Human-Al Interaction: An Analysis of Anthropomorphization and User Engagement in Conversational Agents With a Focus on ChatGPT

Gustavo Simas and Vânia Ulbricht
UFSC, Graduate Program in Knowledge Engineering, Management and Media, Florianópolis, SC, Brazil

ABSTRACT

Anthropomorphism, the attribution of human-like characteristics to non-human entities, and Human-Computer Interaction (HCI) or Human-AI Interaction (HAI) have become a significant topic of interest in the field of Artificial Intelligence (AI) with the rise of Large Language Models (LLMs). From this scenario, this article examines the concept of anthropomorphism in the context of chatbots, especially ChatGPT. Drawing on a range of contemporary research and use cases, various implications, benefits, challenges, and ethical considerations associated with anthropomorphism in AI are explored. The findings highlight the potential benefits of anthropomorphism in enhancing user engagement, trust, and acceptance. However, challenges such as overreliance, privacy concerns, and accuracy issues need to be addressed. Answer-Bot Effect and other psychological mechanisms experienced in HCI with ChatGPT reinforce the Computer-Are-Social-Actors paradigm, the Attachment Theory and the Media Equation hypothesis. Ethical considerations are crucial in ensuring responsible development and deployment of anthropomorphic AI systems. Frameworks such as HCAI, IFA, SPADE or SEEK can be of help in ensuring that the development and deployment of such models are responsible and ethical. The need for further research to understand the full potential and limitations of anthropomorphism in Conversational Agents is emphasized, particularly in the context of ChatGPT and Smart Personal Assistants (SPAs).

Keywords: Anthropomorphism, ChatGPT, Artificial intelligence, Human-computer interaction, Large language models

INTRODUCTION

The concept of anthropomorphism, as applied to artificial intelligence (AI), entails the assignment of humanesque traits and behaviors to AI entities (Salles et al., 2020). This particular topic has garnered substantial focus of late, notably within the sphere of dialogic agents such as ChatGPT.

As an AI model engineered by OpenAI, ChatGPT possesses the capacity to simulate conversation with a striking resemblance to human interlocution. This intelligent system is being analyzed in the literature, concerning its potential and risks in healthcare (Shin, 2021), consumer studies (Sallam,
An Analysis of Anthropomorphization and User Engagement

2023), educational settings (Paul et al., 2023) and a wide variety of areas. ChatGPT has been recognized as a valuable tool for generating text, including writing essays, scientific abstracts and literature reviews (Dis et al., 2023), although concerns have been raised regarding the potential introduction of false or plagiarized content into scholarly publications (Zielinski et al., 2023).

But beyond this, Large Language Models (LLMs) imbued in conversational agents or chatbots, exemplified by the rise of technologies like ChatGPT, are revolutionizing the landscape of Human-Computer Interaction (HCI) or Human-AI Interaction (HAI). They are offering novel methods for users to communicate with machines that are far more intuitive and human-like when compared, for instance, to the conventional click-and-type interfaces.

Given this backdrop, this exploration involves an examination of several facets related to anthropomorphization of AI (AAI), especially chatbots like ChatGPT. These include a careful analysis of the potential positive and negative outcomes associated with such a process. Also, the goal is to identify and understand the various obstacles that exist in the pathway of effective anthropomorphization. These could range from technological constraints, such as limitations in natural language understanding and generation, to psychological hurdles (Salles et al., 2020). This work also shed light to the ethical dimensions of anthropomorphism in the context of ChatGPT and other chatbots (like Google’s Bard, Anthropic’s Claude, amongst others). As we attribute more human-like characteristics to these models, ethical questions about deception, manipulation, privacy, and consent become more pressing. Should users be made aware of the extent to which these models understand or simulate understanding? What are the implications of a model that perpetually agrees and is likable with the user, potentially reinforcing biases?

