

Galactic Companions from Star Wars: C-3PO or R2-D2? Understanding Dual Types of Anthropomorphism in Humanoids

Joohee Kim
Yonsei University
jzk5664@gmail.com

Il Im
Yonsei University
il.im@yonsei.ac.kr

Abstract

Recent advancements in generative AI and humanoid robotics have facilitated the rapid development of humanoids that think, move, and interact in ways that closely resemble human behavior, significantly contributing to the study of anthropomorphism in human-AI interaction. Despite the extensive research in this area, the dynamic nature of anthropomorphism—how it evolves through cognitive processes and the effects of these changes—remains largely unexplored. Therefore, we propose dual types of anthropomorphism: reflexive anthropomorphism (automatic attribution of humanlike characteristics) and reflective anthropomorphism (conscious attribution of humanlike qualities) based on the dual process theory of higher cognition. A preliminary study was conducted with 120 participants using a scenario survey that included images of humanoids with and without human-like faces, as well as a video of a simulated chat interface between the user and a chatbot. The study result highlights that an AI agent's appearance significantly influences both reflexive and reflective anthropomorphism, with reflective anthropomorphism being particularly important for shaping users' anthropomorphic responses.

Keywords: Reflexive anthropomorphism, reflective anthropomorphism, artificial intelligence, humanoid, robot

1. Introduction

Earlier research on humanoids often focused on their physical similarity to humans in examining human-robot interaction, primarily because the technology for physical embodiment was more advanced than that for humanlike intelligence (Yogeeswaran, 2016; Bartneck et al., 2010; Hegel et al., 2008). With recent strides in generative AI, however, we are witnessing a remarkable convergence: humanoids are not only embodying more humanlike physical characteristics but also exhibiting

cognitive abilities that appear akin to humanlike cognition, as demonstrated by Tesla's latest creation, Optimus Gen 2 (Su, 2023). This advanced humanoid robot, engineered to efficiently tackle repetitive human tasks (Lambert, 2024), signifies a significant milestone in the fusion of AI and physical embodiment.

The intriguing aspect of studying humanoids with generative AI intelligence lies in the natural inclination of humans to ascribe humanlike characteristics or social traits to these entities (Nass and Moon, 2020; Epley et al., 2007; Gong, 2008), also known as anthropomorphism (attributing human likeness to non-human agents) in human-AI and human-robot interaction (Epley et al., 2007; Takayama, 2009; Kim et al., 2024). Numerous early studies have highlighted the critical role of anthropomorphism, particularly in its complex function as a moderator in human-AI interactions, where it influences relationships and ethical considerations, often mitigating negative impacts and shaping user perceptions (Waytz et al., 2010; Kim and Im, 2023). However, the differences in how anthropomorphism is perceived during initial, immediate interactions with AI agents compared to how these perceptions evolve over extended, long-term engagements have not been thoroughly explored in existing research.

A compelling example of the importance of understanding changes in perception can be seen in the Star Wars saga through the iconic characters C-3PO and R2-D2. C-3PO is a humanoid protocol droid that is designed to resemble human physical features with a humanlike face. In contrast, R2-D2 is a cylindrical robot with a domed head and three legs, appearing less humanlike (Figure 1). At first glance, C-3PO may appear more humanlike due to his humanoid physical embodiment, in contrast to R2-D2's mechanical design. However, despite their distinct appearances, both droids exhibit highly humanlike cognitive abilities—C-3PO with his expertise in translation and diplomacy, and R2-D2 with his problem-solving skills and technical acumen. This illustrates how judgments based solely on appearance can lead to over-attribution of human-likeness, as seen with C-3PO, and under-attribution, as



Figure 1. Picture of C-3PO (left) and R2D2 (right)

seen with R2-D2, highlighting the importance of considering more than just physical characteristics in assessing anthropomorphism.

