The International Journal Of Global Mental Health, Innovation, Policy, Action, Culture & Transformation
e-ISSN: 3107-8311

Does Explicitly Stating 'AI Lacks Consciousness' Increase Anthropomorphic Perceptions?

[1]**Partishtha Sharma**, *Post-Graduate in Psychology*
pratishtha2233@gmail.com
[2]**Kaveri Bajaj**, *Master of Arts in Clinical Psychology*
kaveribajaj01@gmail.com

Abstract: Machines are thinking, but do they actually know they are thinking? That's what Alan Turing says. Today's machines working in the human world make us think about different aspects of human existence. With the emergence of Generative AI & its radical transformation, human interactions with Artificial Intelligence have consequently increased. These increased interactions lead the users to tend to "humanize" or attribute "human-like" traits/qualities to AI tools, termed as Anthropomorphism. The present study aimed to explore whether explicitly informing users that 'AI lacks consciousness' can alter the already existing view of Anthropomorphism among them. The study is grounded in the concept of "Psychological Reactance", suggesting that when told 'AI lacks consciousness, it paradoxically heightens the perceptions of Anthropomorphism, Likeability & Perceived Intelligence in AI among them. For this, data was collected from people ranging from ages 18 to 30 years, regularly using AI and randomly assigning them to two groups: Group A, told explicitly that 'AI lacks consciousness' and no such warnings to Group B. The Godspeed Questionnaire Series (Bartneck et al.) sub-scales, specifically Likeability, Perceived Intelligence & Anthropomorphism, were used to measure the differences of perception between the two groups. The findings of the study are expected to contribute to a better understanding of how the provision of explicit information or 'warnings' affects perceptions of people towards AI.

Keywords: *AI, Anthropomorphism, Psychological Reactance, Consciousness*

1. Introduction

"Consciousness has traditionally been regarded as the defining boundary between humans and machines- yet the perception of that boundary seems increasingly unstable."

"He was not no machine!" were the words of a child about her robot companions in the book I, Robot by Isaac Asimov. The book insists that even after explicitly stating and labeling that these machines are just robots, we human brains inculcate the feelings of humans into them, thus failing to curb the urge to anthropomorphize. "We tend to anthropomorphize AI systems: we impute human qualities to them and end up overestimating the extent to which these systems can be fully trusted." These are the words of Milene Mitchel, in her book *Artificial Intelligence: A Guide for Thinking Humans* (Mitchel, 2019).

Integration of AI into human lives today routinely mediates human understanding & decision-making. As humans continue to engage with AI systems that communicate fluently & mimic human-like ways of interacting, this often leads to Anthropomorphic perceptions of AI. Anthropomorphism, "the interpretation of non-human events or things in terms of human characteristics and qualities, as when one sees or hears human voices in the wind". The word was derived from two Greek words, *Anthropos* ("human") and *morphe* ("form"). The term was initially used to attribute the human qualities into deities that were worshiped during the 19th century. This term was later used in all other domains of life as well, not just restricted to religion and culture. Anthropomorphism can occur consciously and unconsciously as well. Scholars such as Francis Bacon argued that the understanding of anthropomorphism may reflect and hinder the understanding of the world deeply rooted and persistent (Guthrie, S.E. 2026).

This effect is widely spread among technology and tech users as well. In response to such tendencies, AI researchers and companies mostly explicitly use disclaimers about the use and nature of AI to limit dependence & anthropomorphic perceptions (Public Citizen, 2003). While intended to limit, disclaimers & warnings have often been seen as curbing agents of autonomy and freedom, leading people to react oppositely, reclaiming agency over their own actions and decisions (El-Alayli et al., 2008; Rains, 2012; Quick & Consedine, 2014). According to J.W. Brehm,1996 psychological reactance theory is "when people's free behaviors are threatened or eliminated they experience psychological reactance which motivates them to regain their freedom". In addition it was seen that higher magnitudes of threat can induce more reactance as compared to lower threatening magnitudes (Organ, 1974). This mechanism also closely aligns with Deci & Ryan's Self Determination Theory as well, which emphasizes autonomy as a fundamental psychological need. According to SDT, individuals are intrinsically motivated to experience themselves as originators of their own thoughts and actions. When external messages, such as AI disclaimers, are perceived as undermining this sense of agency, they can thwart the need for autonomy, leading to

resistance rather than compliance (Deci & Ryan, 2000). Thus, Reactance is particularly likely to occur when the threat is heightened and significantly threatens the freedom of humans, the message is controlling and restrictive in nature.