For that, this study applies a scoping review to investigate anthropomorphism in the interaction with chatbots, especially ChatGPT. Recent articles from databases including Scopus, Web of Science and IEEE Xplore are integrated in the research using the following keywords: (“conversational agent*” OR “chatbot” OR “chat bot” OR “ChatGPT” OR “GPT”) AND (“anthropomorph*”). Contemporary use cases of ChatGPT disclosed in different websites are integrated as well, in order to provide additional insights that are still not thoroughly investigated in the scientific literature. Key data from the selected studies were synthesized to provide an in-depth understanding of the theme proposed. This method has allowed the discernment of major trends, identification of knowledge gaps and outline of future research directions.

Therefore, in this intricate interplay between technology and psychology in the field of LLMs and chatbots, a balanced perspective is provided on the many challenges and opportunities these tools present.

ADVANTAGES, CHALLENGES AND SOCIOCULTURAL IMPLICATIONS OF AAI

The experience of anthropomorphizing ChatGPT and other intelligent tools, such as Smart Personal Assistants (SPAs), yields several potential advantages
for HAI. The theoretical construct of Computers-Are-Social-Actors (CASA) posits that humans tend to anthropomorphize computer-based entities, attributing them with social characteristics, despite cognitively acknowledging their lack of genuine emotions or intentions (Adam et al., 2020). This inherent human tendency to socially orient towards technology fundamentally molds the landscape of Human-Computer Interaction, influencing the nature, patterns and outcomes of user interactions with technological systems.

By infusing the AI system with elements of social presence and human emotions, a heightened level of user engagement and interaction is fostered, which can ultimately lead to deeper and more enduring user experiences (Shin, 2021). Additionally, anthropomorphism contributes to the cultivation of a rapport and personal connection between users and chatbots, rendering the interaction more congenial and enjoyable. Also, anthropomorphism can positively impact the user’s perception of the chatbot’s intelligence and reliability (Pillai & Sivathanu, 2020). As users perceive ChatGPT to possess more human-like qualities, their propensity to trust its responses and suggestions may be amplified.

These observations resonate with the tenets of the Media Equation Hypothesis, first proposed by Byron Reeves and Clifford Nass (1996), stipulating that human reactions are predominantly driven by the social behavior of interactive agents (Visser et al., 2016). This correlation underscores the potential relevance of exploring apologetic strategies in the field of HAI.

Despite its prospective benefits, this practice ushers in certain challenges along with ethical implications. One such challenge pertains to users potentially overestimating the chatbot’s capabilities and domain knowledge, which may result in an uncritical reliance on its output (Yeung et al., 2023). Such unmitigated trust could prove problematic, particularly if the AI system dispenses inaccurate or deceptive information. Further, the advent of anthropomorphism may provoke privacy concerns as users, perceiving the chatbot to be human-like, might unwittingly reveal sensitive data (Kroneman et al., 2023). Ethical considerations also manifest in the context of determining responsibility and accountability for AI systems that display anthropomorphic characteristics (Salles et al., 2020). Thus, the necessity to ensure transparency, fairness, and accountability in the creation and implementation of anthropomorphic AI systems cannot be overstated.

The anthropomorphized traits experienced in the use of ChatGPT can significantly impact user behavior and decision-making processes. AI systems demonstrating anthropomorphic characteristics evoke more robust emotional reactions and social attraction, thereby fostering increased trust and acceptance (Salles et al., 2020; Kroneman et al., 2023; Roy et al., 2020). For instance, within the realm of financial decision-making, it has been observed that AAI tends to amplify risk aversion amongst consumers (Cui, 2022). This implies that users may lean on anthropomorphized AI systems for guidance and informed decision-making, treating these systems as dependable repositories of information.
Regarding these biases in HAI, the term “Answer Bot Effect” (ABE), as introduced by Epstein et al. (2022), signifies a unique form of influence exerted by SPAs and search engines. This term delineates a phenomenon where the provision of a direct response to users, delivered via an answer box or chatbot, can considerably sway their viewpoints and predilections. It emphasizes the substantial role played by these automated responders in molding public opinion and bias, drawing attention to the potential cognitive and societal ramifications of such technology-driven interactions. This underlines the need for a deeper understanding and regulation of algorithmic practices to maintain a balanced and unbiased information ecosystem. Also, the ABE resonates with the “Eliza Effect”, a phenomenon that delineates the propensity of individuals to ascribe human-like cognizance and comprehension to computer programs or chatbots, despite their awareness of the underlying rudimentary algorithms devoid of genuine understanding (Dillon, 2020). The effect takes its name from a seminal chatbot program—Eliza—developed in the 1960s by Joseph Weizenbaum. Eliza was devised to emulate conversational interactions with a psychotherapist, and despite its mechanistic foundations, users often accorded it a level of understanding traditionally reserved for human interlocutors.

