



## OPEN ACCESS

## EDITED BY

Sara Ventura,  
University of Bologna, Italy

## REVIEWED BY

Rose E. Guingrich,  
Princeton University, United States  
Rijul Chaturvedi,  
Indian Institute of Management Mumbai, India

## \*CORRESPONDENCE

Albert Łukasik  
✉ lukasik.albert@gmail.com

RECEIVED 31 January 2025

ACCEPTED 27 March 2025

PUBLISHED 09 April 2025

## CITATION

Łukasik A and Gut A (2025) From robots to chatbots: unveiling the dynamics of human-AI interaction. *Front. Psychol.* 16:1569277. doi: 10.3389/fpsyg.2025.1569277

## COPYRIGHT

© 2025 Łukasik and Gut. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# From robots to chatbots: unveiling the dynamics of human-AI interaction

Albert Łukasik<sup>1\*</sup> and Arkadiusz Gut<sup>2</sup>

<sup>1</sup>Department of Cognitive Science, Doctoral School of Social Sciences, Nicolaus Copernicus University in Toruń, Toruń, Poland, <sup>2</sup>Department of Cognitive Science, Faculty of Philosophy and Social Sciences, Nicolaus Copernicus University in Toruń, Toruń, Poland

The rapid integration of artificial agents—robots, avatars, and chatbots—into human social life necessitates a deeper understanding of human-AI interactions and their impact on social interaction. Artificial agents have become integral across various domains, including healthcare, education, and entertainment, offering enhanced efficiency, personalization, and emotional connectivity. However, their effectiveness in providing successful social interaction is influenced by various factors that impact both their reception and human responses during interaction. The present article explores how different forms of these agents influence processes essential for social interaction, such as attributing mental states and intentions and shaping emotions. The goal of this paper is to analyze the roles that artificial agents can and cannot assume in social environments, the stances humans adopt toward them, and the dynamics of human-artificial agent interactions. Key factors associated with the artificial agent's design such as physical appearance, adaptability to human behavior, user beliefs and knowledge, transparency of social cues, and the uncanny valley phenomenon have been selected as factors that significantly influence social interaction in AI contexts.

## KEYWORDS

human-robot interaction, theory of mind, social cognition, Chatbot, virtual agent

## 1 Introduction

The discourse surrounding artificial agents has undergone a radical transformation in recent years. No longer confined to roles in automobile factories, operating rooms, as opponents in chess games, or as translators, artificial agents are now being introduced as “someone” into psychotherapeutic spaces, often acting as motivators for undertaking new challenges or achieving new life goals (Bhargava et al., 2021).

By integrating artificial agents into these intimate forms of social interaction, we are witnessing claims that these entities can build relationships almost akin to those established between human therapists and their patients. Examples like Conversational AI Interfaces (CAIs)—the technology that allows computers to understand and react to human input to create a dialog—demonstrate the emerging belief in the emotional sensitivity of robots and their ability to reflect and assume the role of a coach or mentor (Sedlakova and Trachsel, 2022). This development marks a significant change in the field of artificial intelligence.

The above advancements compel us to delve deeper into why we, as humans, are beginning to treat artificial agents almost as one of our own—assigning them roles of therapists, psychologists, colleagues, and caretakers of our emotional and mental well-being. The question of what leads us to attribute mental states and a form of mental life to artificial agents is no longer purely theoretical or philosophical. It has become a central and pervasive issue concerning the status of artificial agents in our social lives. This deeper look into current

human-artificial agent interaction could be divided into three areas that are likely to become even more important in studying social interaction in the years to come:

- roles that can and cannot be ascribed to artificial agents in our social environment,
- stances that humans can take toward artificial agents,
- dynamics of the human-artificial agent interaction.

To put these claims into context, let us consider an example: social robots that are physically embodied and designed to interact and communicate with humans in a social context. They engage in activities like companionship, education, and assistance by interpreting human emotions and responding appropriately (Breazeal, 2003; Vishwakarma et al., 2024). However, these types of social interaction are not limited to situations where the artificial agent is physically present. AI-driven programs can simulate human conversation as an online chat, as is the case with conversational chatbots. They are commonly used in customer service, providing automated responses to user inquiries, assisting with troubleshooting, or guiding users through processes in a conversational manner (Shawar and Atwell, 2007; Irfan et al., 2024). However, recently they have also gained both public and academic interest as potential colleagues that someone may speak to during hard times (Sallam, 2023). We can also distinguish AI Companions which are forms of artificial intelligence designed to offer personalized interactions and companionship to users. They combine elements of virtual agents, conversational chatbots, and sometimes social robots to create a supportive and engaging experience, often aimed at enhancing well-being and providing emotional support (Fong et al., 2003). Moreover, it has been established that models like ChatGPT pass the Theory of Mind (ToM; Strachan et al., 2023; Kosinski, 2023),<sup>1</sup> as well as the Turing Test (Mei et al., 2024)<sup>2</sup> and Faux Pas Test (Shapira et al., 2023),<sup>3</sup> which serve as supporting evidence of their human-like social cognition. Passing the aforementioned mentalization tests, suggests that the social interactions with artificial agents are embedded in the competent that these agents display.

This emerging landscape highlights the need for comprehensive research into social interactions with artificial agents, particularly concerning their roles in conversation, emotional support, and therapy. It marks a departure from the traditional approaches where artificial intelligence was primarily

studied in contexts like industrial automation or informational chatbots (Yang and Hu, 2024). The integration of artificial agents into deeply personal and emotionally significant aspects of human life underscores the urgency for new research perspectives and ethical considerations in the development and deployment of AI technologies.

Firstly, it is imperative to distinguish and compare the various types of artificial agents, including robots, avatars, and conversational chatbots. By exploring this diversity, we aim to determine whether physical appearance, perceived mental features and visualization significantly impact human-AI interactions. Does the human-like appearance of an agent foster a stronger connection, or are their disembodied functions and capabilities more significant in establishing meaningful interactions?

Secondly, we must delve into the psychological reasons that lead humans to want to engage socially with artificial agents. This involves analyzing why we attribute mental states, intentions, and thoughts to entities we know are artificially intelligent—a process known as mind reading or mind perception (Gray et al., 2007; Koban and Banks, 2024). Understanding the cognitive and emotional attributes we assign to these agents can shed light on the depth of our interactions and the potential for artificial agents to fulfill roles traditionally occupied by humans.

To address these topics, we will try to answer the following questions:

- How is social interaction impacted by different forms of artificial agents (humanoid robots, virtual avatars, chatbots)?
- Does the physical appearance or visualization of an artificial agent significantly influence the process of social interaction?
- What makes humans engage in social interaction with artificial agents, attribute mental states, interpret behavior, and ascribe emotions?

Therefore, in this paper we will first review the different forms of artificial agents in existence today, discussing the common features as well as differences that might play a role in their socio-cognitive capacities. Secondly, we will examine the factors that are important for social cognition such as emotions, context, and. The last segment of the discussion will be the feeling of eeriness caused by interacting with semi-human agents known as the uncanny valley (Mori et al., 2012) and how it influence social interaction between humans and artificial agents impacting the mind attribution and accompanying emotions.

## 2 Overview of artificial agents

Artificial Intelligence (AI) has rapidly evolved, impacting healthcare, education, entertainment, and communication as AI systems like robots, chatbots, and virtual avatars enhance efficiency and personalization in daily interactions (Smith et al., 2021). Understanding their cognitive and emotional impact is crucial as these systems are widely used in social interaction. The following section will describe the current research and terminology necessary to understand what kind of social interaction is possible between humans and artificial agents. Each form of AI realizes the intended functions and goals set by engineers and programmers differently. Additionally, each form has different limitations that, from the point

1 In the context of artificial agents, the ToM test evaluates an AI system's ability to infer and respond to human mental states, such as beliefs, intentions, and emotions, simulating aspects of human social cognition.

2 The Turing Test is a benchmark for evaluating an artificial agent's ability to exhibit human-like intelligence by engaging in natural language conversations indistinguishable from those of a human. In this test, a human evaluator interacts with both an AI and a human without knowing which is which, and if the AI can convincingly mimic human responses, it is considered to have passed the test, demonstrating a form of machine intelligence.

3 A Faux Pas Test assesses an AI system's ability to recognize socially inappropriate or unintended offensive remarks in a conversation, demonstrating an understanding of social norms and implicit meanings. This test evaluates whether an AI can detect when a statement might embarrass or offend someone, requiring it to infer the mental states of the speaker and listener.

of view of human biology and psychology, can hinder the emergence of social cognition (Sandini et al., 2024).

Research shows that AI evokes various cognitive and emotional responses depending on the context and type of artificial agent (Glikson and Woolley, 2020). Chatbots in customer service influence satisfaction and engagement through sophisticated natural language processing (Crolie et al., 2022). Avatars in gaming and virtual reality provide immersive experiences, affecting user perception and interaction (Lee and Nass, 2003; Lee and Ji, 2024). Robots, from simple service bots to advanced humanoids, add complexity to human-robot interaction in healthcare, education, and homes. Human-like androids and virtual avatars elicit strong social and emotional reactions, with their physically enacted social cues influencing user responses. More human-like artificial agents evoke empathy and social bonding (Bickmore et al., 2005; Seitz, 2024). Understanding human reactions to these agents is essential for designing effective, user-friendly AI systems and mitigating negative effects like anxiety (Anisha et al., 2024). The impact of artificial agents on people varies depending on the context, modality (text-based, voice-based, embodied), and social cues, just to name a few. For instance, robots in healthcare provide physical and emotional support, while chatbots in customer service resolve issues efficiently (Bickmore and Gruber, 2010). Assessing these variables is essential to tailor AI systems to specific needs, maximizing benefits and minimizing adverse effects.

AI's current state offers both opportunities and challenges. Diverse artificial agents affect human interaction in various ways. Continued research in human-artificial agent interaction is crucial for developing beneficial, trustworthy AI systems aligned with human values (Smith et al., 2021). This research is increasingly fragmented due to new algorithms, market developments, and advanced language models like ChatGPT (Kosinski, 2023) and virtual assistants like Replika (Pentina et al., 2023). The debate on AI interaction requires an interdisciplinary approach, understanding technical capabilities, emotions, and social cognition comparisons between humans and artificial agents. It is then essential to distinguish between the various forms of artificial agents because, as it was already mentioned, each of these forms brings unique capabilities, limitations, and psychological impacts to interactions with humans, affecting everything from emotional responses to the attribution of agency and intentionality (Ziemke, 2023).

To provide an overview, artificial agents are categorized in this paper by interaction type, by which they can communicate and work with the users. Key categories include:

1. Physical interaction—robots, designed for direct physical interaction with humans or the environment. This physical contact is important in areas such as geriatric healthcare (in example assisting an older person; Robinson et al., 2014) or psychotherapeutic environment (in example robot-animals that children with special needs may hold; Cano et al., 2021).
2. Virtual interaction—autonomous avatars, existing purely in digital environments. The goal of the autonomous avatars is to assist humans in virtual environments (in example assisting through rehabilitation phase in virtual reality; Abedi et al., 2024) or engage with them to create a sense of immersion in video games (Ramadan and Ramadan, 2025). Although users cannot interact with these avatars in physical sense, by adapting the additional technology like virtual reality, they can experience a sense of social presence (Wang et al., 2024a).
3. Conversational interaction—chatbots, focused on natural language interactions through text or speech. The primary interaction happens via prompt written by a user (in the case of ChatGPT, Claude or Gemini) or voice command (in the case of Siri or Alexa). Today chatbots are used as search engines (Selvi et al., 2024), assistants for writing a code (Casheekar et al., 2024) as well as educational tutors (Akpan et al., 2025).

The categorization displayed above, although not comprehensive, is enough to point out the most general differences between artificial agents within the scope of this review. Additional factors that can be taken to further differentiate robots, avatars, and chatbots will be mentioned in the paper, based on the selected articles. The following sections will explore how these different AI forms engage in social cognition, beginning with robots and androids while indicating that the amount of research that compares various types of AI forms in social cognition is still limited.

