When Do Humans See AI as Humanlike? The Role of Perceived Performance in AI Anthropomorphism

Anonymous Authors*

As human-AI collaboration becomes increasingly important, anthropomorphizing AI is a key question for future AI interactions. However, few studies have explored the conditions under which humans anthropomorphize GenAI and how this perception influences performance improvement. Therefore, this study examines the relationship between perceived performance and anthropomorphizing AI and its impact on collaborative effectiveness. We designed a group decision-making process using the NASA Exercise, with ChatGPT acting as the GenAI agent. Participants' anthropomorphized attitudes and perceived performances were measured and statistically analyzed, focusing on synergy. As a result, participants were more likely to anthropomorphize ChatGPT when they perceived its performance as higher. Additionally, a greater proportion of anthropomorphized attitudes was observed in the synergy group, suggesting a possible link between anthropomorphism and collaborative success. This research provides experimental insights into the conditions under which humans anthropomorphize GenAI, potentially leading to more effective human-AI collaboration.

CCS CONCEPTS • Insert your first CCS term here • Insert your second CCS term here • Insert your third CCS term here

Additional Keywords and Phrases: Human-AI interaction, Anthropomorphism, Synergy, Group decision making

1 INTRODUCTION

Anthropomorphism of Artificial Intelligence (AI) is one of the key topics in human-AI collaboration [12, 18, 22]. Based on various theories, numerous studies have investigated humanlike AI-enabled technology and identified several factors associated with anthropomorphism.

However, few studies have explored the relationship between anthropomorphism and the perceived performance of automation in the context of Generative AI (GenAI). Users' perceptions play a crucial role in technology acceptance [5], and in particular, perceived performance is associated with perceived trust in AI [2]. Also, since human interactions with GenAI differ from those with traditional AI [1, 24], further research is needed in the context of GenAI.

Therefore, our research aims to explore the relationship between perceived performance and anthropomorphism and its effectiveness in human interactions with ChatGPT. Human attitudes toward ChatGPT—whether they anthropomorphize it or not—may vary based on perceived performance. Additionally, its effectiveness can be quantitatively measured and assessed using the concept of synergy.

We measured participants' attitudes and estimated performance toward ChatGPT both before and after the exercise, which was structured as a group decision-making process using the NASA Exercise. Attitude-related questions included four humanlike categories (teammate, assistant, coach, and manager) and one non-humanlike category (robot/machine). The experiment consisted of a human-human interaction phase followed by a human-ChatGPT interaction phase.

After preprocessing the estimated performances to relative perceived performances, we statistically compared them with attitudes and synergy groups. The relationship between perceived performance and attitude was analyzed using a t-test or Wilcoxon rank-sum exact test. A chi-squared test was conducted to compare synergy groups and assess the effectiveness of anthropomorphism.

As a result, participants were more likely to anthropomorphize ChatGPT when its perceived performance was high, and the synergy group exhibited a higher proportion of anthropomorphism. Specifically, participants perceived ChatGPT-4o's performance superior to ChatGPT-3.5, and more participants regarded ChatGPT-4o as humanlike. Furthermore, an anthropomorphizing attitude toward ChatGPT-4o was associated with improved performance outcomes.

We experimentally explored the relationship between perceived performance and anthropomorphized attitudes and their effectiveness. Our contributions to the HCI field are as follows:

- Utilizing ChatGPT, we explored conditions when humans anthropomorphize GenAI with respect to perceived performance, addressing the research gap in human-AI interaction.
- If GenAI's perceived performance is designed to be sufficiently high, it can naturally foster anthropomorphized attitudes without the existence of human-like characteristics, ultimately enhancing human-AI team performance.

2 RELATED WORKS

Anthropomorphism refers to the attribution of human characteristics to inanimate or non-social entities [4, 6]. Due to its complexity and nuanced nature [7], various theories have emerged to explain this phenomenon [14]. Examples include media equation theory [21], the Computers Are Social Actors (CASA) paradigm [3], social response theory [9, 17], and the three-factor theory of anthropomorphism [8].

Building on these theories, research on anthropomorphism has been conducted in diverse environments using AI tools that interact with humans. When humans engage with anthropomorphized AI, trust [19], hedonic connection [11], and emotional attachment [23] can increase. However, anthropomorphism may also blur the distinctions between human intelligence and generative intelligence, potentially leading to negative effects such as unrealistic expectations and ethical concerns [12, 20].