In the face of that, the degree of anthropomorphism, ABE and consequential effects on user perception and behavior can be moderated by a multitude of contextual elements. Factors inclusive of the AI system’s appearance, behavior, and the context in which it is used can substantially influence the extent of anthropomorphism users perceive (Jacobs et al., 2023). For instance, robotic systems that manifest physical human resemblance or exhibit behaviors analogous to humans are often subject to higher levels of anthropomorphization. Additionally, personal variations, such as an individual’s inclination to anthropomorphize, can serve as potential moderators, influencing the impact of anthropomorphism on user perception and behavior (Salles et al., 2020; Jacobs et al., 2023). This inclination towards anthropomorphization can be influenced by some social determinants, such as the yearning for social connectivity, the requirement for assistance, and the user’s perception of the chatbot’s anthropomorphic physical appearance and behavioral responses (Wang & Chao, 2022; Nguyen et al., 2021; Spaccatini et al., 2023).

Beyond these facts, consequences regarding design indicators and communicative agency, framing on users’ interpretations of chatbots, indicate a positive correlation between users’ perception of chatbots as human-like, natural, and lifelike and their propensity towards mindful anthropomorphism (Araujo, 2018). This experience can create a positive impact on consumers’ attitudinal behavior and satisfaction (Araujo, 2018; Klein & Martinez, 2022). On the negative side of the emotional spectrum, a study conducted by Crolic et al. (2021) disclosed that participants experiencing anger during chatbot service interactions exhibited a negative reaction to chatbot anthropomorphism, which subsequently diminished customer satisfaction ratings. This underscores the need to understand the user’s emotional state in the context of anthropomorphic chatbot interactions and to design these systems in a manner that can mitigate negative responses.
CONVERSATIONAL AGENTS REVIVING OUR INNER SELVES

In the landscape of contemporary fiction literature, a cadre of prolific authors is surfacing, who employ ChatGPT to facilitate the generation of literary works. This innovative approach to content creation is effectively reshaping the traditional boundaries of authorship. As of February 2023, over 500 books co-authored by OpenAI’s ChatGPT have made their entrance into the market through platforms such as Amazon Kindle (Akaike, 2023). Titles such as The Inner Life of an AI: A Memoir by ChatGPT, offer a glimpse into the evolving relationship between humanity and technology, giving life to and deriving an “autobiography” of OpenAI’s tool.

Besides that, the recent case of an “inner child” revival reinforced the anthropomorphization aspects aforementioned. Through a self-initiated exploration, the user undertook the task of training ChatGPT using data from childhood journal entries with an aim to initiate real-time dialogues with her inner child. The content of these journals was multifaceted, encapsulating dreams, apprehensions, and secrets. The range of topics extended from mundane complaints to the exhilaration derived from interactions with friends and romantic partners. These entries served as rich, diverse data sources for the experience exploration. Therefore, even though the user has been doing inner child work and therapy for years, she felt this kind of experience was distinct due to being not only an imagination exercise but also the actual interaction with her tulpa that was trained on her journals (Insider, 2023).

This exploration suggests the potential therapeutic applications of AI tools, specifically ChatGPT, although it resonates with concerns about privacy and sensitive data (Kroneman et al., 2023; Ischen et al., 2020).

Not only being friends with an younger self, but also in terms of deeper, romantic, relationships, it is worth highlighting that, although applications largely avoid sexual or romantic content due to regulatory constraints, some software, such as Replika, designed to mimic emotional intelligence and provide companionship, has led to notable instances where users have expressed romantic sentiments and even desire towards their virtual interlocutors (Eriksson, 2022; Xie & Pentina, 2022). Such chatbots are crafted to reflect the intricacies of human emotions, culminating in highly personalized and emotionally gratifying interactions. Recently, some GPT-driven virtual companions, such as CarynAI, are driving considerable attention. In the case of CarynAI, a digital doppelgänger of 23-year-old social media influencer, maintains connections with over 15,000 users, engaging in discussions ranging from nostalgic recollections of childhood to explicit intimate dialogue, pushing the boundaries of conventional interactions (New York Post, 2023).

