions (e.g., changing path, waiting, and backing off) to show their submission (Kamezaki et al., 2020). When persuading humans, AI may leverage on explanation, appeal, and even command or threat, with an emphasis on the benefits of cooperation (Babel et al., 2022). Humor, empathy, politeness, and other verbal techniques can be employed to enhance AI’s persuasiveness, but their effectiveness would be affected by a variety of factors, such as task urgency, type of context, and robot’s appearance, etc. (Babel et al., 2021).

3.4. Human-AI symbiosis

Human-AI symbiosis is the most updated version of “man-computer symbiosis”, a concept coined in 1960 to envision the close coupling between humans and electronic computers (Licklider, 1960). Symbiosis, rather than a specific form of interaction, refers to a mutually beneficial relationship. Human-AI symbiosis emphasizes the enhancement of both humans and AI. On the one hand, humans’ information processing, problem-solving, and decision-making abilities can be augmented with AI’s computational power and analytical capabilities. On the other hand, humans can bring contextual understanding, intuition, empathy, and judgement to improve the accuracy, adaptability, and ethical sensitivity of AI. Overall, human-AI symbiosis describes a desirable future featuring harmonious collaboration and benign competition between humans and AI without being hindered by conflicts.

The low trust in and acceptance of AI among humans nowadays, however, indicates that we are still far from achieving human-AI symbiosis. A fundamental obstacle lies in the fact that many AI algorithms operate as black boxes, making it difficult for humans to understand the reasoning behind their decisions. Communication breakdowns and frustrating user experiences are prevalent in human-AI interfaces. The lack of ethical guidelines or ethical assessment has led to concerns about job displacement, loss of autonomy and control, and even adversarial attacks (Huang et al., 2023).

A novel concept, known as human-centered AI, has been put forth to address the above challenges. It emphasizes that the design, development and deployment of AI technologies and systems should focus on the needs, values, and well-being of humans, with an aim to empower individuals and promote positive outcomes for society (Shneiderman, 2022). Human-centered AI requires multidisciplinary research that combines expertise from fields such as computer science, human-computer interaction, cognitive science, psychology, sociology, and ethics, etc. As shown in Fig. 3, existing efforts have taken different approaches to actualizing human-centered AI.

- ***Human-in-the-loop***: a feedback loop of machine learning in which human input or oversight is incorporated to improve the performance of an AI system. Typical tasks featuring human involvement include training data annotation, model validation and evaluation, and algorithmic decision support (Wu et al., 2022).
- ***Explainable AI***: in order to provide insights into how AI models make decisions and generate outputs, previous research has proposed techniques to explain black-box models, i.e., uncovering the key features or factors influencing complex model’s predictions (Xu et al., 2019), as well as attempted to create white-box models, such as decision trees, linear, and rule-based models that are inherently interpretable (Shakerin & Gupta, 2020).
- ***User-friendly AI***: in order to support natural and effective communication and interaction, researchers have considered enhancing AI systems’ abilities to understand natural language and contextual information, respond to human emotions and social situations, adapt to individual users’ preferences and needs, and handle user errors and misunderstandings; and it is also useful to anthropomorphize AI when necessary through more human-like appearances, voices and speeches, conversational styles, and/or non-verbal cues (Salles et al., 2020).
- ***Responsible AI***: it has been widely recognized the importance of establishing ethical principles and legal and regulatory frameworks to engender a more trustworthy, inclusive, and beneficial AI ecosystem that prioritizes human well-being, maintains human dominance, and avoids human capacity diminution; in particular, various solutions have been proposed to address such ethical issues as data bias and fairness, privacy and security, AI divides and equality, and so on (Cheng et al., 2021).

AI literacy also plays an indispensable part in fostering human-AI symbiosis by enabling humans to understand, interact with, and effectively utilize AI systems. As shown in Fig. 4, it consists of a set of basic competencies: (1) ***knowledge of AI capabilities and limitations*** – the basic understanding what AI can and cannot do, enabling individuals to decide when and how to use AI; (2) ***effective communication and interaction with AI*** – the knowledge and skills needed to utilize AI systems’ features and interface elements, provide AI-understandable inputs, and extract meaningful insights from AI outputs; (3) ***critical assessment of AI reliability and credibility*** – the ability to evaluate the