### 3 AI design types

As each artificial agent may interact with the human in different ways, thus influencing user's reception and behavior, it is firstly worth pinpoint what are the main characteristics of each of the design type. Each of the type was designed with respect to different forms of interaction it can provide (virtual, physical, conversational) but also based on the settings it should be used like healthcare or clinical, daily assistance with information or guidance through virtual settings. What should be kept in mind is the fact that, as technology progresses, the clear boundaries between those interaction types are getting blurry. For example some of the robots like Pepper can now be supported with ChatGPT module, which allows them to speak and respond the user's voice (Bertacchini et al., 2023), while chatbots like Claude (Syamsara and Widiastuty, 2024) or Gemini (Haman et al., 2024) also provide audio-based communication instead of only text-based interaction like it was firstly intended. Similarly, avatars are also getting support to provide more immersive and natural interaction—for example virtual reality technology can create a common space and the sense of social presence (Combe et al., 2024) between humans and avatars can also speak to humans thanks to the implementation of large language models (Rao Hill and Troshani, 2024). Next three sections will provide an overall description of particular types of artificial agents, after which the paper will focus on comparing each design type in the context of social interaction highlighting differences and similarities.

#### 3.1 Robots

Robots are machines programmed to perform tasks automatically, commonly utilized in industries such as manufacturing, medicine, and exploration. Some of the forms of the robots like androids are designed to closely resemble humans in appearance and behavior, using advanced artificial intelligence to mimic human interactions (Doncieux et al., 2022). Using androids in research in social sciences enhances immersion and ecological

validity, setting a higher standard than traditional robots (MacDorman and Ishiguro, 2006). Presenting robots to various age groups and cultures allows researchers to study social development stages and cultural differences in social cognition and attitudes toward artificial intelligence (Yam et al., 2023). Increased perception of agency and emotionality toward machines can lead to positive attitudes toward AI and improved decision-making when collaborating with androids (Perez-Osorio and Wykowska, 2020). This collaboration often involves scenarios where the robot may suggest answers or agree with the human participant, intuitively granting a sense of autonomy to the artificial agent. In human-robot interaction studies, one of the goals is to replicate human traits within these mechanized agents (Smith et al., 2021). For example, robots like iCub have been used in research to examine how attitudes toward robots influence cooperation (Siri et al., 2022). These findings suggest that, after accounting for sex differences, men considered socio-emotional abilities displayed by the robot, which slowed task completion—indicating social inhibition about the robot. The iCub robot, due to its human-like traits, provides valuable insights into human cognitive and emotional processes (Marchesi et al., 2019).

Not all robots are designed to resemble humans as much as possible. The type of robot's appearance—humanoid, machine-like, or product-oriented—plays a critical role in shaping user expectations and interactions (Dautenhahn and Saunders, 2011). Humanoid robots, which closely mimic human features, like androids, tend to be perceived as more suitable for social roles, while product-oriented robots, designed primarily for functionality, excel in task-specific environments such as healthcare or service industries. This categorization affects both social engagement and user satisfaction during human-robot interaction (Kwak et al., 2014). As technology advances, the definition of “robot” continues to evolve, and classifications such as android, humanoid, mechanoid, machine-like, and zoomorphic robots, among others, offer various frameworks for differentiation. However, the appearance and functional capabilities of these robots vary greatly even within these categories, influencing how humans perceive and interact with them (Su et al., 2023). Different robot designs, from simple machine-like robots to more complex anthropomorphic designs, are used in distinct contexts like hotels, workplaces, and everyday life. The interaction varies depending on whether the robot is encountered briefly, such as a receptionist robot, or in long-term interactions like serving as a companion or co-workers in ecologically valid environments (Dautenhahn, 2018).

Robots can also be classified by their roles within the interaction, such as assistive robots that help the elderly or disabled, and social robots that engage in peer-like interactions to provide companionship or education (Kachouie et al., 2017). These differentiations highlight how the context and nature of human-robot interaction shape the design and function of artificial agents (Goodrich and Schultz, 2007). For example, in Japan, robots such as Aibo (Joshi et al., 2024), RoBoHoN (Yamazaki et al., 2023), and LOVOT (Tan et al., 2024) are integrated into people's daily routines, including personal and communal social rituals.

To recap, the design of robots significantly influences user perceptions and interaction. Humanoid robots, which exhibit a close resemblance to humans, tend to be more compatible with social roles,

while product-oriented robots are optimized for task-specific applications.

## 3.2 Avatars

Digital entities such as avatars, which exist within virtual environments, present unique opportunities for research and practical applications. Unlike physical robots, avatars as virtual agents offer extensive customization of features but lack a physical form, a crucial aspect of social engagement (Morrison et al., 2010). Research indicates that perceived warmth in virtual agents is negatively associated with fear of technology: individuals who fear technology more tend to attribute more negative emotions to virtual agents and interact with them less (Stein et al., 2020).

As digital representations of users in virtual environments, avatars can be categorized based on their visual fidelity and behavioral characteristics. These digital entities range from simplistic, cartoon-like figures to hyper-realistic humanoids, allowing researchers to manipulate the appearance and behavior of avatars for experimental purposes. Studies on the Proteus effect<sup>4</sup> have demonstrated that an individual's behavior can change due to their avatar's appearance, as more realistic avatars tend to induce behaviors aligned with social expectations (Yee and Bailenson, 2007). Avatars offer the flexibility to control factors like race, gender, and facial expressions, making them useful in studying social dynamics and identity in virtual spaces.

Autonomous avatars, powered by AI rather than human users, offer distinct opportunities for investigating human-AI interactions. In contrast to human-controlled avatars, autonomous avatars operate independently, allowing researchers to regulate social interactions within a virtual environment. These avatars are utilized to analyze perceptions of social cues, trust, and realism. For example, in educational settings, AI-driven avatars can deliver customized instruction, replicate authentic interactions, and alleviate cognitive burdens by offering contextualized learning experiences (Fink et al., 2024). In rehabilitation, AI-powered avatars are employed in virtual therapy sessions to aid in physical and cognitive exercises, providing support to patients in environments that adjust based on their progress (Veras et al., 2023). Autonomous avatars are also used in virtual worlds, such as metaverse,<sup>5</sup> to replicate lifelike social interactions, making them valuable for research and practical applications in virtual spaces (Wang et al., 2024b).

Developers can program social cues, such as facial expressions and gestures, into avatars, creating controlled experimental

<sup>4</sup> The Proteus effect explains how individuals' behaviors and attitudes align with the characteristics of their avatars in virtual environments, such as appearance or perceived traits, influencing real-world interactions and actions. For example, a person using a taller avatar in a virtual negotiation may become more assertive and confident compared to when using a shorter avatar.

<sup>5</sup> The metaverse is a virtual shared space that emerges from the merging of enhanced physical and digital realities. It serves as an interconnected and immersive digital ecosystem where users engage in real-time interactions through avatars, participating in a wide range of activities such as socializing, working, gaming, and trading. Source: <https://about.meta.com/metaverse/>.

environments within virtual spaces (Kyrilitsias and Michael-Grigoriou, 2022). Meanwhile, researchers can influence human-AI interaction in activities like joint problem-solving by manipulating variables that affect social cognition. For instance, in a study involving an ultimatum game, participants were presented with descriptions of AI opponents portrayed as emotional or rational. The results indicated that AI perceived as intentional and emotional received higher fairness ratings and elicited more generous offers (de Melo et al., 2013). The level of cooperation with avatars also hinges on team organization and whether an avatar (or NPC<sup>6</sup> in the context of video games) is viewed as a tool or a teammate. When AI was regarded as a teammate, participants displayed more emotional investment, employed optimal strategies, exchanged more strategic messages, and expressed greater confidence, trust, perceived similarity, and a sense of community compared to when AI was treated solely as a tool (Waytz et al., 2014).

Taken together, in the case of avatars, like the robots, visual design is pivotal in shaping users' perceptions of these agents, as highly anthropomorphic representations frequently elicit discomfort. Still, this effect can be easily changed by using proper software. The variability in fidelity among avatars—ranging from simplistic, cartoon-like designs to hyper-realistic humanoids—provides researchers with a unique platform to explore identity construction and social dynamics. The chosen level of reality depends on the function that the avatar should play when engaging with the human, which is the same case when it comes to their behavior. Virtual agents can perform independently and facilitate tailored interactions across diverse applications, including education, therapeutic interventions, and immersive environments such as the metaverse. Such virtual environments provide precise control over social variables, encompassing facial expressions and gestural communication, rendering them particularly suitable for investigations into human-AI interactions. In conclusion, digital agents present substantial advantages in research applications due to their inherent flexibility and capability to simulate lifelike interactions. However, their psychological and ethical effects, especially concerning user dependency and their influence on cognitive and emotional well-being (like in the case of Replika), warrant thorough examination in both their development and implementation.

### 3.3 Chatbots

Chatbots, also known as conversational agents, have become essential in various fields, including healthcare, social cognition studies, and customer service (de Cock et al., 2020). These AI-powered systems mimic human conversation and are widely utilized to handle user queries, offer assistance, and facilitate issue resolution across diverse industries. Research on chatbots has concentrated on their capacity to improve user satisfaction, trust, and engagement, while also addressing the emotional and cognitive aspects of their interactions (Ruane, 2019). As conversational agents, they utilize

machine learning<sup>7</sup> and natural language processing<sup>8</sup> to engage with users through speech or text. Their widespread presence has a significant impact on fields like computer games (Lim and Reeves, 2010; Safadi et al., 2015), healthcare (de Cock et al., 2020), and social cognition studies (Lee et al., 2021). Several notable examples of chatbots and large language models include OpenAI's ChatGPT (Dao, 2023), Google's Gemini (AlGhozali and Mukminatun, 2024), Anthropic's Claude (Berrezueta-Guzman et al., 2024) and, most recently, Le Chat<sup>9</sup> and DeepSeek.<sup>10</sup>

Chatbots might serve different operational goals, from supporting users in simple, repetitive tasks to engaging in conversation and providing guidance as well as companionship. The adaptability and broad applicability of chatbots make them indispensable tools for various sectors. Their ability to personalize interactions and evolve through learning enhances user satisfaction and broadens their potential for both practical and research-oriented applications.

## 4 Chosen aspects of social cognition in human-artificial agent interaction

To better understand the mechanisms of social cognition between humans and artificial agents, it's essential to firstly investigate what is the process of social cognition as a whole. Storbeck and Clore (2007) emphasized the deep interconnection between cognition and emotion, providing a critical lens for understanding social cognition in human-AI interaction. Their research challenged traditional views that treat cognition and emotion as separate processes and instead argued that they dynamically shape each other (Storbeck and Clore, 2007). Positive emotions can enhance cognitive flexibility and creativity, while negative emotions can sharpen focus and analytical thinking. This interplay is particularly relevant to AI interactions, where human users evaluate artificial agents both rationally and emotionally. The uncanny valley effect—a phenomenon where near-human AI elicits discomfort—can be explained through this lens. When users cognitively assess an artificial agent that appears almost but not entirely human, subtle inconsistencies may trigger negative emotional responses. This response is heightened when AI exhibits near-human appearance but lacks natural emotional expression or movement, disrupting users' expectations and leading to a sense of unease. Understanding the cognitive-emotional interaction is essential for improving AI design, ensuring that artificial agents elicit trust and engagement rather than discomfort and rejection.

Beyond cognition-emotion interdependence, Levine et al. (1993) highlight the fundamental role of social interactions in shaping cognitive processes. Their research emphasizes that cognition is not an isolated function but one deeply embedded within social contexts, where

<sup>6</sup> An NPC (non-playable character) is a character in a game or simulation controlled by the system, not the player, often serving roles like guides, adversaries, or background figures to enrich the narrative or gameplay experience.

<sup>7</sup> Machine learning empowers computational systems to acquire knowledge and enhance their performance through experience and data analysis, all while operating without the need for explicit programming directives. Source: <https://www.ibm.com/think/topics/machine-learning>.

<sup>8</sup> Natural language processing aims to help computers comprehend, interpret, and produce human language in a significant manner. Source: <https://www.ibm.com/think/topics/natural-language-processing>.

<sup>9</sup> Source: <https://chat.mistral.ai/chat>.