Despite these insights, few studies have examined the relationship between perceived performance and anthropomorphism in the context of GenAI. Additionally, previous research has primarily focused on traditional AI systems, such as chatbots, voice assistants, and social robots [14, 15]. Since interactions with GenAI differ from those with conventional AI [1, 24], further research is needed to refine these factors.

3 METHOD

A total of 15 teams comprising 47 participants took part in the experiment. The group decision-making process was designed using the NASA Exercise [10], consisting of four sequential sessions: individual, group discussion, ChatGPT-3.5, and ChatGPT-4o. In the individual session, participants completed the NASA Exercise independently, using their own knowledge to generate their answer sheets individually. In the group discussion session, each team collaboratively reached a consensus and created a shared answer sheet. Finally, in the ChatGPT sessions, teams interacted freely with ChatGPT-3.5 and ChatGPT-4o, generating additional answer sheets for each version.

All participants evaluated the estimated performance of themselves, other humans, and each version of ChatGPT. Additionally, they responded to their attitudes toward ChatGPTs by selecting from teammate, assistant, coach, manager, or robot/machine, both before and after the experiment.

We scored each team’s answer sheet to determine whether synergy occurred. Based on prior research [13], we defined synergy as occurring when a group’s score exceeded the highest individual or ChatGPT score within that group. Our research questions and hypotheses are as follows:

- RQ1: How does humans’ anthropomorphizing attitude toward ChatGPT relate to its perceived performance?
  - H1: Humans are more likely to anthropomorphize ChatGPT when they perceive its performance exceeds their own.
- RQ2: Is anthropomorphizing ChatGPT associated with improved performance?
  - H2: Participants in the synergy groups are more likely to anthropomorphize ChatGPT than non-synergy groups.

4 RESULT

For clarity, participants who perceive other participants or ChatGPTs as Teammates or Assistants are designated as the TA group. Similarly, those who view them as Coaches or Managers belong to the CM group, while those who regard them as Robots or Machines are classified as the RM group. Detailed statistic summaries are presented in Table 1.

4.1 Perceived Performance and Anthropomorphism

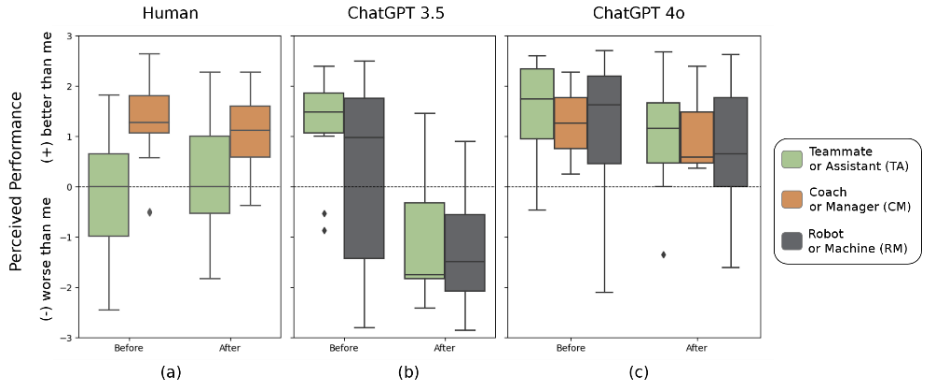

Figure 1: Perceived performance and attitudes toward (a) other humans and (b, c) ChatGPTs before and after the experiment. Participants perceived ChatGPT-4o as superior to ChatGPT-3.5 both before and after the experiment. Moreover, ChatGPT-3.5's perceived performance declined sharply following the interaction.

In the case of human performance perception (Figure 1a), the TA group perceived others' performance as similar to their own, whereas the CM group perceived others as having higher performance. These differences were statistically significant both before ($t = -4.554$, $p < 0.001$) and after the experiment ($p = 0.012$).

For ChatGPT-3.5 (Figure 1b) and ChatGPT-4o (Figure 1c), the TA or TA+CM group perceived its performance as higher than the RM group in all interactions. However, all conditions are not statistically significant. In addition, both before ($p < 0.001$) and after ($p < 0.001$) the experiment, participants perceived ChatGPT-4o as performing better than ChatGPT-3.5. Also, perceived performance dropped significantly in both ChatGPT-3.5 ($p<0.001$) and ChatGPT-4o ($p=0.013$). Specifically, participants believed ChatGPT-3.5 performed worse than themselves. In contrast, ChatGPT-4o maintained the trend of being rated as superior to participants' own performance.