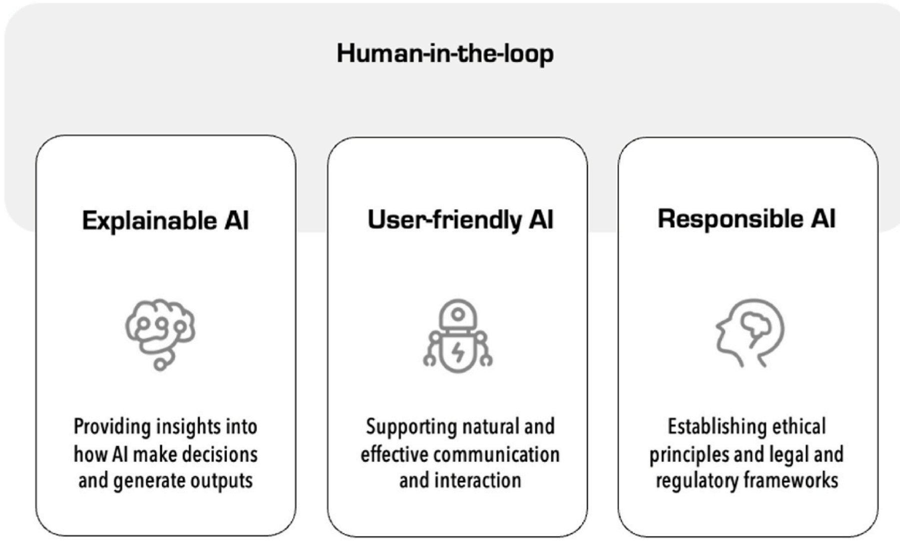

Fig. 3. Approaches to actualizing human-centered AI.

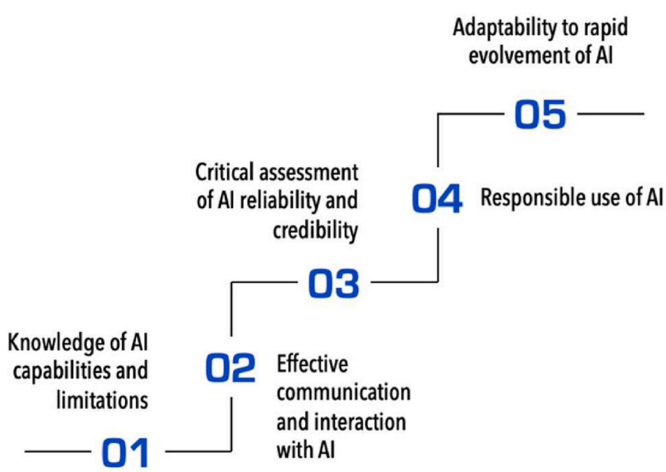

Fig. 4. Basic competencies of AI literacy.

potential benefits and risks of AI recommendations or predictions for making informed decisions and maintaining human control; (4) ***responsible use of AI*** – the awareness of ethical and legal considerations and the efforts to contribute to addressing fairness, security, equality, and other ethical challenges; and (5) ***adaptability to rapid evolvement of AI*** – the readiness to embrace new AI tools and applications, keeping pace with the changing technological landscape (Süße et al., 2021).

4. Human-AI interaction theoretical foundations

4.1. Media equation theory/Computers Are Social Actors

The Media Equation Theory (MET) suggests that humans tend to treat media like real people rather than tools (Reeves & Nass, 1996). The essence of the theory consists in the Computers Are Social Actors (CASA) paradigm, i.e., users perceiving computers as if they were social beings and possessed human-like qualities and applying social rules in their interaction with computers, such as being polite (Nass et al., 1994). The CASA paradigm has been increasingly used as a fundamental theoretical framework for understanding users' social responses to various AI systems. Users have a natural inclination to treat AI systems as social actors for their stronger abilities to process and generate information than traditional computers. Moreover, anthropomorphic AI agents are designed with rich social cues, including human-like appearances, voices, names, and even personalities (Liew & Tan, 2021). Despite the lack of genuine social cognition, AI systems can be driven by algorithms to adhere to social norms and exhibit behaviors that are consistent with social expectations in HAII (Ribino, 2023).

4.2. Social presence theory

Social presence, a concept closely related to the CASA paradigm, is a psychological state in which virtual social actors are experienced as actual social actors (Lee, 2006). The Social Presence Theory (SPT) originally explores how communication technologies vary in their abilities to convey a sense of interpersonal connection and immediacy. The SPT contends that users are motivated to choose media where they perceive a higher level of social presence because of their inherent desire for being present with others. It has been found that the perceptions of social presence can be influenced by individual, technological, and contextual characteristics (Oh, Bailenson, & Welch, 2018). Existing HAII research has investigated social presence widely as a mediating variable between various AI-related factors, such as design elements (e.g., anthropomorphism), conversational cues (e.g., responsiveness), and interaction patterns (e.g., proactivity), and users' attitudes, experiences, and behaviors (Chien et al., 2022). The SPT can provide useful insights into the underlying mechanism of HAII and informs the design and development of AI systems as effective social actors that align with human expectations.