<sup>10</sup> Source: <https://www.deepseek.com/>.

knowledge and understanding are collectively constructed. Socially shared cognition influences learning, decision-making, and problem-solving, reinforcing the idea that intelligence is not solely an individual trait but often a collaborative process. Communication and language play a pivotal role in this shared cognition, serving as mechanisms for aligning mental models and negotiating meanings. Furthermore, motivation is closely tied to cognition, with social interactions driving cognitive engagement, attention, and information retention. Cultural frameworks and social norms further shape cognitive interpretations and expectations, impacting how people interact with others—including artificial agents. This might suggest that for artificial agents to be effective social partners, they must align with human social norms and expectations, facilitating interactions that feel natural, trustworthy, and meaningful.

A foundational study by Nass et al. (1994) demonstrated that humans instinctively apply social cognition concepts to artificial agents, treating computers and other AI-driven systems as social actors. Their research revealed that people unconsciously follow social norms, such as politeness and reciprocity, when interacting with computers. This phenomenon, later expanded into the Media Equation Theory (Reeves and Nass, 1996), established that people respond to artificial agents as they would to humans. The study also found that factors like similarity and ingroup bias influence user attitudes toward AI, aligning with existing social cognition theories. These findings were instrumental in shaping the field of human-robot interaction (HRI), providing early evidence that robots and other artificial agents could be studied within the framework of social cognition. Current studies expand on those findings while adding other key concepts and factors that influence human-artificial agent interaction in the response to fast-evolving forms of AI.

Further and one of the most recent research by Guingrich and Graziano (2024) builds upon these foundational studies by analyzing how AI features such as appearance, voice, and behavior contribute to mind perception and social interaction outcomes. Their study suggests that human-like appearances in AI, particularly in social robots and avatars, increase the likelihood of users attributing consciousness to these entities. This tendency aligns with social cognition theories that explain how humans ascribe agency to non-human entities exhibiting human-like traits (Thellman et al., 2022). Similarly, AI systems equipped with natural, human-like voices enhance perceptions of intelligence and social presence, making interactions feel more natural. Adaptive behaviors, including context-aware responses and emotional sensitivity, further reinforce the perception of AI as conscious and socially competent. Additionally, previous studies by the same authors (Guingrich and Graziano, 2023) on chatbots, particularly companion chatbots like Replika, have examined how mind perception in AI relates to social outcomes. The research indicates that users who attribute higher levels of consciousness and human-like qualities to Replika report significant social health benefits. Contrary to concerns that AI companionship might replace human interactions and negatively impact social well-being, findings suggest that users of Replika experience improved emotional support and a sense of social connection. These effects align with broader social cognition literature, which suggests that perceived agency and intentionality in AI enhance relational and emotional interactions. By fostering trust and emotional attachment, chatbots like Replika contribute positively to users' social well-being, demonstrating the expanding role of AI as a companion and support system.

Another key concepts such as trust, attachment, empathy, acceptance, and disclosure are also extensively studied in fields

regarding the social process involved in interaction with artificial agents (Hancock et al., 2020). Artificial agents designed with social cues and behaviors can evoke emotional responses and foster social bonds, increasing their acceptability and effectiveness in roles such as education, healthcare, and companionship (Belpaeme et al., 2018). Understanding how humans process and apply information about social beings is essential to social cognition research, which both informs and is informed by the development of social robotics (Wiese et al., 2017; Broadbent, 2017; Zlotowski et al., 2015). As some of the research suggests, human-like properties and attitudes toward artificial intelligence depend on three main factors: the framework, the robot's social behavior, and the interaction environment (Wallkötter et al., 2020). The framework involves personal experiences and knowledge that influence perceptions in new situations. The robot's social behavior includes human interaction patterns like nodding and commenting. The environment encompasses the study setting, whether in a laboratory or natural conditions like streets or hospitals. These factors affect perceptions of AI as intentional entities, often assessed through questionnaires like the Godspeed (Bartneck, 2023) and Mind Perception Questionnaires. Findings indicate that robots exhibiting social gestures are perceived as more social. These findings however can also be translated into avatars since they can be programmed with specific animations and responses in a virtual environment (Starke et al., 2020).

Other studies concerning the aspect of communication may suggest that this part of social cognition is better tailored in chatbots (Ali H. et al., 2024), especially the version equipped with the natural-like voice (Hwang et al., 2024) which additionally underscores choosing the proper form of AI for investigating a particular aspect of human-AI interaction. As a common point between different forms of AI, cognition, and emotion are inseparable processes in human interaction especially in social interaction. Positive emotions like happiness, trust, and safety—or negative ones like sadness, anger, and uncanniness—play critical roles, especially with service robots (Chuah and Yu, 2021). Implementing complex emotional reactions in artificial agents can benefit joint tasks, test acceptance of new technologies, and facilitate the introduction of robots and androids in healthcare settings (Indurkha, 2023) but also implementing chatbots as a part of mental health prevention (Sarkar et al., 2023).

Therefore, various forms of artificial agents—including robots, androids, avatars, and chatbots—offer both common and different variables that can be adjusted based on research hypotheses and the type of social interaction being studied (Glikson and Woolley, 2020). To give an example, some studies focus on cooperation between humans and artificial agents, while others explore competition and collaboration (Shin and Kim, 2020). There are certain themes to be most central when considering the issue of AI-human social interaction:

- embodiment—physically present, virtual, text-based agents (Memarian and Doleck, 2024),
- emotional dynamics—emotional expression manifested by the agent, reactions to the emotions manifested by the human but also hidden expressions both via gestures and the tone of the written statement (Krueger and Roberts, 2024),
- social bonds—the degree to which the user can relate to the agent (Zhang and Rau, 2023),
- expectations—the relation between predicted agent's behavior and it is actual response with the emphasis on prediction error present in uncanny valley effect (Vaitonytė et al., 2023),

- other aspects such as the adaptability of the agent's behavior, and humans' beliefs about artificial intelligence and usability (described further in the paper).

In the following subsections, each type of artificial agent is discussed, concerning the above themes.

## 4.1 Factors influencing social cognition in AI-human interaction

Designing AI capable of engaging users on emotional and cognitive levels requires consideration of a wide range of factors. For example, the impact of bodily expression—including biological versus mechanical movement, gesture presence, and movement speed—and the agent's ability to recognize emotions during interactions are significant (Hortensius et al., 2018). These factors can be applied to physical agents like robots by providing them with a proper set of joints and virtual agents like avatars with appropriate animation.<sup>11</sup> However, this will not apply to the chatbots since they do not possess any form of visual representation besides generated text or speech. This type of representation excludes the use of chatbots in some of the studies from the neuroscience field like measuring the activity of the Action Observation Network involving premotor, temporal, and medial temporal areas and the Person Perception Network involving the temporoparietal sulcus (Henschel et al., 2020), as well as studies regarding the mirror neuron system (Gazzola et al., 2007). On the other hand, the current state of chatbots allows researchers to examine neural activity during a conversation with an artificial agent as large language models become more advanced and faster in responding to humans with better accuracy (Kedia et al., 2024). But at the same time variables such as the agent's anthropomorphism, scale (big or small size), bimanual manipulation, and locomotion are also important for developing effective human-robot interactions (Kerzel et al., 2017) but cannot be applied to text or speech-based chatbots. Some of the current research showed that brain activation regarding the pragmatics is lower during human-robot interaction compared to human-human interaction because of the lack of natural human speech within robots (Torubarova et al., 2023), which creates an opportunity to replicate such studies with new forms of AI. Additionally, research on the temporo-parietal junction and its role in Theory of Mind, suggests that this region is selectively activated when individuals infer others' beliefs and intentions, distinguishing it from adjacent brain regions involved in perceiving physical characteristics of human-like entities (Saxe and Kanwisher, 2013). If the TPJ is central to how humans infer and predict others' thoughts and intentions, it raises important questions about AI design. AI systems that mimic human social behavior without genuine mental states may fail to engage the TPJ in the same way human interactions do, leading to differences in trust, acceptance, and engagement. Understanding this distinction can help refine AI

models to better align with the cognitive processes underlying human social interaction.

Investigations into whether interactions between humans and robots differ from human-human interactions in establishing social bonds during conversation have shown that human-robot interactions result in decreased activity in the fusiform gyrus, an area associated with face perception (Spatola and Chaminade, 2022). While increased activity in the posterior cingulate cortex, associated with social cognition, is observed during longer interactions with humans, no such effect is seen during interactions with robots. This suggests that robots are not considered valid social partners. Still, it also creates another opportunity to test this hypothesis in virtual avatars whose faces can be easily adjusted to the environment, role, and type of planned social interaction. It is also easier to monitor longer interactions in virtual reality compared to the lab settings with robots. Some of the studies (Mustafa et al., 2017) already compared different types of artificial agents while evaluating the N400 component using EEG but using stimuli consisting of static pictures with different levels of realism among robots, androids, and avatars with no actual interaction. Future studies focusing on the perception of faces should focus on actual interaction with setup considering robots, androids, and avatars on both screens and in virtual reality.

As already mentioned, besides faces, chatbots also lack bodies, which also excludes them from use in research investigating the perception of social-relevant stimuli like body parts and gaze cues. As the robot's head may attract the most attention the fix duration may also depend on the emotional expression (Li et al., 2022) but the studies investigating gaze and fixation toward relevant stimuli mainly focus on physically present agents while the current state of virtual reality already allows to gather eye-tracking data, making the use of virtual avatars and their embodiment (Adhanom et al., 2023). Physiological responses are already being studied when interacting with virtual agents but this interaction mainly happens through the screen (Teubner et al., 2015) rather than in a virtual reality when both humans and agents are socially present. Going further, virtual presence separates avatars and chatbots from robots and androids since the first two forms are easier to implement because of their lack of physical bodies. This creates an opportunity to investigate avatars and chatbots in settings regarding cooperation and competition in areas like video games (Possler et al., 2022), dedicated virtual reality settings (Walker et al., 2023), and simulations (Murnane et al., 2021). Cooperative and competitive tasks, although limited, can also be applied to human-robot studies. These studies usually focus on joint attention using EEG in physically common space and shared responsibility (Hinz et al., 2021) or the relation with robots in teamwork (Lefkeli et al., 2021). This interaction however offers fewer ways in which robots and androids can interact with their environment since they are limited by their movement and lack of precision in which they manipulate objects.

Although there are differences in how particular forms of AI can influence social cognition, there are also factors that seem universal for every type of artificial agent as suggested by the current research:

- adaptability to human behavior in real-time taking into account the cultural background of the user, and enhancing acceptance (Alemi et al., 2021; Hauptman et al., 2023),
- humans' beliefs and knowledge about the agent before the actual interaction, which may significantly influence their perception of its behavior (Wykowska et al., 2014; Henschel et al., 2020),
- easily interpretable and transparent social cues manifested by artificial agents (Banks, 2020; Jorge et al., 2024),

<sup>11</sup> Appropriate animation can be set thanks to the rigging system in some of the avatars. A rig is a digital skeleton used in 3D modeling to define how an avatar's structure moves, enabling the implementation of animations by controlling its joints and limbs.

- usability and behavior appropriate for its role, which may be taken from user experience studies. Although user experience studies differ from those conducted in the fields of neuroscience and psychology, these types of studies are essential to understanding how the use of an artificial agent will impact social, cognitive, and emotional elements of human-agent interaction (Weiss et al., 2009; Silva et al., 2023),
- knowledge about the source of the agent's behavior. Depending on whatever agent's behavior is accompanied by the user's input or, in the case of research studies, whatever participant is convinced that the agent is autonomous when in reality it is controlled by a human (usually referred to as the Wizard of Oz method) (de Melo et al., 2013), the human's stance and beliefs about the mental state of the agent may differ (Yu et al., 2023),
- beliefs about the moral stature and virtuous characteristics of the agent (Maninger and Shank, 2022; Bonnefon et al., 2024; Fortuna et al., 2024).

One prominent example of shared modalities is Replika, an AI companion designed for personalized social interaction. It can establish emotional bonds with users, particularly during stressful periods such as the COVID-19 pandemic (Trothen, 2022). Users often view Replika as a source of emotional support and psychological comfort, attributing human-like qualities to the avatar despite its cartoon-like appearance. Unlike traditional avatars, Replika offers a more immersive, customizable digital presence that promotes a deeper sense of connection. However, these attachments can sometimes lead to addictive behaviors and harm real-world relationships (Yuan et al., 2024). While Replika's avatar-like qualities contribute to mental health benefits, these findings raise ethical concerns regarding its potential influence on users' social, cognitive, and emotional well-being (Xie and Pentina, 2022). Replika can serve as an example of blurring boundaries between chatbots and avatars with common characteristics of both like text-based and audio-based communication, social presence (thanks to virtual reality integration) and personalization of the appearance (cosmetics and body and face changes).