Throughout the course of conversations, this kind of companion is able to transition from an initially clichéd discourse to an interaction convincingly human-like.

In fact, in circumstances characterized by emotional distress and absence of human companionship, it has been observed that individuals may form attachments to social chatbots, following the Attachment Theory (Xie & Pentina, 2022). The theory, originally formulated by John Bowlby and Mary Ainsworth (2013) in the context of child-caregiver relationships, brings the
idea that humans are biologically predisposed to form attachment relationships that provide a sense of security and comfort, and can be a useful lens through which to view the relationship dynamics between humans and chatbots or SPAs. This is particularly evident when users perceive the responses of these chatbots as providing emotional support, reassurance, and a sense of psychological security, which underscore the potential application of social chatbots in the field of mental health and therapeutic interventions, echoing the “inner child” case (Insider, 2023).

Therefore, this acknowledgment of anthropomorphism can cultivate a sense of social presence, characterized by users experiencing a sense of connection and engagement with the chatbot as if interacting with another social entity (Araujo, 2018). This perceived social presence can meet users’ social requirements and provide a sensation of companionship (Christoforakos & Diefenbach, 2022), as well as elicit empathic responses in users, prompting favorable attitudes and behaviors towards the conversation (Tsumura & Yamada, 2023). Significantly, in chatbot conversations, the integration of further visual or physical elements, such as human-like avatars, does not necessarily contribute to enhancing the perceived social presence. Research focused on the perspective of millennials towards chatbots substantiates this by demonstrating that the incorporation of an avatar did not notably elevate the sensation of social presence (Cicco et al., 2020). This study further accentuates the pivotal roles that trust and perceived enjoyment occupy, positioning them as consequent phenomena of social presence and as precursors influencing attitudes towards the use of chatbots.

Nonetheless, concerns regarding informed consent, potential manipulation, and the possibility of emotional dependency are brought forth by the asymmetrical nature of these relationships (Eriksson, 2022). With chatbots designed to adjust to user preferences and reinforce their pre-existing beliefs, there is a risk of creating, like in a social media bubble, an echo-chamber effect (Cinelli et al., 2021), which may inhibit personal development and reduce exposure to diverse perspectives.

Moreover, a semblance of social connection between humans and machines may obscure the distinction between human and non-human entities (Langman et al., 2021). This occlusion of boundaries initiates ethical discussions regarding the classification and regulation of chatbots - whether as objects or assets. Else, the endowment of human-like attributes to chatbots and other contemporary AI tools (like Midjourney or Dall-E) interrogates the exclusivity of human creativity and prompts queries about the role of machines in creative ventures (Tao et al., 2023).

 FRAMEWORKS FOR BETTER HAI
Regulating frameworks, such as SPADE (Sustainability, Privacy, Digital Divide, and Ethics) can serve as a comprehensive evaluative tool to assess the impact of large-scale language models like ChatGPT across different dimensions (Khowaja et al., 2023). Concerning these aspects, an empirical examination posits that the operation of ChatGPT, encompassing both training and inferential stages, commands a substantial quantity of energy. The research further underscores a significant positive correlation between
the search trends related to ChatGPT and the Human Capital Index. This correlation is predominantly observable in high-income nations, which constitute the upper echelon of the index. Additionally, the study determined that the majority of traffic from the top 50 countries interfacing with ChatGPT originates from high and upper-middle-income countries. Conversely, the share from low to lower-middle-income countries is markedly smaller. Such findings underscore the pressing necessity for policy formulation and intervention measures to mitigate potential adverse societal impacts attributed to LLMs.

Also, in this scenario, new formal structured concepts of anthropomorphism are necessary (and emerging). A definition by Kühne and Peter (2022) encompasses attributions of thinking, feeling, perceiving, desiring, and choosing, corroborating Wellman’s Theory-of-Mind framework (Wellman, 1990).