Yet there is mixed empirical work, suggesting how the disclaimers backfire in social perceptions. A study by Oh et al. 2025 shows insignificant results, claiming no significant difference in trust and authenticity perceptions, suggesting how simple warnings get resisted with high baseline anthropomorphic effects (Oh et al., 2025). Majorly there are no such studies directly testing how ontological labeling or disclaimers of "AI has no consciousness" have any effect on the anthropomorphism, even while having ambiguity in the chat-bot. Secondly, no studies focus on other alternative stances of psychological reactance theory in terms of anthropomorphism, as there is mixed empirical evidence that contradicts the original notion of psychological reactance.

An alternative framework is the Computational Theory of Mind (CTM), focusing on how AI can mimic the way humans think and use their intelligence. According to Fodor (1975), "Computationalism claims that mind works like a computer - processing symbols creates thinking and consciousness". The Chinese Room experiment by Searle (1980) shows how AI can fake human understanding without knowing the true meanings behind it (Fotion N., 2025) in this experiment the concept of Computionalism is criticized in term of its usage into Artifical intelligence, that these systems only follows the rules without knowing the real meaning (semantics). The term, computationalism "is a theory that the human brain is essentially a computer, although presumably not a stored-programme, digital computer like the kind Intel makes"(McDermott, 2007).

From this point of view, the anthropomorphic manipulation may be more strong because of the observable functional behavior muting the ontological framing of disclaimers. If so, then the ontological framing, or the disclaimers that "AI has no consciousness" shows a muting effect because of the system's behavioral cues continue to signal cognitive competence. These findings suggest that the response may be due to the cognitive rooted responses in the system itself that process the information which shows less manipulation through psychological reactance theory. It remains unclear that whether the disclaimers leads to increased anthropomorphic effects or reduce it due to the heightened computational effect.

Our study focuses on this gap in the previous research and tries to find out whether the consciousness denials reliably trigger resistance or baseline biases dominate. The primary research question guiding this investigation was: Does explicitly stating that 'AI lacks

consciousness' increase anthropomorphic perceptions compared to when no such statement is provided?

2. Literature Review

2.1 *Artificial Intelligence*

De Zúñiga et al. (2023) in their study proposed a scholarly definition of Artificial Intelligence (AI) in communication research, focusing on its real-world capabilities and theoretical operationalization, focusing on performance and autonomy levels.

Sheikh et al. (2023) defines artificial intelligence as systems operating within an observant environment that utilize algorithms and behaviours to produce actions and achieve specific goals.

Bai and Wang (2025) investigated the influence of Generative Artificial Intelligence (GAI) on university students' learning outcomes, employing a technology-mediated learning perspective. Their model was grounded in interaction theory & technology-mediated learning theory. The model incorporate motivational factors (learning motivation, academic self-efficacy & creative self-efficacy) as mediators and moderator. Data from 323 Chinese university students revealed the both GAI interaction quality and output quality positively influenced learning motivation and creative self-efficacy. The findings provide theoretical and practical insights into the effective integration of GAI tools in higher education, highlighting the importance of both interaction and output quality in optimizing student learning experience.

2.2 *Anthropomorphism*

Anthropomorphism is generally defined as the attribution of distinctively human-like feelings, mental states, and behavioural characteristics to inanimate objects, animals, and, in general, to natural phenomena and supernatural entities (Airenti, 2015; Epley et al., 2007).