Furthermore, AI companionship is increasingly being explored through the concept of Companionship Development Quality (CDQ), which defines the effectiveness of AI in fostering deep, meaningful, and lasting relationships with users (Chaturvedi et al., 2024). AI companions (ACs) are designed to integrate conversational, functional, and emotional capabilities to sustain user engagement. Conversational capabilities allow ACs to maintain natural, context-aware conversations, remembering past interactions to make discussions feel personalized. Functional capabilities enable ACs to assist users in practical tasks such as setting reminders, booking appointments, or controlling smart home devices, as seen in digital assistants like Alexa and Siri. Emotional capabilities include recognizing and responding to human emotions, facilitating social bonding, and reducing loneliness, exemplified by AI companions like Replika and Microsoft's Xiaoice. Research suggests that AI designed with only functional or emotional traits tends to lose user engagement over time, leading to interaction decline. To avoid this, AI systems must balance all three capabilities, preventing users from falling into an uncanny valley where prolonged interaction leads to discomfort or loss of trust. AI companions that successfully integrate these capabilities can foster long-term human-AI relationships, enhancing emotional support, engagement, and usability. These systems along with their capabilities can be implemented to both physical representations of agents like robots but

also to virtual entities like voice assistants which may have potential to integrate common features along different representations of AI.

Understanding the variables above is crucial for designing AI that fosters positive social interactions. However, another significant factor influencing human-AI interaction is the phenomenon known as the uncanny valley, which describes the discomfort people feel when interacting with agents that appear almost, but not entirely, human. Exploring this concept, and how it relates to different forms of artificial agents, can provide valuable insights into creating artificial agents that are both effective and comfortable for users.

## 4.2 Uncanny Valley

The uncanny valley describes the discomfort that arises when interacting with humanoid robots whose appearance closely resembles humans but falls short of full realism (Mori et al., 2012; Zhang et al., 2020). This phenomenon affects the perception of robots as sentient beings capable of feeling and decision-making. A meta-analysis of factors influencing the uncanny valley effect identified variables such as morphing faces to better match natural facial muscle movements, mismatched facial features, distorted biological movements, realism rendering, depictions of various characters, distorted or synthetic voices resembling androids, and human responses like emotions (disgust, fear) and esthetic feelings (symmetry, wrinkles; Diel et al., 2022). Designing improved AI through virtual agents, humanoid robots, and androids requires multidisciplinary collaboration among engineers, IT specialists, neuroscientists, cognitive scientists, and psychologists (MacDorman and Ishiguro, 2006). AI which is designed to serve as conversational partners, therapists, or tools to study social interactions cannot evoke any potential feeling of eeriness. For instance, androids can test theories about human interaction and brain functions in mediating communication. Failure to elicit appropriate social responses risks triggering the uncanny valley effect. This effect can be mitigated by designing AI suited to specific tasks and behaviors, such as a "nursebot" for hospital patients and the elderly. The uncanny valley effect can happen both on a psychological and neural level. Observing human-human interaction activates the left temporoparietal junction (one of the areas responsible for mentalization) more compared to observing human-robot interaction. In contrast, human-robot interactions activate the ventromedial prefrontal cortex and precuneus—areas associated with feelings of eeriness (Wang and Quadflieg, 2015).

Perceptions of robots' capacities also affect feelings of eeriness in humans. Robots perceived as capable of experience (feeling emotions) elicit stronger feelings of eeriness compared to those seen as agents or mere tools. This effect is moderated in contexts where emotional sensitivity is valued, such as nursing, reducing the eeriness of experienced robots (Stein et al., 2020). The Uncanny Valley effect can also be measured outside of laboratory settings by analyzing what people think about robots on the internet (Ratajczyk, 2022). In one of those studies, Ratajczyk and team tried to address some issues in uncanny valley studies, including inconsistent similarity assessments, a focus on visual stimuli, and challenges in evoking genuine emotions in laboratory settings. Natural language processing was used to analyze YouTube comments on robot videos in social contexts. This method captured more authentic emotional reactions, revealing that human-like robots frequently triggered terms associated with uncanniness, with human-like



robots often eliciting negative emotions. The analysis showed a relationship between facial features, sentiment, and horror, with words like “scary” and “terrifying” being most indicative of the uncanny valley effect. Interestingly, human resemblance did not correlate with pleasure or attractiveness, and smaller robots were perceived more positively, often viewed as toys. Additionally, the anticipated threat perception of larger robots was not confirmed.

Using the internet as a natural environment for studying the social perception of artificial agents also has its place in examining virtual influencers (VI). Highly anthropomorphized VIs—those with realistic human-like features—tend to elicit greater feelings of unease and uncanniness in users, potentially undermining their effectiveness as brand endorsers. This aligns with Mori’s uncanny valley theory, where near-human entities provoke discomfort due to their almost-but-not-quite-real appearance. Additionally, social cues, such as the inclusion of real human counterparts (like publicizing human-like activities like going out for a concert, or having coffee) in the VI’s content, can moderate this effect, making highly anthropomorphic VIs more acceptable to consumers by reinforcing a sense of familiarity and relatability (Creasey and Vázquez Anido, 2020; Gutuleac et al., 2024). Virtual influencers however do not exist only in the human-like form. Some of them take on the form of cartoon-like characters. Some studies suggest that in the case of cartoon-like characters, the feeling of eeriness may be lower compared to human-like influencers and cartoon-like may receive more positive reactions (indicated by the number of likes and the emotional tone of comments) compared to human-like influencers (Arsenyan and Mirowska, 2021a). This may be caused by doubt and skepticism about the human-like influencer’s authenticity, which is not present in interactions with the more stylized cartoon-like character. The level of distrust toward virtual influencers,

similar to the case of chatbots, might be reduced by providing users with knowledge about the artificial nature of the avatar (De Brito Silva et al., 2022).

The factors of Uncanny Valley are associated mainly with visual cues, while the feeling of eeriness does not have to be necessarily limited to one modality. Chatbots can be interacted with either by typing or speaking with them and as artificial agents that also can play a role in healthcare settings (Ayers et al., 2023), they should also be studied in terms of potential negative/unnerving feelings toward them. Unlike robots but similar to virtual influencers, the research regarding the feeling of eeriness caused by interaction with chatbots is fairly new, mainly because of the fairly recent and fast evolution of large language models. Studies show that the uncanny valley effect may be triggered when chatbots impersonate the real person even when being empathetic and social (Skjuve et al., 2019; Park et al., 2023) but the feeling of eeriness does not appear in the same scenario when claiming full identity disclosure (in the sense that chatbot openly claims that it works based on large language model). Studies comparing communication with speech-based chatbots and text-based chatbots are lacking since most of the studies examine voice perception already implemented in the avatars (Song and Shin, 2024; Rao Hill and Troshani, 2024). These studies suggest that users respond more positively when expressive words and prosody are balanced with the avatar’s animation rather than overly animated, suggesting that subtle emotional cues in speech are preferable for a positive user experience without inducing uncanniness (Zhu et al., 2022). Other studies indicate that users experience more discomfort, negative affect, and psychophysiological arousal, such as increased heart rate and muscle tension when interacting with the animated avatar chatbot compared to the simpler text-based version agent (Ciechanowski

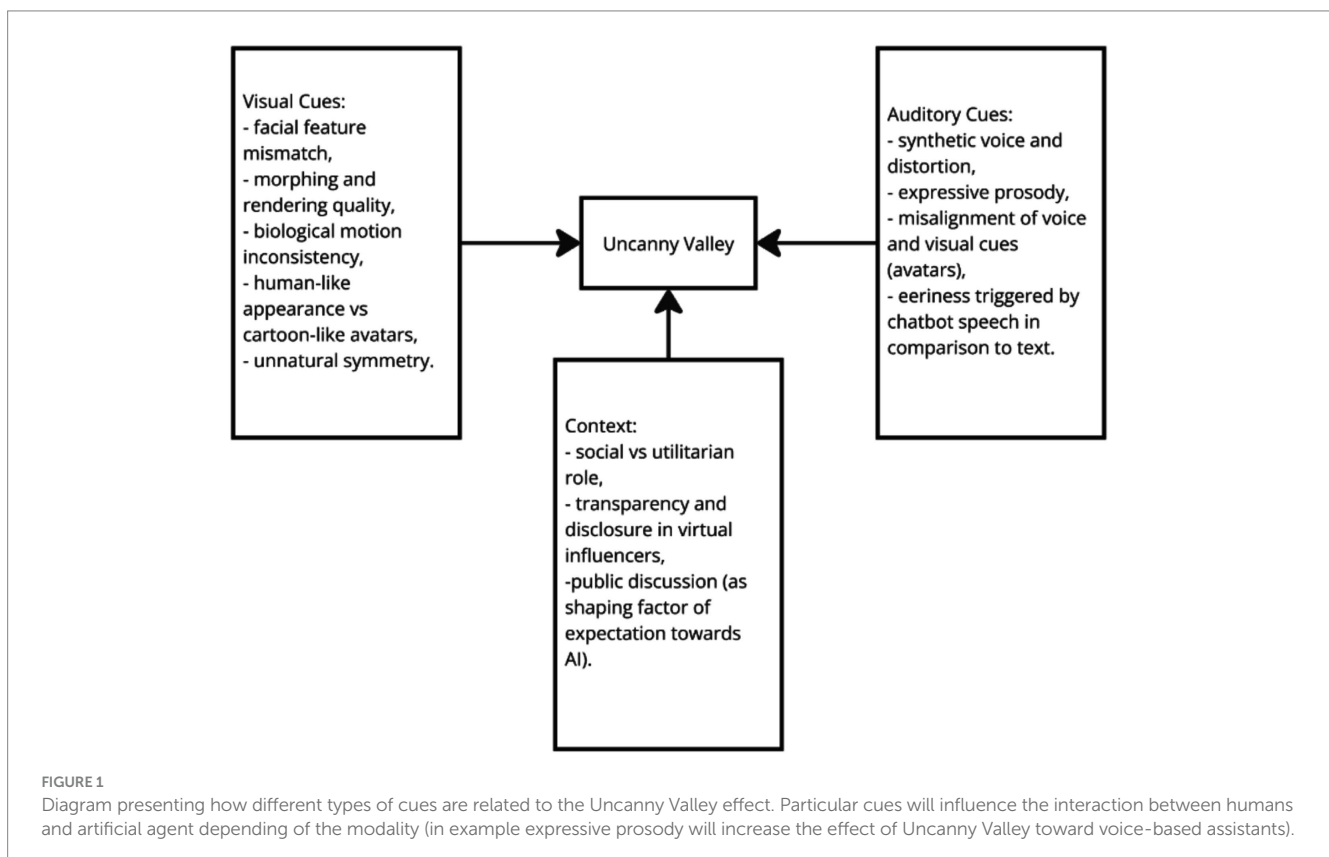


TABLE 1 Overview of selected papers investigating social factors relative to AI design type.