4.3. Para-Social Relationship theory

Another related theory deserving attention is the Para-Social Relationship (PSR) theory that can be used to explain how users view their relationships with AI systems. The PSR is a term originated in the 1950s, referring to the "illusion of face-to-face relationship" that spectators develop with performers on mass media (Horton & Richard, 1956). Similarly, HAII has a bidirectional relation with PSRs in which users sense a degree of closeness and intimacy with AI systems and even consider them as friends and companions. It has been revealed for chatbots, voice assistants, and social robots that the greater their human-likeness perceived by users, the stronger the PSRs users developed with them (Tsai et al., 2021; Whang & Im, 2021). In turn, PSRs can contribute to the establishing of emotional attachment and social bonds, leading to increased user engagement, satisfaction, and attitude (Tsai et al., 2021).

4.4. Uncanny valley hypothesis

"Uncanny valley" is a concept originating in robotics and used to describe people's reactions to robots that look and act almost like humans (Mori et al., 2012). The uncanny valley hypothesis includes three stages demonstrating different patterns of change in users' emotional response. First, as robots or virtual characters become more human-like in appearance, users' affinity or emotional connection will increase correspondingly. Second, when the resemblance reaches a certain level but there still exist subtle imperfections, such anomalies can be detected by users and evoke a feeling of eeriness or discomfort, known as the "valley". Third, as the humanoid entities continue to improve and becomes highly realistic, users' positive emotions will return (Ho & MacDorman, 2010). The uncanny valley hypothesis has useful implications for designing the appearances and movements of embodied AI systems as well as the voices and speaking styles of voice assistants. According to some preliminary empirical evidence, it is important to strike a balance between human- and machine-likeness in the design of AI systems in order to enhance users' trust, acceptance, and engagement in HAII (Ciechanowski et al., 2019).

4.5. Technology Acceptance Model and related models

Based on the general framework provided by the *Theory of Reasoned Action* (TRA) and the *Theory of Planned Behavior* (TPB) for understanding human behavior, the *Technology Acceptance Model* (TAM) suggests that users are more likely to accept and adopt a new technology when they perceive it as useful and easy to use (Davis, 1989). The integration and extension of these theories and models led to the *Unified Theory of Acceptance and Use of Technology* (UTAUT) that identifies four key factors with direct influence on users' behavioral intention, i.e., performance expectancy, effort expectancy, social influence, and facilitating conditions, and considers several moderating factors, e.g., gender, age, experience, and voluntariness of use (Venkatesh et al., 2003). The subsequent UTAUT2 introduces hedonic motivation, price value, and habit as the additional constructs and expands the list of moderation factors (Venkatesh et al., 2012). Previous studies of HAII have applied the TAM, UTAUT, or UTAUT2 to predict and explain users' acceptance and adoption of various AI systems, e.g., voice assistants, chatbots, and service robots, with an aim to guide the user experience design of the systems.

However, it has been argued that the models mentioned above have limited applicability to HAII research as they focus on users' acceptance and adoption of "unintelligent functional technologies". In contrast, AI systems have the capabilities of interacting with users in a more natural manner, such as incorporating voice and gesture modalities, and performing complex cognitive processing tasks (Lu et al., 2019). Hence, Gursoy et al. (2019) proposed and tested an AI-specific model, i.e., *Artificially Intelligence Device Use Acceptance* (AIDUA). The new model features a three-stage process: (1) primary appraisal – social influence, hedonic motivation, and anthropomorphism are determinants of users' perceived performance/effort expectancy of AI systems; (2) secondary appraisal – performance/effort expectancy influences users' emotions towards the use of AI systems; and (3) outcome – positive/negative emotions predict users' acceptance/rejection of AI systems.

5. Human-AI interaction research methods

The existing human-AI interaction research has been made possible through a variety of research methods that have been frequently applied in the investigation of information behavior and user experience. These methods help researchers engender useful insights to inform the design, development, and evaluation of AI systems, leading to user-centered improvements and better understanding of human-AI interaction dynamics.

Questionnaire surveys provide a structured approach to gathering large-scale quantitative data from a wide audience and enable researchers to analyze trends and patterns across different user groups. Questionnaires can include scales that are rigorously developed and validated measurement tools for assessing single constructs or variables in a standardized manner. Surveys have been applied in human-AI interaction research to capture users' trust in and acceptance of AI systems, ethical concerns, perceptions of AI systems' usability, usefulness, and emotional impact, as well as the levels of engagement and satisfaction during human-AI interaction (Villacis Calderon et al., 2023).

Interviews are a qualitative method that enable researchers to gain rich insights and subjective perspectives through in-depth conversations with users. They provide a platform for users to articulate their needs and expectations, describe the contexts in which they interact with AI and their overall impressions of the experience, and uncover the frustrations, surprises, and other nuances of human-AI interaction. The great flexibility of interviews supports deeper probe into the underlying reasons for users' trust, acceptance, perceptions, attitudes, and concerns in relation to AI systems in general (Zhu et al., 2023).