Due to the presence of some non-significant results, we accept H1 with a revision, suggesting that humans' criteria for anthropomorphizing ChatGPT are based on comparisons with other ChatGPT models rather than their own performance.

Revised H1: Humans are more likely to anthropomorphize ChatGPT when they perceive its performance exceeds *that of another ChatGPT model rather than* their own.

Table 1: Statistical summary of perceived performance and attitude distribution across interactions

Interaction	Human				ChatGPT-3.5				ChatGPT-4o			
	Before		After		Before		After		Before		After	
Group	TA	CM	TA	CM	TA	RM	TA	RM	TA+CM	RM	TA+CM	RM
Mean	-0.145	1.218	0.163	1.076	1.282	0.208	-1.153	-1.185	1.495	1.237	0.934	0.698
S.D.	1.059	0.896	0.933	0.915	0.949	1.818	1.236	1.103	0.984	1.301	1.052	1.185
n (%)	86.1	13.9	91.1	8.9	29.8	70.2	21.3	78.7	46.8	53.2	53.2	46.8
Total Mean	0.044		0.244		0.528		-1.178		1.358		0.823	

4.2 Anthropomorphism and Synergy

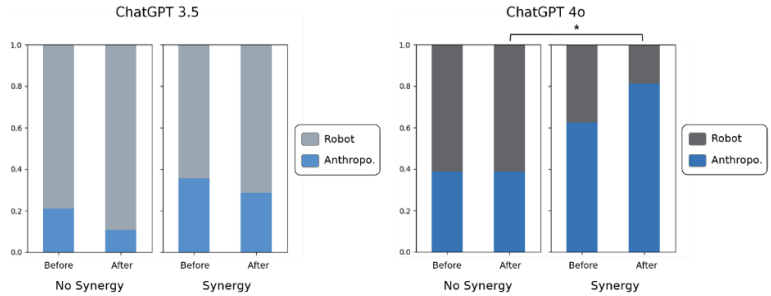

Figure 2: ChatGPT-3.5 had no significant differences in anthropomorphized proportions between the synergy and non-synergy groups. In contrast, for ChatGPT-4o, a higher proportion of participants perceived it as humanlike, with a significant difference observed between the synergy and non-synergy groups after the interaction.

To focus more on anthropomorphizing, we combined the TA and CM groups into a single anthropomorphizing group (ATP). Regardless of synergy, the proportion of ATP participants was higher for ChatGPT-4o than for ChatGPT-3.5, with a statistically significant difference observed after the interaction ($\chi^2 = 8.922$, $p < 0.001$).

We compared ATP proportions within the synergy and non-synergy groups to examine the effect of anthropomorphizing. Generally, the proportion of ATP participants was higher in the synergy group than in the non-synergy group for both versions of ChatGPT. Specifically, ChatGPT-3.5's proportion of ATP participants decreased after the experiment, though the change was not statistically significant. In contrast, ChatGPT-4o's proportion of ATP participants increased in the synergy group after the experiment. Notably, a statistically significant difference emerged between the synergy and non-synergy groups after the interaction ($\chi^2 = 6.057$, $p = 0.014$).

Therefore, due to some non-significant results for ChatGPT-3.5, H2 can be accepted with revisions. It suggests that a higher proportion of anthropomorphizing is associated with synergy, particularly when interacting with ChatGPT, which is perceived as highly performant.

Revised H2: *When interacting with ChatGPTs perceived as highly performant*, participants in the synergy groups are more likely to anthropomorphize ChatGPT than non-synergy groups.

5 DISCUSSION AND CONCLUSION

In summary, when humans interact with GenAI perceived as highly performant, they are more likely to anthropomorphize it. Also, anthropomorphism can enhance team performance, particularly in interactions with high-performance GenAI.

Additionally, attitudes toward humans and ChatGPTs differed. Specifically, when participants perceived another human as more capable, they categorized them as a coach or manager. In contrast, when participants perceived ChatGPT as superior, they viewed it as a teammate or assistant—roles that reflect a higher degree of anthropomorphism. This finding aligns with prior research suggesting that higher perceived intelligence can lead to increased anthropomorphism [14, 16]. Moreover, when ChatGPTs' performance did not meet participants' expectations, many responded by saying, *"It is just a robot!"* In particular, when ChatGPT struggled to understand the environment fully or failed to empathize with humans, participants sometimes underestimated its actual performance.