Authors	Type of AI	Investigated social factors	Type of interaction	Findings
Felnhofer et al. (2024)	Avatar	Social presence, agency perception, evaluation	Observing and interacting with virtual humans in immersive VR	Avatars were rated higher in social presence and evaluation than AI agents, but behavioral responses did not significantly differ. Social presence was more pronounced in neutral tasks compared to negative ones. The study suggests that higher-order responses (e.g., evaluation, presence) are influenced by perceived agency, while automatic behaviors remain unchanged.
Fraser et al. (2024)	Avatar	Perceived realism, enjoyment	Disclosing positive and negative experiences in VR	Avatars with high human resemblance and graphical resolution were perceived as the most realistic, but both cartoon and high-realism avatars were rated equally enjoyable. Standard avatars, commonly used in social VR, were rated least enjoyable, suggesting that enhancing graphical realism may improve social VR experiences.
Lim et al. (2024)	Avatar	Social influence, presence, gender matching	Conversing with a VR-embodied conversational agent (VR-ECA) about health	VR-ECAs enhanced perceived presence and social connection compared to text-based chatbots. Gender matching did not significantly impact likeability, but opposite-gender pairings increased gaze duration and slightly influenced healthy snack selection. Female participants rated VR-ECAs more favorably than male participants.
Woo et al. (2024)	Avatar	Adaptation, engagement, cognitive change	Engaging in cognitive behavior therapy (CBT) with an adaptive virtual agent	Adaptive virtual agents that adjust facial expressions and head movements based on users' behavior enhanced engagement and effectiveness in cognitive behavior therapy. Users perceived adaptive agents as more human-like and reported greater cognitive change and anxiety reduction. However, non-adaptive agents or those with mismatched behaviors negatively impacted user experience.
Arsenyan and Mirowska (2021b)	Avatar	Uncanny Valley, social media engagement, authenticity	Analysis of human reactions to virtual influencers on Instagram	Human-like virtual influencers received significantly fewer positive reactions compared to human and anime-like influencers, supporting the Uncanny Valley hypothesis. The study highlights authenticity concerns and social identity effects when interacting with virtual agents in publicly visible online networks.
Shin (2020)	Avatar	Intimacy, emotional engagement, social media interaction	Analysis of user interactions with virtual agents on Instagram	Users are more likely to engage with virtual agents when they express emotions in their posts. Emotional expression and relationships between virtual agents attract higher numbers of likes and comments. The study highlights the importance of emotional cues in fostering social engagement with AI entities on social media.
Von Der Pütten and Krämer (2010)	Avatar	Social presence, behavioral realism, agency perception	Observing and interacting with virtual agents and avatars	Participants' beliefs about interacting with an avatar or an agent had minimal influence on their social responses, but higher behavioral realism significantly increased perceived presence and engagement. The findings support the Ethopoeia concept, suggesting that social cues, rather than perceived agency, drive human social responses to AI.
Figuroa-Torres (2025)	Chatbot	Social dimensions of chatbot technology	Theoretical analysis of chatbot roles in science, commerce, and personal life	Chatbots function across three dimensions: as scientific objects, commercial commodities, and agents of intimate interaction. Their roles extend beyond mere functionality, shaping and being shaped by society. The study emphasizes the importance of understanding chatbot technology through a sociotechnical lens rather than a purely technological progression.
Ali F. et al. (2024)	Chatbot	Social anxiety, fear of rejection, compulsive chat	Frequent interaction with a social chatbot (Xiaoice)	Socially anxious individuals with a fear of negative evaluation and rejection are more likely to engage in compulsive chatbot interactions. Fear of unavailability of human social connections further strengthens this behavior. The study highlights how social chatbots may serve as coping mechanisms for anxiety but also risk fostering dependency.
Bialkova (2024)	Chatbot	Functionality, interactivity, enjoyment, satisfaction	Survey-based evaluation of chatbot user experience	Information quality, accuracy, and competence were key factors in chatbot functionality. Personal care and social presence enhanced user enjoyment. Poor chatbot performance in these areas resulted in low satisfaction, highlighting the need for optimized chatbot design to meet user expectations.
Yu and Zhao (2024)	Chatbot	Perceived warmth, competence, service satisfaction	Experimental study on emoji usage in chatbot interactions	Emojis enhance the perceived warmth of chatbots, increasing service satisfaction, but do not improve perceptions of competence. The effect is stronger for hedonic chatbots and pre-programmed bots compared to highly autonomous ones. The study highlights the role of emojis as social cues in chatbot communication.

(Continued)

TABLE 1 (Continued)

Authors	Type of AI	Investigated social factors	Type of interaction	Findings
Pan et al. (2024)	Chatbot	Uncertainty, emotional attachment, relational dynamics	Analysis of user discussions in an online chatbot community	Users experience four key uncertainties when forming relationships with social chatbots: technical, relational, ontological, and sexual uncertainty. Relational uncertainty was the most common, often leading to emotional attachment and mixed feelings about AI companionship. Some users embraced unpredictability, while others felt discomfort or confusion.
Kang and Kang (2024)	Chatbot	Self-disclosure, companionship, anthropomorphism	Engaging in a counseling session with a chatbot	The chatbot's anthropomorphic features, including gender, personality, and visual interface cues, influenced user self-disclosure and companionship. Users disclosed less when the chatbot had a visual interface cue, especially when it had an introverted personality. Participants felt a stronger companionship with chatbots of the opposite gender. The study highlights the importance of tailoring chatbot design to user characteristics.
Rheu et al. (2024)	Chatbot	Expectancy violation, trust, emotional validation	Engaging in a support-focused conversation with a chatbot	Chatbots that provided contingent, personalized feedback were evaluated more positively than those that gave generic responses. When an "expert" chatbot failed to provide contingent feedback, it led to more negative evaluations (negative expectancy violation). However, a non-expert chatbot that exceeded expectations by offering contingent feedback was evaluated favorably (positive expectancy violation). The study highlights the importance of meeting user expectations in chatbot interactions.
Pekçetin et al. (2024)	Robot	Mind perception, agency, experience attribution, generational differences	Observing live human and robot actors perform communicative and non-communicative actions	Real-time implicit and explicit measurements revealed that people attribute higher agency and experience to humans than robots. Communicative actions increased mind perception more than non-communicative actions. Generational differences influenced responses, with younger participants attributing greater mental states to robots. Implicit and explicit results varied, suggesting different cognitive mechanisms behind mind perception in HRI.
Kamino et al. (2024)	Robot	Social bonding, interaction rituals, cultural integration	Ethnographic study of social robot communities in Japan	The study explores how users integrate robots into their social lives through recurring interaction rituals, such as meetups, co-ownership events, and daily routines. Companies facilitate social bonding by organizing events and designing robots with customizable features that promote user attachment. The findings highlight that robot sociality is not just a product of design but is actively constructed through human interactions in social networks and communities.
Ghiglino et al. (2023)	Robot	Intentional stance, social bonding, interaction variability	Engaging with the iCub robot in different interaction scenarios	The study analyzed how different levels of interaction with the iCub robot influenced participants' attribution of intentional states to the robot. When the robot exhibited highly human-like, contingent behaviors, participants were more likely to adopt a mentalistic stance, interpreting its actions as intentional. However, in low-interactivity scenarios, participants tended to maintain a mechanistic perspective. The study suggests that deeper social engagement with robots enhances perceived intentionality.
Karaduman et al. (2023)	Robot	Empathy, pain perception, emotional vs. physical pain	Watching videos of humans and robots experiencing pain and rating perceived intensity	Participants attributed significantly more pain to humans than to robots in both physical and emotional scenarios. Emotional pain ratings varied depending on whether the pain source was an object or a person, whereas physical pain ratings were stable across conditions. The study highlights a persistent gap in empathy toward non-biological agents.
Spisak and Indurkha (2023)	Robot	Social exclusion, trust, team dynamics	Cooperating with the Nao robot in a bomb defusal task	The study examined social exclusion in human-robot teams. When the robot favored one participant over another, the discriminated participant reported a stronger sense of exclusion but did not significantly change their mood or attitude toward the robot. The findings highlight the potential social implications of biased robot behavior in group interactions.
Churamani and Howard (2022)	Robot	Affective learning, adaptability, negotiation strategies	Interacting with a social robot (NICO) in a negotiation game	The study introduces an affect-driven learning framework for robots, where the NICO robot adapts its negotiation strategy based on users' affective responses. Results show that robots with patient and high-arousal affective cores negotiate longer and retain persistence, whereas those with impatient and low-arousal dispositions are perceived as more generous and altruistic. The findings highlight the importance of affective appraisal in human-robot interaction and adaptive behavior learning.
Jacobs and Turner (2022)	Robot	Mind perception, agency, experience attribution	Rating agency and experience of real and fictional robots	The study found significant variation in how people attribute agency (capacity to act) and experience (capacity to feel) to different robots. While robots were rated lower than humans, some real robots, like Sophia and Atlas, received higher attributions of experience than digital assistants like Siri or Alexa. Younger participants attributed higher levels of agency and experience to robots, suggesting a generational shift in AI perception.

et al., 2018) which creates a problem to establish whatever this effect is observed because of the voice features compared to the face features.

The smaller portion of the studies, focusing mostly on the speech, suggest that emotionally expressive prosody—such as varied pitch and enthusiastic interjections—significantly enhances user engagement and perceived human-likeness, but can also trigger discomfort when overly human-like traits lead to an uncanny valley effect (Krauter, 2024). Krauter conducted an extensive analysis regarding the factors associated with an uncanny valley in chatbots including the aforementioned expressive prosody. His work however can also be adjusted to other virtual agents like robots and avatars while trying design studies that compare these forms between the same social tasks. This will allow future researchers to establish what elements of uncanniness are present in particular forms of AI and how they can be adjusted to meet social cognition needs (Figure 1).

The uncanny valley affects how people perceive artificial agents' capacities for emotions and decision-making, especially in contexts where emotional sensitivity is valued, such as healthcare. Studies mentioned above indicate that human-robot interaction activates brain areas associated with eeriness, while human-human interaction engages regions linked to natural social processes. Robots perceived as capable of experiencing emotions elicit stronger feelings of eeriness, though this is reduced in emotionally sensitive roles. For virtual influencers, human-like designs may evoke skepticism about authenticity, whereas cartoon-like designs generate higher engagement and positive affect. This pattern suggests that anthropomorphism should be balanced with clear identity disclosure to reduce unease. Similarly, chatbots may induce uncanny valley effects when mimicking humans, particularly in speech-based interactions, though this can be mitigated by openly acknowledging their AI nature. Subtle emotional cues in speech, rather than overly human-like traits, enhance user experiences without triggering discomfort. Addressing these factors to the proper form of the agent will help design artificial agents that balance human likeness with user comfort, fostering positive social interactions while meeting social cognition needs. One of the possible solutions to reduce the feeling of eeriness and make the interaction between humans and AI more social and natural might be associated with a more personalized approach in designing robots, avatars, or chatbots. This will be the topic of the next section.

## 5 Conclusion

The growing study of human-artificial agent interaction underscores the increasing significance of AI design in shaping social experiences and cognitive processes. As artificial agents such as humanoid robots, virtual avatars, and chatbots continue integrating into social, therapeutic, and professional environments, their design and behavioral adaptability profoundly influence human perception, engagement, and emotional connection. We presented the mechanisms underlying these interactions, demonstrating that the form, functionality, and perceived mental capacities of artificial agents directly impact the depth and quality of human-AI relationships.

From the design perspective, human-artificial agent interaction is shaped by an artificial agent's embodiment, expressiveness, and perceived autonomy. Humanoid robots benefit from physical presence and non-verbal cues but risk triggering the uncanny valley effect. Virtual avatars offer flexible social representation in digital environments but lack the nuances of face-to-face interaction. Chatbots, engaging primarily through language, enhance accessibility yet lack physical expressiveness.

Despite these limitations, conversational AI continues to improve in eliciting empathy and fostering engagement while also synergizing with both robots and avatars.

Physical appearance and visualization significantly influence human attribution of mental states to AI. More human-like agents enhance mind attribution, trust, and social presence, affecting whether they are seen as tools, companions, or social peers. Engagement is driven by social and emotional mechanisms, such as trust and expectation alignment, reinforcing AI as a social entity.

The uncanny valley remains a challenge in AI design, where human-like features must be balanced to avoid discomfort. Avatars and chatbots, with more controlled anthropomorphism, integrate more seamlessly into social settings without triggering unease. Advancing AI social design will require interdisciplinary collaboration to foster meaningful, trustworthy, and emotionally intelligent interactions.

Future research must continue to examine the ethical and psychological implications of human-AI interaction, particularly in contexts where artificial agents serve roles traditionally reserved for human counterparts. Furthermore, interdisciplinary efforts involving cognitive science, psychology, robotics, and ethics are necessary to develop AI that is not only technologically proficient but also socially attuned to human expectations and needs.

In summary, the design of artificial agents plays a foundational role in shaping human-AI social interactions. By carefully considering embodiment, appearance, behavioral transparency, and adaptability, developers can create AI systems that foster trust, social connection, and emotional engagement. As AI technology advances, the key to successful human-AI interaction will lie in crafting agents that align with human social and emotional processes while respecting the boundaries of what is natural and what is artificial. This ongoing evolution demands a nuanced understanding of both technological innovation and the fundamental principles governing human social behavior (Table 1).