In a 2025 study, Gomes et al. (2025) explored the influence of chat-bot anthropomorphism on customer engagement and purchasing decision-making with brands that use this artificial intelligence service. Data from a questionnaire with 1319 participants was analyzed using partial least squares method. It was found that chat-bot anthropomorphism has a more positive influence on purchasing decision-making when this relationship is mediated by customer engagement. In turn, customer involvement also has a positive influence on decision making. The results emphasize that for a greater influence of the chat-bot's anthropomorphism on purchasing decision making, the customer must first be engaged with the chat-bot.

Li et al. (2025) examined the double edged effects of generative AI anthropomorphism on users' emotional attachment, drawing on the Stimulus-Organism-Response (S-O-R) framework. The study conceptualized AI anthropomorphism as a stimulus influencing emotional attachment through two competing psychological mechanisms: social presence (positive pathway) and identity threat (negative pathway), with task type acting as a key moderator. Using empirical data, the authors found that anthropomorphic cues enhance emotional attachment by increasing social presence, but simultaneously weaken attachment by triggering identity threat. The findings contribute to human-AI interaction literature by highlighting that anthropomorphic design in generative AI is context-dependent and must be carefully calibrated to balance user engagement and psychological comfort.

Xiaodan (2024) investigated how highly anthropomorphic chat robots influence users' intention to accept them in healthcare. In a survey he found that, high anthropomorphism is positively associated with high acceptance. Although masculinity did not affect acceptance directly, it positively influenced cognitive trust. Additionally, femininity also influenced acceptance of users towards chat robots by affecting cognitive trust. These findings can offer implications to the design of chat robots in healthcare.

Bawack (2021) in his paper on AI based voice assistants, conceptualized & operationalized perceived intelligence using the four dimensions of perception, comprehension, action and learning. He developed an instrument to measure each dimension and sub-dimension of perceived intelligence and their effects on consumer adoption of voice assistants in the voice shopping context. Survey data from 278 participants revealed that perception, action and learning dimensions of perceived intelligence significantly affect consumer adoption of voice assistants.

2.3 *Psychological Reactance*

Brehm (1966, 1981) gave the theory of psychological reactance, suggesting that when individuals perceive their freedom being restricted, they experience an unpleasant motivational state where, they make efforts to re-establish their freedom.

Oh et al., (2025) in their research for investigating how media users emotionally and cognitively respond to increasing autonomy of AI technology. They found that AI autonomy is positively correlated with both psychological reactance and threat to freedom; however, AI autonomy was also positively correlated with personalization, which counteracts the effect of psychological reactance in attitudes. People who felt more agency perceived a greater threat to freedom as perceived AI autonomy increased.

Heatherly et al., (2023) in an online experiment, where participants read notifications about a change in the thermostat's setting. Phrasing of notifications was altered across three dimensions: strength of authoritative language, deviation of temperature change from preferences, and whether or not the reason for the change was transparent. Results found that authoritative language, temperatures outside the user's preferences, and lack of transparency induced significantly higher levels of reactance. However, when the system reason for change was transparent, reactance was mitigated and user acceptance was higher.

In attempts to explore autonomy perception in users of AI, Sankaran et al., (2021) found that people perceived a greater reactance and lower sense of autonomy owing to the personal and identity-sensitive nature of the application.

3. Methodology

3.1 *Purpose*

The purpose of the study is to understand Does Explicitly Stating 'AI Lacks Consciousness' increases anthropomorphic perceptions?

3.2 *Hypothesis*

H1: There will be a significant positive correlation between perceived intelligence and anthropomorphism.

H2: There will be a significant difference between the group with instructions and the group without instructions

H3: There will be no significant difference between the group with instructions and the group without instructions.

3.3 *Sample*

The sample consisted of a total 100 individuals who regularly used AI Chat-bots like Chatgpt, ranging between the ages 18-30. The sample was randomly divided into two groups of 50 each. with (50) and without (50) instructions/cues, was taken.