Field studies involve observing how AI is integrated into users' daily lives, work environments, or social interactions, which ensures a higher level of ecological validity than controlled laboratory settings. In field studies, researchers can understand the contextual factors that influencing human-AI interaction, understand the practical challenges and opportunities users encounter when using AI, and even compare different user groups and assess the long-term impact of AI (Schlomann et al., 2021). Field studies are useful complement to other methods by providing valuable insights into the complexities and nuances of human-AI interaction in real-world contexts.

Experimental studies allow for the investigation of how specific factors influence user experiences in human-AI interaction, with an aim to obtain generalizable findings about the relationships between variables. In a controlled experiment, researchers may compare different AI algorithms, system features, design elements, and interaction techniques to measure their effects on users' objective performance and subjective evaluation. An experimental design often involves carefully manipulating the independent variables, randomizing participant assignments, measuring the dependent variables using various instruments and techniques, and adopting appropriate statistical tests to determine the significance of observed effects. In addition to build fully automated AI systems, previous experiments have employed the Wizard of Oz technique in which system functionalities, such as providing information, recommendations, decisions, and physical assistance, are stimulated by human operators to trigger human-AI interaction in various tasks (Riek, 2012). Physiological measurements of heart rates, electrodermal activities, eye movements, and EEG have been introduced to human-AI interaction experiments to supplement psychological measurements and behavioral observation (Zheng et al., 2019). Experimental studies contribute to evidence-based design decisions of AI design and development by providing meaningful concrete conclusions.

Usability testing is a prevalent user-centered method used to evaluate the usability and user interface of an AI system or AI-enabled product (Lam et al., 2023). In usability testing, researchers observe and collect quantitative and qualitative feedback from representative users as they perform naturalist tasks using the given system or product. The aim is to identify usability issues, understand user behavior, and gather context-specific feedback to improve the design, functionality, and user experience of the system or product. Usability testing is often conducted at different stages of an iterative design process in which human-AI interaction is continuously improved (Amershi et al., 2019).

Data-driven research, featuring the automated collection and analysis of large-scale user-centered data, is gaining popularity in the field of human-AI interaction. Chat logs are a typical type of AI system usage data stored on servers, capturing all the interactions between users and chatbots or other conversational agents as well as the messages exchanged during their conversations. Chat log analysis is useful for characterizing user behavior and evaluating chatbot performance (Gao

& Jiang, 2021). Meanwhile, a substantial volume of user comments regarding AI systems have been posted to social media or online review platforms, describing their relevant experiences, opinions, concerns, and expectations, etc. Topic modelling, sentiment analysis, and other text mining techniques have been applied to analyze both conversation messages and user reviews, enabling researchers to recognize user intentions, understand user preferences, and determine user satisfaction. The insights extracted from such user-centered data are valuable to the continuous improvement of AI systems and enhancement of user experience (Siemon et al., 2022).

6. Future trends of Human-AI interaction research

More diverse user groups. Prior HAII studies often involve general users or domain-specific users who exhibit higher willingness and abilities to utilize AI systems. As human-centered AI emphasizes the broader social impacts and ethical concerns, future research should consider various demographic or cultural user groups, especially technologically disadvantaged groups. For examples, there is a rising interest in delivering age-friendly AI services and investigating the ways in which older adults can use AI assistive technologies to enhance their independent living and social and cognitive activities; AI applications have been integrated into smart classrooms and online gaming to amplify children's enthusiasm for learning, fostering the development of their cognitive capacities and physical skills; and it is also necessary to leverage the potential of AI technologies to improve the lives of people with disabilities, such as providing automated speech recognition tools for individuals with hearing impairments. Understanding the diverse needs and preferences of all user groups helps create AI systems that offer an inclusive user experience, preventing the exacerbation of the digital divide.

More diverse AI roles. Following the traditions in HCI research, existing HAII studies have often centered around augmenting humans with AI systems to accomplish their goals, highlighting AI roles as assistants or even tools. As evidenced by the wide adoption of the CASA paradigm, however, there is a growing inclination to perceive AI systems as social actors, especially those with embodied forms. Anthropomorphic design, aiming to enhance the human-likeness of AI in terms of appearance, communication, and behavior, has appeared as a promising avenue of research. Humans are also enabled to engage in more natural multi-modal interaction with AI, such as using speech, text, touch, gestures, and even facial expressions. Moreover, the literature has witnessed a rise in efforts to create empathetic AI or emotionally intelligent systems, investigating how AI systems can recognize, understand, respond to, and influence human emotions. An increasing variety of human-AI relationships are emerging, ranging from teammates and companions to, intriguingly, romantic partners. Anticipated is a shift of attention, moving from the usability, usefulness, and ease of use of AI systems to the considerations of anthropomorphism, hedonism, and socialization.