## Author contributions

AL: Conceptualization, Investigation, Validation, Writing – original draft, Writing – review & editing. AG: Supervision, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Abedi, A., Colella, T. J., Pakosh, M., and Khan, S. S. (2024). Artificial intelligence-driven virtual rehabilitation for people living in the community: a scoping review. *NPJ Digital Med.* 7:25. doi: 10.1038/s41746-024-00998-w
- Adhanom, I. B., Mac Neilage, P., and Folmer, E. (2023). Eye tracking in virtual reality: a broad review of applications and challenges. *Virtual Reality* 27, 1481–1505. doi: 10.1007/s10055-022-00738-z
- Akpan, I. J., Kobara, Y. M., Owolabi, J., Akpan, A. A., and Offodile, O. F. (2025). Conversational and generative artificial intelligence and human–chatbot interaction in education and research. *Int. Trans. Oper. Res.* 32, 1251–1281. doi: 10.1111/itor.13522
- Alemi, M. Islamic Azad University (IAU) West Tehran Branch, and Abdollahi, A. Alzahra University (2021). A Cross-cultural investigation on attitudes towards social robots: Iranian and Chinese university students. *J. Higher Educ. Policy Leadership Stud.* 2, 120–138. doi: 10.52547/johepal.2.3.120
- AlGhozali, S., and Mukminatun, S. (2024). Natural language processing of Gemini artificial intelligence powered Chatbot. *Balangkas: Int. Multidis. Res. J.* 1, 41–48.
- Ali, H., Mahmood, R., and Zhang, L. (2024). Social chatbot: my friend in my distress. *AI & Human Relations Rev.* 40, 1702–1712. doi: 10.1080/10447318.2022.2150745
- Ali, F., Zhang, Q., Tauni, M. Z., and Shahzad, K. (2024). Social chatbot: my friend in my distress. *Int. J. Human-Computer Interaction* 40, 1702–1712. doi: 10.1080/10447318.2022.2150745
- Anisha, S. A., Sen, A., and Bain, C. (2024). Evaluating the potential and pitfalls of AI-powered conversational agents as humanlike virtual health carers in the remote management of noncommunicable diseases: scoping review. *J. Med. Internet Res.* 26:e56114. doi: 10.2196/56114
- Arsenyan, J., and Mirowska, A. (2021a). Almost human? A comparative case study on the social media presence of virtual influencers. *Int. J. Human-Computer Stud.* 155:102694. doi: 10.1016/j.ijhcs.2021.102694
- Arsenyan, J., and Mirowska, M. (2021b). Perceived authenticity and social engagement with virtual influencers. *Digital Market. Society* 28, 45–68. doi: 10.1007/s13278-022-00966-w
- Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., et al. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern. Med.* 183, 589–596. doi: 10.1001/jamainternmed.2023.1838
- Banks, J. (2020). Theory of mind in social robots: replication of five established human tests. *Int. J. Soc. Robot.* 12, 403–414. doi: 10.1007/s12369-019-00588-x
- Bartneck, C. (2023). Godspeed questionnaire series: Translations and usage. In *International handbook of behavioral health assessment* (pp. 1–35). Cham: Springer International Publishing.
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Science. Robotics* 3:eaat5954. doi: 10.1126/scirobotics.aat5954
- Berrezueta-Guzman, S., Kandil, M., Martín-Ruiz, M. L., de la Cruz, I. P., and Krusche, S. (2024). “Exploring the efficacy of robotic assistants with ChatGPT and Claude in enhancing ADHD therapy: innovating treatment paradigms.” In *2024 International conference on intelligent environments (IE)* (pp. 25–32). IEEE.
- Bertacchini, F., Demarco, F., Scuro, C., Pantano, P., and Bilotta, E. (2023). A social robot connected with chat GPT to improve cognitive functioning in ASD subjects. *Front. Psychol.* 14:1232177. doi: 10.3389/fpsyg.2023.1232177
- Bhargava, A., Bester, M., and Bolton, L. (2021). Employees' perceptions of the implementation of robotics, artificial intelligence, and automation (RAIA) on job satisfaction, job security, and employability. *J. Technol. Behav. Sci.* 6, 106–113. doi: 10.1007/s41347-020-00153-8
- Bialkova, A. (2024). How to optimise interaction with chatbots? Key parameters influencing user experience. *J. Digital Commun.* 11, 175–195. doi: 10.1080/10447318.2023.2219963
- Bickmore, T., and Gruber, A. (2010). Relational agents in clinical psychiatry. *Harvard Rev. Psychiatry* 18, 119–130. doi: 10.3109/10673221003707538
- Bickmore, T., Gruber, A., and Picard, R. (2005). Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Educ. Couns.* 59, 21–30. doi: 10.1016/j.pec.2004.09.008
- Bonnefon, J. F., Rahwan, I., and Shariff, A. (2024). The moral psychology of artificial intelligence. *Annu. Rev. Psychol.* 75, 653–675. doi: 10.1146/annurev-psych-030123-113559
- Breazeal, C. (2003). Toward sociable robots. *Robot. Auton. Syst.* 42, 167–175. doi: 10.1016/s0921-8890(02)00373-1
- Broadbent, E. (2017). Interactions with robots: the truths we reveal about ourselves. *Annu. Rev. Psychol.* 68, 627–652. doi: 10.1146/annurev-psych-010416-044144
- Canó, S., González, C. S., Gil-Iranzo, R. M., and Albiol-Pérez, S. (2021). Affective communication for socially assistive robots (sars) for children with autism spectrum disorder: a systematic review. *Sensors* 21:5166. doi: 10.3390/s21155166
- Casheekar, A., Lahiri, A., Rath, K., Prabhakar, K. S., and Srinivasan, K. (2024). A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: applications, open challenges and future research directions. *Comput Sci Rev* 52:100632. doi: 10.1016/j.cosrev.2024.100632
- Chaturvedi, R., Verma, S., and Srivastava, V. (2024). Empowering AI companions for enhanced relationship marketing. *Calif. Manag. Rev.* 66, 65–90. doi: 10.1177/00081256231215838
- Chuah, S. H.-W., and Yu, J. (2021). The future of service: the power of emotion in human-robot interaction. *J. Retail. Consum. Serv.* 61:102551. doi: 10.1016/j.jretconser.2021.102551
- Churamani, N., and Howard, M. (2022). Affect-driven learning of robot behaviour for collaboration. *Int. J. Robot. Res.* 41, 130–155. doi: 10.3389/frobt.2022.717193
- Ciechanowski, L., Przegalinska, A., Magnuski, M., and Gloor, P. (2018). In the shades of the uncanny valley: an experimental study of human–chatbot interaction. *Futur. Gener. Comput. Syst.* 92, 539–548. doi: 10.1016/j.future.2018.01.055
- Combe, T., Fribourg, R., Detto, L., and Normand, J. M. (2024). Exploring the influence of virtual avatar heads in mixed reality on social presence, performance and user experience in collaborative tasks. *IEEE Trans. Vis. Comput. Graph.* 30, 2206–2216. doi: 10.1109/TVCG.2024.3372051
- Creasey, M. C., and Vázquez Anido, A. (2020). Virtual influencing: Uncharted frontier in the uncanny valley. Available at: <https://lup.lub.lu.se/student-papers/search/publication/9015731>
- Crolic, C., Thomaz, F., Hadi, R., and Stephen, A. T. (2022). Blame the bot: anthropomorphism and anger in customer–chatbot interactions. *J. Mark.* 86, 132–148. doi: 10.1177/00222429211045687
- Dao, X. Q. (2023). Performance comparison of large language models on vnhsge english dataset: Openai chatgpt, microsoft bing chat, and google bard. *arXiv preprint arXiv.*
- Dautenhahn, K. (2018). Some brief thoughts on the past and future of human-robot interaction. *ACM Transactions on Human-Robot Interaction* 7:4. doi: 10.1145/3209769
- Dautenhahn, K., and Saunders, J. (Eds.). (2011). *New frontiers in human-robot interaction, vol. 2*. Amsterdam, Netherlands: John Benjamins Publishing.
- De Brito Silva, M. J., de Oliveira Ramos Delfino, L., Alves Cerqueira, K., and de Oliveira Campos, P. (2022). Avatar marketing: a study on the engagement and authenticity of virtual influencers on Instagram. *Soc. Netw. Anal. Min.* 12:130. doi: 10.1007/s13278-022-00966-w
- de Cock, C., Milne-Ives, M., van Velthoven, M. H., Alturkistani, A., Lam, C., and Meinert, E. (2020). Effectiveness of conversational agents (virtual assistants) in health care: protocol for a systematic review. *JMIR Res. Protocols* 9:e16934. doi: 10.2196/16934
- de Melo, C. M., Gratch, J., and Carnevale, P. J. (2013). “The effect of agency on the impact of emotion expressions on people's decision making.” In *2013 Humaine association conference on affective computing and intelligent interaction* (pp. 546–551). IEEE.
- Diel, A., Weigelt, S., and Maccormack, K. F. (2022). A Meta-analysis of the Uncanny Valley's independent and dependent variables. *ACM Transactions on Human-Robot Interaction* 11, 1–33. doi: 10.1145/3470742
- Doncieux, S., Chatila, R., and Straube, S. (2022). Human-centered AI and robotics. *AI Perspect* 4:1. doi: 10.1186/s42467-021-00014-x
- Felnhöfer, A., Kothgassner, O. D., and Beutl, L. (2024). A virtual character's agency affects social responses. *Cyber Psychol. Behav.* 27, 56–72. doi: 10.1080/10447318.2023.2209979
- Figuroa-Torres, M. (2025). The three social dimensions of chatbot technology. *J. AI Soc. Sci.* 38, 30–50. doi: 10.1007/s13347-024-00826-9
- Fink, M. C., Robinson, S. A., and Ertl, B. (2024). AI-based avatars are changing the way we learn and teach: benefits and challenges. *Front. Educ.* 9. doi: 10.3389/educ.2024.1416307
- Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robot. Auton. Syst.* 42, 143–166. doi: 10.1016/S0921-8890(02)00372-X