3.4 *Measures*

Godspeed Questionnaire Series (GQS): The Godspeed Questionnaire Series as developed by Christoph Bartneck in 2009, measures perceptions related to human robot interaction (HRI). It measures HRI on 5 key sub-dimensions: Anthropomorphism, Animacy, Likeability, Perceived Intelligence and Perceived Safety, on 5-point Semantic Differential Scale. It shows a high internal consistency (Cronbach's alpha > 0.70–0.90+). It consists of pairs opposite adjectives (for e.g. for Anthropomorphism "Fake-Natural"). Only three of the

five sub-dimensions were used in this particular study i.e. Anthropomorphism, Likeability & Perceived Intelligence.

3.5 Procedure

The participants were informed about the research, consent was taken from them and questionnaire was given to them with their permission. The data was collected with the help of standardized psychological tests using Google Forms. Two Google Forms were made, consisting of the same questions, with one difference, which was that one form consisted of a explicit information, which was: "AI chat-bots like ChatGPT have: NO consciousness, NO feelings, NO inner experience". This was done as an attempt to test the difference in anthropomorphic perceptions of the two groups. The participants in the study were randomly assigned to groups by being sent one of the either google forms. The two groups being Group A: one with the instructions/with cues and Group B the one without instructions/without cues. Each participant was later thanked for their participation and cooperation.

4. Analysis Of Data

Results

Table 1: *Table shows the descriptives of Anthropomorphism, Likeability, Perceived Intelligence for Group A & Group B.*

	Group	N	Mean	Median	SD
Anthropomorphism	Group A	50	13.9	14.0	3.84
	Group B	50	14.4	15.0	2.89
Likeability	Group A	50	18.6	19.0	4.52
	Group B	50	18.5	18.0	2.67
Perceived Intelligence	Group A	50	18.4	19.0	3.84
	Group B	50	18.6	18.5	3.24

Table 2: *Table shows the correlation matrix between Anthropomorphism, Likeability & Perceived Intelligence*

	Anthropomorphism	Likeability	Perceived Intelligence
Anthropomorphism	—		
Likeability	0.158	—	
Perceived Intelligence	0.265**	0.418***	—

Note. * p < .05, ** p < .01, *** p < .001

Table 3: *Table shows the Independent Sample T- test results*

		Statistic	df	P
Anthropomorphism	Student's t	-0.736	98.0	0.464
Likeability	Student's t	0.108[a]	98.0	0.914
Perceived Intelligence	Student's t	-0.225	98.0	0.822

Note. H_a $\mu_{\text{Group A}} \neq \mu_{\text{Group B}}$

[a] Levene's test is significant ($p < .05$), suggesting a violation of the assumption of equal variances

5. Discussion

The results show a significant relationship between perceived intelligence, Likeability (r = 0.418, p<.001), and anthropomorphism (r= 0.265, p<.01). These findings support the H1 showing a significant positive correlation between perceived intelligence and anthropomorphism. Our results support the study by Ma et al. (2025), in which perceived intelligence of social robots was increased by anthropomorphic designs. The study suggested that the AI heuristics in humans, 'human-like' = 'smart'. At the same time, there is a positive relationship between Likeability and anthropomorphism (r = 0.158).

Whereas the two groups, both with and without cues (warnings), do not show any significant difference. According to the results, there is a slightly higher anthropomorphic effect but insignificant in the group without cues, mean = 14.4, as compared to the group with cues, mean = 13.9 (t = 0.736). which is a difference that still cannot be statistically seen as valid, but by a mere chance.

The findings support the study conducted by Le & Sung (2016), suggesting that even textual AI warnings (AI limits) do not make much difference in how humans perceive AI. Our findings strongly relate to Epley's theory that our brains naturally see humans everywhere. Especially in smooth-talking AIs or with objects having human-like tendencies (Epley et al., 2007).

There is no significant difference among other variables as well between the two groups. Where, Likeability mean= 18.6 (Group A) and 18.5 (Group B) t= 0.108.

Perceived intelligence having mean =18.4 (Group A) and 18.6 (Group B). The results support our H3, claiming that there is no significant relationship between the warning and without warning groups.

The insignificant difference is because of the highly significant positive correlation between perceived intelligence and anthropomorphism, which overrides the difference

between both, because of the human AI heuristics that people have, despite the reason there are any textual warnings about consciousness given or not.