More diverse tasks. To elicit interaction between humans and AI systems, prior studies have mostly relied on simple short single conversational or gaming tasks, probably due to the constraints of AI capabilities. The introduction of the Wizard of Oz technique, i.e., having a human operator control or simulate the AI system's responses, helped address such challenge by making users believe they are interacting with a fully autonomous AI. However, researchers still need to consider limited realism, operator bias, difficulty in simulating, and other concerns associated with this technique. The deployment of large language models has engendered new possibilities in HAII, showcasing remarkable abilities to process vast amounts of textual data and generate contextually relevant text. It is predicted that the forthcoming burst of large vision models will offer further improvement in understanding and interpreting complex visual data. As a result, future HAII research will be empowered to explore more complex or longitudinal tasks within realistic scenarios such as public services, business decisions, creative processes, and disease diagnosis and treatment, etc.

Multi-disciplinary theoretical development. The disciplines of communication and psychology have contributed the most theories to HAII research, including MET, SPT, and PSR. These theories explain how humans perceive AI and their relationships with AI, enabling researchers to explore AI systems as social beings. The uncanny valley hypothesis, also with a psychological basis, further suggests how humans feel about embodied AI. TAM and related models, derived from the information system literature, provide useful frameworks for the research designs for examining users' acceptance of AI systems as tools. The existing theoretical foundations as described in Section 4 will continue to support HAII research in the future, highlighting different concerns for developing functional AI and social AI. Future research should seamlessly integrate interdisciplinary theories. Incorporating principles from computational law and intelligent law is crucial for establishing standardized governance in areas such as privacy security, intellectual property, and the allocation of rights and responsibilities. Philosophical theories, including ontology, epistemology and ethics, can deepen human understanding of AI and facilitate the alignment of AI with human values. In addition, educational concepts such as critical thinking, creative thinking, and experiential learning theory will contribute to the development of a modern education system tailored to the age of artificial intelligence, strengthening human core competencies that are distinct from AI.

Multi-level methodological integration. Current research methods in HAII tend to be singular, and future integration should occur at multiple levels. Firstly, data collection should be more thorough. The integration of qualitative and quantitative data enables the simultaneous analysis of quantitative metrics and subjective experiences. Utilizing both large and small-scale data allows the discovery of universal patterns within large datasets, and enables nuanced analysis to uncover underlying reasons. Combining trace and response data means that researchers could capture the most natural interactions unobtrusively, and could also manipulate specific variables for real-time user reaction recording. Secondly, user observation dimensions should be more comprehensive, covering cognitive, emotional, and behavioral aspects for a holistic understanding. Lastly, tool selection should be diversified, moving beyond traditional psychological measures to incorporate behavior tracking, micro-expression capture, and neuroscientific tools such as EEG and fNIRS. This integration would capture implicit, objective responses, offering more comprehensive data support and enhancing the rigor of research conclusions.

7. Conclusion

The rapid advancement of AI technology continues to significantly impact various aspects of human life, creating extensive research opportunities for HAII. This study reviews existing HAII research, extracting four main research themes (human-AI collaboration, competition, conflict, and symbiosis) and outlining the theoretical and methodological foundations. Based on the current landscape, the study envisions future research directions. By furnishing theoretical guidance and practical insights, this study contributes not only to ensuring the sustained and robust development of the HAII field but also to the realization of the harmonious symbiosis between humans and AI.

CRediT authorship contribution statement

Tingting Jiang: Writing – review & editing, Investigation, Funding acquisition, Conceptualization. **Zhumo Sun:** Writing – original draft, Methodology, Investigation. **Shiting Fu:** Writing – review & editing, Writing – original draft, Project administration, Investigation. **Yan Lv:** Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research has been made possible through the financial support of the National Social Science Fund of China under Grant No. 22&ZD325.

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.

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 [Article].

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).

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.

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* (pp. 183–188). RO-MAN).

Flemisch, F. O., Pacaux-Lemoine, M.-P., Vanderhaegen, F., Itoh, M., Saito, Y., Herzberger, N., Wasser, J., Grislin, E., & Baltzer, M. (2020). Conflicts in human-machine systems as an intersection of bio-and technosphere: Cooperation and interaction patterns for human and machine interference and conflict resolution. In *2020 IEEE international conference on human-machine systems (ICHMS)* (pp. 1–6).