Although humans interact with ChatGPT solely through text, they naturally develop an anthropomorphized attitude based on comparisons with other GenAI tools rather than human performance. We did not incorporate any anthropomorphic characteristics, yet participants still perceived ChatGPT as humanlike solely due to its highly perceived performance. Therefore, if a GenAI tool's relative perceived performance is designed to be significantly higher than that of other tools, it can still elicit an anthropomorphized attitude, even if its absolute performance remains low.

REFERENCES

[1] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N. and Inkpen, K. Guidelines for human-AI interaction. 2019 In Proceedings of the 2019 chi conference on human factors in computing systems.

[2] Bitkina, O. V., Jeong, H., Lee, B. C., Park, J., Park, J. and Kim, H. K. Perceived trust in artificial intelligence technologies: A preliminary study. Human Factors and Ergonomics in Manufacturing & Service Industries, 30, 4 (2020), 282-290.

[3] Brave, S., Nass, C. and Hutchinson, K. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. International journal of human-computer studies, 62, 2 (2005), 161-178.

[4] Caporael, L. R. Anthropomorphism and mechanomorphism: Two faces of the human machine. Computers in human behavior, 2, 3 (1986), 215-234.

[5] Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly (1989), 319-340.

[6] Duffy, B. R. Anthropomorphism and robotics. The society for the study of artificial intelligence and the simulation of behaviour, 20 (2002).

[7] Eddy, T. J., Gallup Jr, G. G. and Povinelli, D. J. Attribution of cognitive states to animals: Anthropomorphism in comparative perspective. Journal of Social issues, 49, 1 (1993), 87-101.

[8] Epley, N., Waytz, A. and Cacioppo, J. T. On seeing human: a three-factor theory of anthropomorphism. Psychological review, 114, 4 (2007), 864.

[9] Fogg, B. and Nass, C. How users reciprocate to computers: an experiment that demonstrates behavior change. City, 1997.

[10] Hall, J. and Watson, W. H. The effects of a normative intervention on group decision-making performance. Human relations, 23, 4 (1970), 299-317.

[11] Han, S. and Yang, H. Understanding adoption of intelligent personal assistants: A parasocial relationship perspective. Industrial Management & Data Systems, 118, 3 (2018), 618-636.

[12] Ji, J. J. Demystify ChatGPT: Anthropomorphism around generative AI. GRACE: Global Review of AI Community Ethics, 2, 1 (2024).

[13] Larson Jr, J. R. Deep diversity and strong synergy: Modeling the impact of variability in members' problem-solving strategies on group problem-solving performance. Small group research, 38, 3 (2007), 413-436.

[14] Li, M. and Suh, A. Anthropomorphism in AI-enabled technology: A literature review. Electronic Markets, 32, 4 (2022), 2245-2275.

[15] Mariani, M. M., Hashemi, N. and Wirtz, J. Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. Journal of Business Research, 161 (2023), 113838.

[16] Moussawi, S. and Koufaris, M. Perceived intelligence and perceived anthropomorphism of personal intelligent agents: Scale development and validation (2019).

[17] Nass, C. and Moon, Y. Machines and mindlessness: Social responses to computers. Journal of social issues, 56, 1 (2000), 81-103.

[18] Obrenovic, B., Gu, X., Wang, G., Godinic, D. and Jakhongirov, I. Generative AI and human–robot interaction: implications and future agenda for business, society and ethics. AI & SOCIETY (2024), 1-14.

[19] Qiu, L. and Benbasat, I. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. Journal of management information systems, 25, 4 (2009), 145-182.

[20] Raji, I. D., Bender, E. M., Paullada, A., Denton, E. and Hanna, A. AI and the everything in the whole wide world benchmark. arXiv preprint arXiv:2111.15366 (2021).

[21] Reeves, B. and Nass, C. The media equation: How people treat computers, television, and new media like real people. Cambridge, UK, 10, 10 (1996), 19-36.

[22] Salles, A., Evers, K. and Farisco, M. Anthropomorphism in AI. AJOB neuroscience, 11, 2 (2020), 88-95.

[23] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M. and Bolton, A. Mastering the game of go without human knowledge. nature, 550, 7676 (2017), 354-359.

[24] Xu, W., Dainoff, M. J., Ge, L. and Gao, Z. Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI. International Journal of Human–Computer Interaction, 39, 3 (2023), 494-518.