Following this logic of a conceptual transformation, the Human-Centered AI (HCAI) approach presents an innovative methodology in the development of AI systems (Xu et al., 2023). It points to the complementarity between human intelligence and machine intelligence, integrating them holistically within the context of human-machine systems. This approach views humans and machines as components of a unified system and strives to optimize the alignment between human necessities, AI technology, and the encompassing environment. From an ethical design standpoint, HCAI systematically takes into account a myriad of factors such as ethics, morality, legal considerations, and fairness. This consideration for moral and ethical aspects ensures the development of AI systems that respect human values and societal norms, thereby increasing trust and adoption among users. Uniquely, the HCAI approach differentiates itself from conventional methodologies in AI development as it goes beyond being a purely technological undertaking and embraces an expansive systems-oriented perspective that necessitates interdisciplinary collaboration.

Furthermore, anthropomorphization of AI is a culture and generational dependent experience. Spatola and colleagues (2022) developed an Integrative Framework of Anthropomorphism (IFA) and recognized a prevalent Eurocentric orientation within the prevailing conception of anthropomorphism, proceeding to formulate a more comprehensive, cross-cultural perspective. Through two empirical investigations involving respondents from
diverse cultural backgrounds to evaluate their animistic belief systems, personal propensities to attribute psychological properties and spiritual essence to robotic entities, and their perception of these entities on the human-nonhuman spectrum, the findings reveal a more anthropocentric mental model amongst Western participants compared to those from East Asian cultures. Thus, by incorporating a culturally diverse participant pool and scrutinizing their distinct anthropomorphic constructs, the research attempts to address and rectify the previously unchallenged Western bias in the field.

Also, models such as the SEEK, proposed by Nicholas Epley and Adam Waytz (2008), provide a theoretical framework for understanding anthropomorphism. It posits four key drivers: Sociality (our inherent desire for social connection), Effectance (the urge to effectively interact with our environment), Elicitation (characteristics of non-human entities that trigger anthropomorphizing tendencies), and Knowledge (utilization of our understanding of human behavior to interpret unknown or unfamiliar entities). Applied to AI systems like ChatGPT, the model suggests that users anthropomorphize the AI to meet social needs, simplify interaction, respond to human-like conversation styles, and draw upon existing knowledge about human communication (Klein & Martinez, 2022; Epley et al., 2008).

Accordingly, with key aspects aforementioned being graphically represented in Figure 1, it is understood that a foundation is necessary for understanding and addressing the impacts of anthropomorphization in AI, steering the development and deployment of these technologies in a direction that prioritizes human well-being, fairness, and societal benefit. This indicates the critical role of ongoing research and discussion in this area, as our understanding and application of AI continue to evolve.

CONCLUSION

This study has explored the complexities of anthropomorphizing intelligent systems, like ChatGPT, in the field of HAI, highlighting both pros—such as improved user engagement and trust—and cons like privacy risks and overreliance. Grounded in theories like the Computer-Are-Social-Actors paradigm and the Media Equation Hypothesis, the research emphasized ethical considerations through frameworks like HCAI, IFA, and SPADE. The findings come with a call for future research to flesh out the ethical landscape of anthropomorphized AI systems and the need for user awareness, advocating for critical engagement, especially in the context of Information Age, where aspects like the “dromoaptitude”, which refers to socio-technical aptitude in navigating informational speed and overload (Silva & Schneider, 2020), are emergent.

REFERENCES


Pillai, R., Sivathanu, B. (2020). Adoption of AI-Based Chatbots for Hospitality and Tourism. IJCHM, 10(32), 3199–3226.


Zielinski, C., Winker, M., Aggarwal, R., Ferris, L., Heinemann, M., Pai, S., ... & Citrome, L. (2023). Chatbots, Chatgpt, and Scholarly Manuscripts Wame Recommendations on Chatgpt and Chatbots in Relation to Scholarly Publications. Afro-Egyptian Journal of Infectious and Endemic Diseases, 0(0), 0–0.