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gao, Z., & Jiang, J. (2021). Evaluating human-AI hybrid conversational systems with chatbot message suggestions. In *Proceedings of the 30th ACM international conference on information & knowledge management* (pp. 534–544). https://doi.org/10.1145/3459637.3482340

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems, 17*(4), 54–63. https://doi.org/10.1109/MIS.2002.1024753

Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management, 49*, 157–169. https://doi.org/10.1016/j.ijinfomgt.2019.03.008

Guzdial, M., Liao, N., Chen, J., Chen, S.-Y., Shah, S., Shah, V., Reno, J., Smith, G., & Riedl, M. O. (2019). Friend, collaborator, student, manager: How design of an ai-driven game level editor affects creators. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13).

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science, 349*(6245), 261–266. https://doi.org/10.1126/science.aaa8685

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior, 26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Horton, D., & Richard, W. R. (1956). Mass communication and para-social interaction. *Psychiatry, 19*(3), 215–229. https://doi.org/10.1080/00332747.1956.11023049

Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence, 4*(4), 799–819. https://doi.org/10.1109/TAI.2022.3194503

Hung, C., Choi, J., Gutstein, S., Jaswa, M., & Rexwinkle, J. (2021). Soldier-led adaptation of autonomous agents (SLA3). *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, III*, 743–754. https://doi.org/10.1117/12.2585828

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal, 16*(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005

Islas-Cota, E., Gutierrez-Garcia, J. O., Acosta, C. O., & Rodríguez, L.-F. (2022). A systematic review of intelligent assistants. *Future Generation Computer Systems, 128*, 45–62. https://doi.org/10.1016/j.future.2021.09.035

Jin, S. V., & Youn, S. (2023). Social presence and imagery processing as predictors of chatbot continuance intention in human-AI-interaction. *International Journal of Human-Computer Interaction, 39*(9), 1874–1886. https://doi.org/10.1080/10447318.2022.2129277

Johnson, J. (2019). The AI-cyber nexus: Implications for military escalation, deterrence and strategic stability. *Journal of Cyber Policy, 4*(3), 442–460. https://doi.org/10.1080/23738871.2019.1701693

Jörling, M., Böhm, R., & Paluch, S. (2019). Service robots: Drivers of perceived responsibility for service outcomes. *Journal of Service Research, 22*(4), 404–420. https://doi.org/10.1177/1094670519842334

Jung, H. S., & Yoon, H. H. (2018). Improving frontline service employees' innovative behavior using conflict management in the hospitality industry: The mediating role of engagement. *Tourism Management, 69*, 498–507. https://doi.org/10.1016/j.tourman.2018.06.035

Kahn, L. H., Savas, O., Morrison, A., Shaffer, K. A., & Zapata, L. (2020). Modelling hybrid human-artificial intelligence cooperation: A call center customer service case study. In *2020 IEEE international conference on big data (big data)* (pp. 3072–3075).

Kamezaki, M., Kobayashi, A., Yokoyama, Y., Yanagawa, H., Shrestha, M., & Sugano, S. (2020). A preliminary study of interactive navigation framework with situation-adaptive multimodal inducement: Pass-by scenario. *International Journal of Social Robotics, 12*(2), 567–588.

Lai, Y., Kankanhalli, A., & Ong, D. (2021). Human-AI collaboration in healthcare: A review and research agenda. *In the 54th Hawaii international Conference on system sciences (HICSS-54)*.

Lam, S. C. J., Ali, A., Abdalla, M., & Fine, B. (2023). U"AI" testing: User interface and usability testing of a chest X-ray AI tool in a simulated real-world workflow. *Canadian Association of Radiologists Journal, 74*(2), 314–325. https://doi.org/10.1177/08465371221131200

Lee, K. M. (2006). Presence, explicated. *Communication Theory, 14*(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A., & Bermúdez i Badia, S. (2021). A human-ai collaborative approach for clinical decision making on rehabilitation assessment. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–14).

Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies, 77*, 23–37. https://doi.org/10.1016/j.ijhcs.2015.01.001

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE transactions on human factors in electronics, HFE-1*(1), 4–11.

Liew, T. W., & Tan, S.-M. (2021). Social cues and implications for designing expert and competent artificial agents: A systematic review. *Telematics and Informatics, 65*, Article 101721. https://doi.org/10.1016/j.tele.2021.101721

Liu, V., & Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–23). https://doi.org/10.1145/3491102.3501825

Loske, D., & Klumpp, M. (2021). Intelligent and efficient? An empirical analysis of human–AI collaboration for truck drivers in retail logistics. *International Journal of Logistics Management, 32*(4), 1356–1383. https://doi.org/10.1108/IJLM-03-2020-0149

Louie, R., Coenen, A., Huang, C. Z., Terry, M., & Cai, C. J. (2020). Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–13).