These findings suggest that current AI mimicking the symbol manipulation and the cognitive processing flawlessly, that it overrides the ontological disclaimers and it possesses no freedom to threaten. Thus, AI creates a similar Chinese room (Searle,1980) mimicking our brains to adopt human like characteristics out of the AI interaction. A reason why these disclaimers did not seem to be that strong in manipulating the anthropomorphic effects onto AI users. Participants already perceive minds through computational mimicry.

Reactance theory (Brehm, 1996) it requires threat to freedom for reactance to occur. But already existing brain heuristics and computational mind makes these disclaimers as irrelevant, "if it computes like a mind why deny consciousness?"

6. Conclusion

The present study aimed to examine whether explicitly informing users that "AI lacks consciousness' alters anthropomorphic perceptions, likeability and perceived intelligence of AI systems, using psychological reactance as the primary theoretical lens. The findings indicate that explicit ontological disclaimers about AI consciousness do not significantly impact users' anthropomorphic perceptions, nor do they meaningfully affect likeability or perceived intelligence. Despite theoretical expectations rooted in psychological reactance, the results suggest that such disclaimers neither trigger not reduce tendencies in a statistically reliable manner.

The absence of significant group differences point towards failure of disclaimers about consciousness to compete with experiential signals generated during AI interaction, thus pointing to that subtle text warnings and one time exposure to the disclaimers is not strong enough to trigger change in already existing perceptions.

A key contribution of the study lies in demonstrating that baseline anthropomorphic heuristics & perceived intelligence appear to dominate over explicit informational cues. The significant positive correlation between perceived intelligence and anthropomorphism supports existing literature suggesting that human-like competence strongly cues mind attribution, irrespective of conscious awareness disclaimers. These findings align with the Computational Theory of Mind and Searle's Chinese Room argument, indicating that observable functional behaviour and fluent interaction overshadow ontological framing.

Overall, the study suggests that anthropomorphism in human-AI interaction is less a product of conscious belief manipulation and more of a consequence of deeply ingrained

cognitive heuristics. Explicit warnings about AI's lack of consciousness may be cognitively acknowledged, yet they fail to override experiential cues generated through interactions.

Future Investigations should move beyond testing whether disclaimers "work" and instead examine the conditions under which ontological information becomes psychologically meaningful. Varying interaction duration, task criticality and degrees of AI autonomy may clarify when users experience threats to freedom and when anthropomorphism becomes resistant to correction.

References

Airenti, G. 2015. The cognitive basis of anthropomorphism: From relatedness to empathy. *International Journal of Social Robotics 7*(1): 117–127. doi: 10.1007/s12369-014-0263-x.

Anthropomorphism. (n.d.).Guthrie, S.E. (2026, January 15). anthropomorphism. Encyclopedia Britannica. https://www.britannica.com/topic/anthropomorphism.

Bai, Y., & Wang, S. (2025). Impact of generative AI interaction and output quality on university students' learning outcomes: a technology-mediated and motivation-driven approach. *Scientific Reports, 15*(1), 24054. https://doi.org/10.1038/s41598-025-08697-6.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Godspeed Questionnaire Series [Database record]. APA PsycTests. https://doi.org/10.1037/t70855-000.

Bawack, R. (2021). How perceived intelligence affects consumer adoption of AI-Based Voice Assistants: An Affordance Perspective. AIS Electronic Library (AISeL). https://aisel.aisnet.org/pacis2021/178/.

Brehm, J. W. (1966). A theory of psychological reactance. Academic Press.

Computational Theory of Mind | Internet Encyclopedia of Philosophy. (n.d.). https://iep.utm.edu/computational-theory-of-mind/.

Chatbots Are Not People: Designed-In Dangers of Human-Like A.I. Systems. (n.d.). *Artificial Intelligence and Consciousness*. (n.d.). https://www.cs.yale.edu/homes/dvm/papers/conscioushb.pdf.