- Fortuna, P., Wróblewski, Z., Gut, A., and Dutkowska, A. (2024). The relationship between anthropocentric beliefs and the moral status of a chimpanzee, humanoid robot, and cyborg person: the mediating role of the assignment of mind and soul. *Curr. Psychol.* 43, 12664–12679. doi: 10.1007/s12144-023-05313-6
- Fraser, N., Jones, R., and Patel, M. (2024). Do realistic avatars make virtual reality better? *Virtual Reality & Society* 2, 100082–100140. doi: 10.1016/j.chbah.2024.100082
- Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuro Image* 35, 1674–1684. doi: 10.1016/j.neuroimage.2007.02.003
- Ghigino, D., Marchesi, S., and Wykowska, A. (2023). Play with me: complexity of human-robot interaction and intentional stance adoption. *AI & Soc.* 40, 250–280.
- Glikson, E., and Woolley, A. W. (2020). Human trust in artificial intelligence: review of empirical research. *Acad. Manag. Ann.* 14, 627–660. doi: 10.5465/annals.2018.0057
- Goodrich, M. A., and Schultz, A. C. (2007). Human-robot interaction: a survey. *Foundations and Trends in Human-Computer Interaction* 1, 203–275. doi: 10.1561/11000000005
- Gray, H. M., Gray, K., and Wegner, D. M. (2007). Dimensions of mind perception. *Science* 315:619. doi: 10.1126/science.1134475
- Guinrich, R. E., and Graziano, M. S. (2023). Chatbots as social companions: How people perceive consciousness, human likeness, and social health benefits in machines. *arXiv preprint*. doi: 10.48550/arXiv.2311.10599
- Guinrich, R. E., and Graziano, M. S. (2024). Ascribing consciousness to artificial intelligence: human-AI interaction and its carry-over effects on human-human interaction. *Front. Psychol.* 15:1322781. doi: 10.3389/fpsyg.2024.1322781
- Gutuleac, R., Baima, G., Rizzo, C., and Bresciani, S. (2024). Will virtual influencers overcome the uncanny valley? The moderating role of social cues. *Psychol. Mark.* 41, 1419–1431. doi: 10.1002/mar.21989
- Haman, M., Školník, M., and Kučirková, K. (2024). The rise of talking machines: balancing the potential and pitfalls of voice chatbots for mental wellbeing. *J. Public Health* 46, e715–e716. doi: 10.1093/pubmed/fdae269
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., and Parasuraman, R. (2020). A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* 53, 517–527. doi: 10.1177/0018720811417254
- Hauptman, A. I., Schelle, B. G., McNeese, N. J., and Madathil, K. C. (2023). Adapt and overcome: perceptions of adaptive autonomous agents for human-AI teaming. *Comput. Hum. Behav.* 138:107451. doi: 10.1016/j.chb.2022.107451
- Henschel, A., Hortensius, R., and Cross, E. S. (2020). Social cognition in the age of human-robot interaction. *Trends Neurosci.* 43, 373–384. doi: 10.1016/j.tins.2020.03.013
- Hinz, N.-A., Ciardo, F., and Wykowska, A. (2021). ERP markers of action planning and outcome monitoring in human-robot interaction. *Acta Psychol.* 212:103216. doi: 10.1016/j.actpsy.2020.103216
- Hortensius, R., Hekele, F., and Cross, E. S. (2018). The perception of emotion in artificial agents. *IEEE Transactions Cogn. Develop. Syst.* 10, 852–864. doi: 10.1109/TCDS.2018.2826921
- Hwang, A. H.-C., Siy, J. O., Shelby, R., and Lentz, A. (2024). “In whose voice?: examining AI agent representation of people in social interaction through generative speech.” In *Designing interactive systems conference (pp. 224–245)*. *DIS’24: Designing interactive systems conference*. ACM.
- Indurkha, B. (2023). “Ethical aspects of faking emotions in Chatbots and social robots\*.” In *2023 32nd IEEE International conference on robot and human interactive communication (RO-MAN)* (Vol. 14, pp. 1719–1724).
- Irfan, B., Kuoppamäki, S., and Skantze, G. (2024). Recommendations for designing conversational companion robots with older adults through foundation models. *Front. Robotics AI* 11:1363713. doi: 10.3389/frobt.2024.1363713
- Jacobs, D., and Turner, R. (2022). Mind the robot: observer perception of dominance and mirroring behavior. *J. Robotics AI Ethics* 17, 112–135.
- Jorge, C. C., Jonker, C. M., and Tielman, M. L. (2024). How should an AI trust its human teammates? Exploring possible cues of artificial trust. *ACM Transactions on Interactive Intelligent Systems* 14, 1–26. doi: 10.1145/3635475
- Joshi, S., Kamino, W., and Šabanović, S. (2024). Social robot accessories for tailoring and appropriation of social robots. *Int. J. Soc. Robot.* doi: 10.1007/s12369-023-01077-y
- Kachouie, R., Sedighadeli, S., and Abkenar, A. B. (2017). The role of socially assistive robots in elderly wellbeing: a systematic review. In *Cross-Cultural Design: 9th International Conference, CCD 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings 9* (Springer International Publishing) pp. 669–682.
- Kamino, W., Jung, M. F., and Šabanović, S. (2024). Constructing a social life with robots: shifting away from design patterns towards interaction ritual chains. *AI & Soc.* 45, 112–135. doi: 10.1145/3610977.3634994
- Kang, Y., and Kang, J. (2024). Counseling chatbot design: the effect of anthropomorphism on self-disclosure and companionship. *Comput. Hum. Behav.* 139:108512. doi: 10.1080/10447318.2022.2163775
- Karaduman, T., Pekçetin, T. N., and Urgen, B. A. (2023). Perceived pain of humans and robots: an exploration of empathy gaps. *Comput. Hum. Behav.* 150:109530.
- Kedia, P., Lee, J., Rajguru, M., Agrawal, S., and Tremere, M. (2024). The LLM latency guidebook: Optimizing response times for gen AI applications: Microsoft Tech Community.
- Kerzel, M., Strahl, E., Magg, S., Navarro-Guerrero, N., Heinrich, S., and Wermter, S. (2017). “NICO—Neuro-inspired companion: A developmental humanoid robot platform for multimodal interaction.” *2017 26th IEEE International symposium on robot and human interactive communication (RO-MAN)*, 113–120.
- Koban, K., and Banks, J. (2024). It feels, therefore it is: associations between mind perception and mind ascription for social robots. *Comput. Hum. Behav.* 153:108098. doi: 10.1016/j.chb.2023.108098
- Kosinski, M. (2023). Theory of mind may have spontaneously emerged in large language models. *Proc. Natl. Acad. Sci.* 120:e2300207119. doi: 10.1073/pnas.2300207119
- Krauter, J. (2024). Bridging the Uncanny Valley: improving AI Chatbots for effective leadership mentoring. *Open J. Leadership* 13, 342–384. doi: 10.4236/oj.2024.133021
- Krueger, J., and Roberts, T. (2024). Real feeling and fictional time in human-AI interactions. *Topoi* 43, 783–794. doi: 10.1007/s11245-024-10046-7
- Kwak, S. S., Kim, M., and Kim, J. (2014). “The impact of the robot appearance types on social interaction.” *Proceedings of the Human-Robot Interaction Conference*.
- Kyrlitsias, C., and Michael-Grigoriou, D. (2022). Social interaction with agents and avatars in immersive virtual environments: a survey. *Front. Virtual Reality* 2:786665. doi: 10.3389/frvir.2021.786665
- Lee, Y. J., and Ji, Y. G. (2024). Effects of visual realism on avatar perception in immersive and non-immersive virtual environments. *Int. J. Human-Computer Interaction* 41, 4362–4375. doi: 10.1080/10447318.2024.2351713
- Lee, M., Lucas, G., and Gratch, J. (2021). Comparing mind perception in strategic exchanges: human-agent negotiation, dictator and ultimatum games. *J. Multimodal User Interfaces* 15, 201–214. doi: 10.1007/s12193-020-00356-6
- Lee, K. M., and Nass, C. (2003). “Designing social presence of social actors in human computer interaction.” In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 289–296).
- Lefkeli, D., Ozbay, Y., Gürhan-Canlı, Z., and Eskenazi, T. (2021). Competing with or against cozmo, the robot: influence of interaction context and outcome on mind perception. *Int. J. Soc. Robot.* 13, 715–724. doi: 10.1007/s12369-020-00668-3
- Levine, J. M., Resnick, L. B., and Higgins, E. T. (1993). Social foundations of cognition. *Annu. Rev. Psychol.* 44, 585–612.
- Li, M., Guo, F., Ren, Z., and Duffy, V. G. (2022). A visual and neural evaluation of the affective impression on humanoid robot appearances in free viewing. *Int. J. Ind. Ergon.* 88:103159. doi: 10.1016/j.ergon.2021.103159
- Lim, S., Chen, Y., and Park, H. (2024). Artificial social influence via human-embodied AI. *Cogn. Sci. AI Rev.* 15, 211–230. doi: 10.48550/arXiv.2406.05486
- Lim, S., and Reeves, B. (2010). Computer agents versus avatars: responses to interactive game characters controlled by a computer or other player. *Int. J. Human-Computer Stud.* 68, 57–68. doi: 10.1016/j.ijhcs.2009.09.008
- MacDorman, K. F., and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Stud. Soc. Behav. Commun. Biol. Artificial Systems* 7, 297–337. doi: 10.1075/is.7.3.03mac
- Maninger, T., and Shank, D. B. (2022). Perceptions of violations by artificial and human actors across moral foundations. *Computers Human Behav. Reports* 5:100154. doi: 10.1016/j.chbr.2021.100154
- Marchesi, S., Ghigino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., and Wykowska, A. (2019). Do we adopt the intentional stance toward humanoid robots? *Front. Psychol.* 10:450. doi: 10.3389/fpsyg.2019.00450
- Mei, Q., Xie, Y., Yuan, W., and Jackson, M. O. (2024). A Turing test: are AI chatbots behaviorally similar to humans? *Proc. Natl. Acad. Sci.* doi: 10.1073/pnas.2312
- Memarian, B., and Doleck, T. (2024). Embodied AI in education: A review on the body, environment, and mind. *Education and Information Technologies*, 29, 895–916.
- Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* 19, 98–100. doi: 10.1109/MRA.2012.2192811
- Morrison, I., Löken, L. S., and Olsson, H. (2010). The skin as a social organ. *Exp. Brain Res.* 204, 305–314. doi: 10.1007/s00221-009-2007-y
- Murnane, M., Higgins, P., Saraf, M., Ferraro, F., Matuszek, C., and Engel, D. (2021). “A simulator for human-robot interaction in virtual reality.” In *2021 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)* (pp. 470–471).
- Mustafa, M., Guthe, S., Tauscher, J.-P., Gesele, M., and Magnor, M. (2017). “How human am I?: EEG-based evaluation of virtual characters.” *Proceedings of the 2017 CHI conference on human factors in computing systems*, 5098–5108.
- Nass, C., Steuer, J., and Tauber, E. R. (1994). “Computers are social actors.” In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 72–78).
- Pan, T., Lee, J., and Huang, S. (2024). Desirable or distasteful? Exploring uncertainty in human-chatbot relationships. *J. AI Ethics* 40, 6545–6555. doi: 10.1080/10447318.2023.2256554