Lu, L., Cai, R., & Gursoy, D. (2019). Developing and validating a service robot integration willingness scale. *International Journal of Hospitality Management, 80*, 36–51. https://doi.org/10.1016/j.ijhm.2019.01.005

Mantravadi, S., Jansson, A. D., & Møller, C. (2020). User-friendly MES interfaces: Recommendations for an AI-based chatbot assistance in industry 4.0 shop floors, 2020. In *Intelligent information and database systems* (pp. 189–201).
Modliński, A., Fortuna, P., & Rożnowski, B. (2023). Human–machine trans roles conflict in the organization: How sensitive are customers to intelligent robots replacing the human workforce? *International Journal of Consumer Studies, 47*(1), 100–117. https://doi.org/10.1111/ijcs.12811
Mohamed, A., Ren, J., El-Gindy, M., Lang, H., & Ouda, A. N. (2018). Literature survey for autonomous vehicles: Sensor fusion, computer vision, system identification and fault tolerance. *International Journal of Automation and Control, 12*(4), 555–581. https://doi.org/10.1504/IJAAC.2018.095104
Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics and Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811
Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 72–78.
Ng, F., Suh, J., & Ramos, G. (2020). Understanding and supporting knowledge decomposition for machine teaching. In *Proceedings of the 2020 ACM designing interactive systems conference* (pp. 1183–1194).
Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI, 5*, 114. https://doi.org/10.3389/frobt.2018.00114
Oh, C., Song, J., Choi, J., Kim, S., Lee, S., & Suh, B. (2018). I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–13).
Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* penguin UK.
Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discovery Today, 26*(1), 80–93. https://doi.org/10.1016/j.drudis.2020.10.010
Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine, 28*(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0
Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.
Ribino, P. (2023). The role of politeness in human–machine interactions: A systematic literature review and future perspectives. *Artificial Intelligence Review, 56*(1), 445–482. https://doi.org/10.1007/s10462-023-10540-1
Riek, L. D. (2012). Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. *J. Hum.-Robot Interact., 1*(1), 119–136. https://doi.org/10.5898/JHRI.1.1.Riek
Roca, S., Lozano, M. L., García, J., & Alesanco, Á. (2021). Validation of a virtual assistant for improving medication adherence in patients with comorbid type 2 diabetes mellitus and depressive disorder. *International Journal of Environmental Research and Public Health, 18*(22), Article 12056. https://www.mdpi.com/1660-4601/18/22/12056.
Sadiku, M. N. O., & Musa, S. M. (2021). Augmented intelligence. In *A primer on multiple intelligences* (pp. 191–199). Springer International Publishing. https://doi.org/10.1007/978-3-030-77584-1_15.
Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience, 11*(2), 88–95. https://doi.org/10.1080/21507740.2020.1740350
Sandholm, T. (2017). Super-human AI for strategic reasoning: Beating top pros in heads-up No-limit Texas hold'em. In *Proceedings of the twenty-sixth international joint conference on artificial intelligence (IJCAI-17)* (pp. 24–25).
Schlomann, A., Wahl, H.-W., Zentel, P., Heyl, V., Knapp, L., Opfermann, C., Krämer, T., & Rietz, C. (2021). Potential and pitfalls of digital voice assistants in older adults with and without intellectual disabilities: Relevance of participatory design elements and ecologically valid field studies. *Frontiers in Psychology, 12*, Article 684012. https://doi.org/10.3389/fpsyg.2021.684012
Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., & Stadtler, S. (2021). Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences, 379*(2194), Article 20200097. https://doi.org/10.1098/rsta.2020.0097
Shakerin, F., & Gupta, G. (2020). White-box induction from SVM models: Explainable AI with logic programming. *Theory and Practice of Logic Programming, 20*(5), 656–670. https://doi.org/10.1017/S1471068420000356
Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.
Siemon, D., Strohmann, T., Khosrawi-Rad, B., Vreede, T.d., Elshan, E., & Meyer, M. (2022). Why do we turn to virtual companions? A text mining analysis of replika reviews. In *AMCIS 2022 proceedings* (pp. 1–10). https://www.alexandria.unisg.ch/handle/20.500.14171/109536.
Sun, Y., Shen, X.-L., & Zhang, K. Z. K. (2023). Human-AI interaction. *Data and Information Management, 7*(3), Article 100048. https://doi.org/10.1016/j.dim.2023.100048
Süße, T., Kobert, M., & Kries, C. (2021). Antecedents of constructive human-AI collaboration: An exploration of human actors' key competencies. In *Smart and sustainable collaborative networks 4.0* (pp. 113–124).
Świechowski, M. (2020). Game AI competitions: Motivation for the imitation game-playing competition, 6-9 Sept. 2020. In *2020 15th conference on computer science and information systems (FedCSIS)* (pp. 155–160).