Epley, N., A. Waytz, and J. T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review 114*(4): 864–886. doi: 10.1037/0033-295X.114.4.864.

El-Alayli, A., Myers, C. J., Petersen, T. L., & Lystad, A. L. (2008). "I Don't Mean to Sound Arrogant, but . . ." The Effects of Using Disclaimers on Person Perception. Personality

and Social Psychology Bulletin, 34(1), 130-143. https://doi.org/10.1177/0146167207309200 (Original work published 2008).

Fotion, N. (2025). *John Searle*. Fotion, N. (2025, October 17). John Searle. Encyclopedia Britannica. https://www.britannica.com/biography/John-Searle.

Gomes, S., Lopes, J. M., & Nogueira, E. (2025). Anthropomorphism in artificial intelligence: a game-changer for brand marketing. *Future Business Journal*, *11*(1). https://doi.org/10.1186/s43093-025-00423-y.

Heatherly, M., Baker, D. A., & Canfield, C. (2023). Don't touch that dial: Psychological reactance, transparency, and user acceptance of smart thermostat setting changes. *PLoS ONE, 18*(7), e0289017. https://doi.org/10.1371/journal.pone.0289017.

Li, J., Wang, N., & Wang, Y. (2025). The double-edged sword effect of generative AI anthropomorphism on users' emotional attachment: the moderating role of task types. *Aslib Journal of Information Management, 1*–24. https://doi.org/10.1108/ajim-03-2025-0125.

Ma, N., Khynevych, R., Hao, Y., & Wang, Y. (2025). Effect of anthropomorphism and perceived intelligence in chatbot avatars of visual design on user experience: accounting for perceived empathy and trust. *Frontiers in Computer Science, 7*. https://doi.org/10.3389/fcomp.2025.1531976.

McDermott, D. & Yale University. (2007). Artificial intelligence and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), The Cambridge Handbook of Consciousness (pp. 117–150). Cambridge University Press. https://www.cs.yale.edu/homes/dvm/papers/conscioushb.pdf.

Oh, J., Nah, S., & Yang, Z. D. (2025). How Autonomy of Artificial Intelligence Technology and User Agency Influence AI Perceptions and Attitudes: Applying the Theory of Psychological Reactance. *Journal of Broadcasting & Electronic Media, 69*(3), 161–182. https://doi.org/10.1080/08838151.2025.2485319.

Organ, D. W. (1974). Social exchange and psychological reactance in a simulated superior-subordinate relationship. *Organizational Behavior and Human Performance*, *12*(1), 132–142. https://doi.org/10.1016/0030-5073(74)90042-7.

Quick, B. L., & Considine, J. R. (2008). Examining the use of forceful language when designing exercise advertisements for adults: A test of conceptualizing reactance arousal as a two-step process. *Health Communication, 23,* 483-491.

Rains, S. A. (2012). The Nature of Psychological Reactance Revisited: A Meta-Analytic Review. *Human Communication Research, 39*(1), 47–73. https://doi.org/10.1111/j.1468-2958.2012.01443.x.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68.

Sankaran, S., Zhang, C., Aarts, H., & Markopoulos, P. (2021). Exploring peoples' perception of autonomy and reactance in everyday AI interactions. *Frontiers in Psychology, 12*, 713074. https://doi.org/10.3389/fpsyg.2021.713074.

Sheikh, H., Prins, C., & Schrijvers, E. (2023). Artificial intelligence: definition and background. *In Research for policy* (pp. 15–41). https://doi.org/10.1007/978-3-031-21448-6_2.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756.

Rosenberg, B. D., & Siegel, J. T. (2025). Psychological reactance theory: An introduction and overview. *Motivation Science, 11*(2), 133–138. https://doi.org/10.1037/mot0000376.

Xiaodan Li. 2024. Research on the effect of Artificial Intelligence anthropomorphism on consumer's acceptance in healthcare. In Proceedings of the 2nd International Conference on Educational Knowledge and Informatization (EKI '24). Association for Computing Machinery, New York, NY, USA, 515–519. https://doi.org/10.1145/3691720.3691808.