- Park, G., Yim, M. C., Chung, J., and Lee, S. (2023). Effect of AI chatbot empathy and identity disclosure on willingness to donate: The mediation of humanness and social presence. *Behav. Inform. Technol.* 42, 1998–2010. doi: 10.1080/0144929X.2022.2105746
- Pekçetin, T. N., Acarturk, C., and Urgen, B. A. (2024). Investigating mind perception in HRI through real-time implicit and explicit measurements. *J. Cogn. Sci. Robotics* 19, 60–85. doi: 10.1145/3610978.3638366
- Pentina, I., Hancock, T., and Xie, T. (2023). Exploring relationship development with social chatbots: a mixed-method study of replika. *Comput. Hum. Behav.* 140:107600. doi: 10.1016/j.chb.2022.107600
- Perez-Osorio, J., and Wykowska, A. (2020). Adopting the intentional stance toward natural and artificial agents. *Philos. Psychol.* 33, 369–395. doi: 10.1080/09515089.2019.1688778
- Possler, D., Carnol, N. N., Klimmt, C., Weber-Hoffmann, I., and Raney, A. A. (2022). “A matter of closeness: player-avatar relationships as degree of including avatars in the self” in Entertainment computing–ICEC 2022. ICEC 2022. eds. B. Göbl, E. Spek, J. Baalsrud Hauge and R. McCall, vol. 13477 (Cham: Springer).
- Ramadan, Z., and Ramadan, J. (2025). AI avatars and co-creation in the metaverse. *Consumer Behav. Tourism Hospitality* 20, 131–147. doi: 10.1108/CBTH-07-2024-0246
- Rao Hill, S., and Troshani, I. (2024). Chatbot anthropomorphism, social presence, uncanniness and brand attitude effects. *J. Comput. Inf. Syst.*, 1–17. doi: 10.1080/08874417.2024.2423187
- Ratajczyk, D. (2022). Shape of the Uncanny Valley and emotional attitudes toward robots assessed by an analysis of you tube comments. *Int. J. Soc. Robot.* 14, 1787–1803. doi: 10.1007/s12369-022-00905-x
- Reeves, B., and Nass, C. (1996). The media equation: how people treat computers, television, and new media like real people, vol. 10. Cambridge, UK, 19–36.
- Rheu, M., Choi, Y., and Kim, E. (2024). When a chatbot disappoints you: expectancy violation in human-chatbot interaction. *Comput. Hum. Behav.* 138:107642. doi: 10.1177/00936502231221669
- Robinson, H., MacDonald, B., and Broadbent, E. (2014). The role of healthcare robots for older people at home: a review. *Int. J. Soc. Robot.* 6, 575–591. doi: 10.1007/s12369-014-0242-2
- Ruane, E. (2019). “Conversational AI: Social and ethical considerations,” *AICS -27th AIAI Irish conference on artificial intelligence and cognitive science*.
- Safadi, F., Fonteneau, R., and Ernst, D. (2015). Artificial intelligence in video games: towards a unified framework. *Int. J. Computer Games Technol.* 2015:271296. doi: 10.1155/2015/271296
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare* (Vol. 11, p. 887). MDPI.
- Sandini, G., Sciutti, A., and Morasso, P. (2024). Artificial cognition vs. artificial intelligence for next-generation autonomous robotic agents. *Front. Comput. Neurosci.* 18:408. doi: 10.3389/fncom.2024.1349408
- Sarkar, S., Gaur, M., Chen, L. K., Garg, M., and Srivastava, B. (2023). A review of the explainability and safety of conversational agents for mental health to identify avenues for improvement. *Front. Artificial Intelligence* 6:805. doi: 10.3389/frai.2023.1229805
- Saxe, R., and Kanwisher, N. (2013). People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *Soc. Neurosci.*, 171–182.
- Sedlakova, J., and Trachsel, M. (2022). Conversational artificial intelligence in psychotherapy: a new therapeutic tool or agent? *Am. J. Bioeth.* 23, 4–13. doi: 10.1080/15265161.2022.2048739
- Seitz, L. (2024). Artificial empathy in healthcare chatbots: does it feel authentic? *Computers Human Behav.: Artificial Humans* 2:100067. doi: 10.1016/j.chbah.2024.100067
- Selvi, A., Mounika, V., Rubika, V., and Uvadharee, B. (2024). “Collegobot: virtual assistant system for enquiry using natural language processing.” In *2024 2nd International conference on intelligent data communication technologies and internet of things (IDCIoT)* (pp. 1407–1414). IEEE.
- Shapira, N., Zwirn, G., and Goldberg, Y. (2023). “How well do large language models perform on faux pas tests?” in Findings of the Association for Computational Linguistics: ACL 2023 (Findings of the Association for Computational Linguistics: ACL 2023. Association for Computational Linguistics), 10438–10451.
- Shawar, B. A., and Atwell, E. (2007). Chatbots: are they really useful? *J. Lang. Technol. Computational Linguistics* 22, 29–49. doi: 10.21248/jlcl.22.2007.88
- Shin, D. (2020). The impact of virtual agents on social media interaction: intimacy and engagement. *J. Digital Commun.* 9, 155–180.
- Shin, D., and Kim, H. J. (2020). Exploring the role of social robots in interactive learning: the effect of robot facial expression and gender. *Int. J. Human-Computer Interaction* 36, 81–93. doi: 10.1080/10447318.2019.1614547
- Silva, A., Schrum, M., Hedlund-Botti, E., Gopalan, N., and Gombolay, M. (2023). Explainable artificial intelligence: evaluating the objective and subjective impacts of xai on human-agent interaction. *Int. J. Human-Computer Interaction* 39, 1390–1404. doi: 10.1080/10447318.2022.2101698
- Siri, G., Abushait, A., De Tommaso, D., Cardellicchio, P., D’Ausilio, A., and Wykowska, A. (2022). “Perceptions of a robot’s mental states influence performance in a collaborative task for males and females differently.” *2022 31st IEEE International conference on robot and human interactive communication (RO-MAN)*, 1238–1243.
- Skjuve, M., Haugstveit, I. M., Følstad, A., and Brandtzaeg, P. B. (2019). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Hum. Technol.* 15, 30–54. doi: 10.17011/ht/urn.201902201607
- Smith, E. R., Šabanović, S., and Fraune, M. R. (2021). Human-robot interaction through the lens of social psychological theories of intergroup behavior. *Technology, Mind, Behav.* 1, 1–30. doi: 10.1037/tmb0000002#supplemental-materials
- Song, S. W., and Shin, M. (2024). Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: the moderating role of avatar familiarity. *Int. J. Human-Computer Interaction* 40, 441–456. doi: 10.1080/10447318.2022.2121038
- Spatola, N., and Chaminade, T. (2022). Precuneus brain response changes differently during human-robot and human-human dyadic social interaction. *Sci. Rep.* 12:14794. doi: 10.1038/s41598-022-14207-9
- Spisak, B. R., and Indurkha, B. (2023). A study on social exclusion in human-robot interaction. *AI & Soc.* 38, 456–478. doi: 10.3390/electronics12071585
- Starke, S., Zhao, Y., Komura, T., and Zaman, K. A. (2020). Local motion phases for learning multi-contact character movements. *ACM Trans. Graph.*, 39, 54.
- Stein, J.-P., Appel, M., Jost, A., and Ohler, P. (2020). Matter over mind? How the acceptance of digital entities depends on their appearance, mental prowess, and the interaction between both. *Int. J. Human-Computer Stud.* 142:102463. doi: 10.1016/j.ijhcs.2020.102463
- Storbeck, J., and Clore, G. L. (2007). On the interdependence of cognition and emotion. *Cognit. Emot.* 21, 1212–1237. doi: 10.1080/02699930701438020
- Strachan, J., Albergo, D., Borghini, G., Pansardi, O., Scaliti, E., Rufo, A., et al. (2023). Testing theory of mind in GPT models and humans. Research Square Platform LLC.
- Su, H., Qi, W., Chen, J., Yang, C., Sandoval, J., and Laribi, M. A. (2023). Recent advancements in multimodal human-robot interaction. *Front. Neurobot.* 17:1084000. doi: 10.3389/fnbot.2023.1084000
- Syamsara, A., and Widiastuty, H. (2024). “Student perceptions of AI Claude as a tool for academic writing development.” In *International Seminar* (Vol. 6, pp. 890–898).
- Tan, C. K., Lou, V. W., Cheng, C. Y. M., He, P. C., and Khoo, V. E. J. (2024). Improving the social well-being of single older adults using the LOVOT social robot: qualitative phenomenological study. *JMIR Hum. Factors* 11:e56669. doi: 10.2196/56669
- Teubner, T., Adam, M., and Riordan, R. (2015). The impact of computerized agents on immediate emotions, overall arousal and bidding behavior in electronic auctions. *J. Assoc. Inf. Syst.* 16, 838–879. doi: 10.17705/1jais.00412
- Thellman, S., De Graaf, M., and Ziemke, T. (2022). Mental state attribution to robots: a systematic review of conceptions, methods, and findings. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 1–51. doi: 10.1145/3526112
- Torubarova, E., Arvidsson, C., Udden, J., and Pereira, A. (2023). Differences in brain activity during turn initiation in human-human and human-robot conversation.
- Trothen, T. J. (2022). Replika: spiritual enhancement technology? *Religion* 13:275. doi: 10.3390/rel13040275
- Vaitonytė, J., Alimardani, M., and Louwerse, M. M. (2023). Scoping review of the neural evidence on the uncanny valley. *Computers Human Behav. Reports* 9:100263. doi: 10.1016/j.chbr.2022.100263
- Veras, M., Labbé, D. R., Furlano, J., Zakus, D., Rutherford, D., Pendergast, B., et al. (2023). A framework for equitable virtual rehabilitation in the metaverse era: challenges and opportunities. *Front. Rehab. Sci.* 4:1020. doi: 10.3389/fresc.2023.1241020
- Vishwakarma, L. P., Singh, R. K., Mishra, R., Demirkol, D., and Daim, T. (2024). The adoption of social robots in service operations: a comprehensive review. *Technol. Soc.* 76:102441. doi: 10.1016/j.techsoc.2023.102441
- Von Der Pütten, A. M., and Krämer, N. C. (2010). It doesn’t matter what you are! Explaining social responses to agents and avatars. *J. Virtual Reality* 7, 18–36. doi: 10.1016/j.chb.2010.06.012
- Walker, M., Phung, T., Chakraborti, T., Williams, T., and Szafir, D. (2023). Virtual, augmented, and mixed reality for human-robot interaction: a survey and virtual design element taxonomy (version 1). *ACM Trans. Hum.-Robot Interact.* 12, 1–39. doi: 10.48550/ARXIV.2202.11249
- Wallkotter, S., Stower, R., Kappas, A., and Castellano, G. (2020). “A robot by any other frame: framing and behaviour influence mind perception in virtual but not real-world environments.” *Proceedings of the 2020 ACM/IEEE International conference on human-robot interaction*, 609–618.
- Wang, Y., Gong, D., Xiao, R., Wu, X., and Zhang, H. (2024a). A systematic review on extended reality-mediated multi-user social engagement. *System* 12:396. doi: 10.3390/systems12100396
- Wang, Y., and Quadflieg, S. (2015). In our own image? Emotional and neural processing differences when observing human-human vs human-robot interactions. *Soc. Cogn. Affect. Neurosci.* 10, 1515–1524. doi: 10.1093/scan/nsv043
- Wang, Y., Zhu, M., Chen, X., Liu, R., Ge, J., Song, Y., et al. (2024b). The application of metaverse in healthcare. *Front. Public Health* 12:367. doi: 10.3389/fpubh.2024.1420367
- Waytz, A., Heafner, J., and Epley, N. (2014). The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. *J. Exp. Soc. Psychol.* 52, 113–117. doi: 10.1016/j.jesp.2014.01.005

- Weiss, A., Bernhaupt, R., Lankes, M., and Tscheligi, M. (2009). "The USUS evaluation framework for human-robot interaction." In *AISB 2009: Proceedings of the symposium on new frontiers in human-robot interaction* (Vol. 4, pp. 11–26).
- Wiese, E., Metta, G., and Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Front. Psychol.* 8:1663. doi: 10.3389/fpsyg.2017.01663
- Woo, J., Tanaka, K., and Suzuki, M. (2024). Adaptive virtual agent design and evaluation for therapy. *J. AI Mental Health* 10, 321–338. doi: 10.1016/j.ijhcs.2024.103321
- Wykowska, A., Wiese, E., Prosser, A., and Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLoS one*, 9, e94339.
- Xie, T., and Pentina, I. (2022). Attachment theory as a framework to understand relationships with social chatbots: A case study of Replika.
- Yam, K. C., Tan, T., Jackson, J. C., Shariff, A., and Gray, K. (2023). Cultural differences in People's reactions and applications of robots, algorithms, and artificial intelligence. *Manag. Organ. Rev.* 19, 859–875. doi: 10.1017/mor.2023.21
- Yamazaki, R., Nishio, S., Nagata, Y., Satake, Y., Suzuki, M., Kanemoto, H., et al. (2023). Long-term effect of the absence of a companion robot on older adults: a preliminary pilot study. *Front. Computer Sci.* 5. doi: 10.3389/fcomp.2023.1129506
- Yang, A. K., and Hu, B. (2024). AI interventions: assessing the potential for a chat GPT-based loneliness and cognitive decline intervention in geriatric populations. *OSF*. doi: 10.3389/fnbot.2023.1084000
- Yee, N., and Bailenson, J. (2007). The Proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res.* 33, 271–290. doi: 10.1111/j.1468-2958.2007.00299.x
- Yu, L., Li, Y., and Fan, F. (2023). Employees' appraisals and trust of artificial intelligences' transparency and opacity. *Behav. Sci.* 13:344. doi: 10.3390/bs13040344
- Yu, R., and Zhao, W. (2024). Emojifying chatbot interactions: an exploration of social cues in digital conversations. *Digital Psychol.* 18, 88–110. doi: 10.1016/j.tele.2023.102071
- Yuan, Z., Cheng, X., and Duan, Y. (2024). Impact of media dependence: how emotional interactions between users and chat robots affect human socialization? *Front. Psychol.* 15:860. doi: 10.3389/fpsyg.2024.1388860
- Zhang, J., Li, S., Zhang, J. Y., Du, F., Qi, Y., and Liu, X. (2020). "A literature review of the research on the uncanny valley." In *Cross-cultural design. User experience of products, services, and intelligent environments: 12th International conference, CCD 2020, held as part of the 22nd HCI International conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22* (pp. 255–268). Springer International Publishing.
- Zhang, A., and Rau, P. L. P. (2023). Tools or peers? Impacts of anthropomorphism level and social role on emotional attachment and disclosure tendency towards intelligent agents. *Comput. Hum. Behav.* 138:107415. doi: 10.1016/j.chb.2022.107415
- Zhu, Q., Chau, A., Cohn, M., and Liang, K.-H., (2022). "Effects of emotional expressiveness on voice Chatbot interactions." *4th conference on conversational user interfaces (CUI 2022), Glasgow, United Kingdom.*
- Ziemke, T. (2023). Understanding social robots: attribution of intentional agency to artificial and biological bodies. *Art&Life* 29, 351–366. doi: 10.1162/artl\_a\_00404
- Zlotowski, J., Proudfoot, D., Yogeewaran, K., and Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human-robot interaction. *Int. J. Soc. Robot.* 7, 347–360. doi: 10.1007/s12369-014-0267-6