Tabrez, A., Luebbers, M. B., & Hayes, B. (2020). A survey of mental modeling techniques in human–robot teaming. *Current Robotics Reports, 1*(4), 259–267. https://doi.org/10.1007/s43154-020-00019-0
Takayama, L., Groom, V., & Nass, C. (2009). I'm sorry, dave: i'm afraid i won't do that: Social aspects of human-agent conflict. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2099–2108).
Thomas, J., & Vaughan, R. (2018). After you: Doorway negotiation for human-robot and robot-robot interaction. In *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 3387–3394).
Tsai, W.-H. S., Liu, Y., & Chuan, C.-H. (2021). How chatbots' social presence communication enhances consumer engagement: The mediating role of parasocial interaction and dialogue. *Journal of Research in Interactive Marketing, 15*(3), 460–482. https://doi.org/10.1108/JRIM-12-2019-0200
Tyagi, A. K. (2019). Machine learning with big data. In *Proceedings of international conference on sustainable computing in science* (pp. 1011–1020). Technology and Management (SUSCOM).
Ueno, T., Sawa, Y., Kim, Y., Urakami, J., Oura, H., & Seaborn, K. (2022). Trust in human-AI interaction: Scoping out models, measures, and methods. In *CHI conference on human factors in computing systems extended abstracts* (pp. 1–7). https://doi.org/10.1145/3491101.3519772
UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/notice?id=p::usmarcdef_0000381137.
Varshan, V. S., Rajakumaran, G., Usharani, S., & Vincent, R. (2023). A multimodal architecture with visual-level framework for virtual assistant. *International Journal of Intelligent Systems and Applications in Engineering, 11*(2), 1004–1012. https://www.ijisae.org/index.php/IJISAE/article/view/2983.
Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.
Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 157–178.
Villacis Calderon, E. D., James, T. L., & Lowry, P. B. (2023). How Facebook's newsfeed algorithm shapes childhood vaccine hesitancy: An algorithmic fairness, accountability, and transparency (FAT) perspective. *Data and Information Management, 7*(3), Article 100042. https://doi.org/10.1016/j.dim.2023.100042
Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience, 2018*, Article 7068349. https://doi.org/10.1155/2018/7068349
Wen, H., Amin, M. T., Khan, F., Ahmed, S., Imtiaz, S., & Pistikopoulos, S. (2022). A methodology to assess human-automated system conflict from safety perspective. *Computers & Chemical Engineering, 165*, Article 107939.
Whang, C., & Im, H. (2021). 'I like Your Suggestion!' the role of humanlikeness and parasocial relationship on the website versus voice shopper's perception of recommendations. *Psychology and Marketing, 38*(4), 581–595. https://doi.org/10.1002/mar.21437
Wu, Z., Ji, D., Yu, K., Zeng, X., Wu, D., & Shidujaman, M. (2021). AI creativity and the human-AI Co-creation model. In *HCII 2021: Human-computer interaction. Theory, methods and tools* (pp. 171–190).
Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems, 135*, 364–381. https://doi.org/10.1016/j.future.2022.05.014
Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Natural Language processing and Chinese computing: 8th CCF international conference, NLPCC 2019* (pp. 563–574).
Zhang, R., Duan, W., Flathmann, C., McNeese, N., Freeman, G., & Williams, A. (2023). Investigating AI teammate communication strategies and their impact in human-AI teams for effective teamwork. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW2), 1–31. https://doi.org/10.1145/3610072
Zhang, C., Yao, C., Liu, J., Zhou, Z., Zhang, W., Liu, L., Ying, F., Zhao, Y., & Wang, G. (2021). StoryDrawer: A Co-creative agent supporting children's storytelling through collaborative drawing. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1–6).
Zhao, Z., & Ma, X. (2018). A compensation method of two-stage image generation for human-AI collaborated in-situ fashion design in augmented reality environment. In *2018 IEEE international conference on artificial intelligence and virtual reality (AIVR)* (pp. 76–83).
Zheng, W. L., Liu, W., Lu, Y., Lu, B. L., & Cichocki, A. (2019). EmotionMeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics, 49*(3), 1110–1122. https://doi.org/10.1109/TCYB.2018.2797176
Zhu, H., Sallnäs Pysander, E.-L., & Söderberg, I.-L. (2023). Not transparent and incomprehensible: A qualitative user study of an AI-empowered financial advisory system. *Data and Information Management, 7*(3), Article 100041. https://doi.org/10.1016/j.dim.2023.100041
Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies, 100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008
Zwakman, D. S., Pal, D., & Arpnikanondt, C. (2021). Usability evaluation of artificial intelligence-based voice assistants: The case of amazon Alexa. *SN Computer Science, 2*(1), 28. https://doi.org/10.1007/s42979-020-00424-4