This example underscores the importance of distinguishing and understanding between the initial attributions of humanlike qualities to humanoid robots and how these attributions evolve through interaction. It is essential to separately analyze how the over- or under-attribution of humanlike attributes, both prior to and following interaction, influences the development of anthropomorphism. Therefore, we ask: *"How do initial perceptions of appearance and intelligence evolve through interaction with AI agents, and how does this evolution shape the anthropomorphism of robots and humanoids?"* To address this research question, we apply the dual-process theory of higher cognition, a framework well-supported by foundational research in psychological science (Evans, 2007; Evans, 2008; Hecht et al., 2012). Based on this theory, we propose two corresponding forms of anthropomorphism: reflexive anthropomorphism, driven by Type 1 processing (which is characterized by implicit, automatic responses requiring minimal cognitive effort), and reflective anthropomorphism, guided by Type 2 processing (which involves explicit, deliberate thinking that engages working memory).

By leveraging the dual model of higher cognitive theory, we aim to understand how anthropomorphism evolves in response to factors such as appearance and intelligence, distinguishing between reflexive (Type 1 processing) and reflective (Type 2 processing) anthropomorphism. Furthermore, this study examines the relationship between anthropomorphic responses—referring to users' perceptions and behaviors toward AI agents, including connection, helpfulness, trust, empathy, and satisfaction—and perceived anthropocentric threat (PAT), which arises when AI systems are perceived to surpass human capabilities,

thus challenging the foundational principles of anthropocentrism (Kim and Im, 2024). Specifically, PAT can be broken down into three key dimensions: intelligence superiority threat (when AI surpasses human capabilities), autonomy superiority threat (when AI violates human authority), and ethical superiority threat (when AI transgresses human ethics and privilege). By investigating these specific dimensions of PAT, we contribute to a deeper understanding of how these perceived threats influence users' anthropomorphic responses and the ethical challenges that emerge from anthropomorphism. Furthermore, as the first study to empirically apply and examine both reflexive and reflective anthropomorphism, we aim to offer a more nuanced framework for understanding how individuals perceive and interact with humanoids equipped with AI intelligence. This approach allows us to explore the automatic, instinctive responses (reflexive anthropomorphism) as well as the more deliberate, conscious evaluations (reflective anthropomorphism) that shape human-AI interactions. By doing so, we seek to advance current models of anthropomorphism in the context of AI, offering theoretical and practical contributions to the fields of human-computer interaction and AI ethics.

2. Dual Types of Anthropomorphism

The concept of a dual process of thinking—one fast and instinctive, the other slow and analytical—has deep historical roots and is widely recognized in both philosophical and psychological literature (Tversky and Kahneman, 1970; Evans, 1990; Solman, 1996). This theory has been widely used in research because it effectively addresses the complexity and diversity of cognitive processes by distinguishing two systems of thought, known as System 1 (fast and instinctive) and System 2 (slow and analytical) (Evans, 2007). Among the various types of proposed dual-process models of thinking (Table 1), we focus on understanding anthropomorphism through the lens of dual-process theories of higher cognition (Evans and Stanovich, 2013). This approach is particularly relevant because anthropomorphism—attributing humanlike qualities to non-human agents—aligns with the dual-process theories' emphasis on the interplay between intuitive and reflective cognitive processes. These theories suggest that our cognitive system allows us to quickly attribute human characteristics based on initial impressions while also enabling more reflective consideration, where we may reassess or refine these attributions based on further interaction or analysis (Zlotowski, 2018). By applying dual-process theories of higher cognition by Evans and Stanovich (2013), we can better understand the cognitive mechanisms behind

Table 1. Summary of Dual Process of Thinking

Authors	Model	First Term Name	Second Term Name
Stanovich (1999)	Dual System	System 1	System 2
Evans and Stanovich (2013)	Dual Process Theories of Higher Cognition	Type 1 processes	Type 2 processes
Fodor (1983, 2001)	The Modularity of Mind	Input Modules	Higher Cognition
Schneider & Schiffrin (1977)	Two Process Theory in Psychology	Automatic	Controlled
Chaiken (1980), Chen & Chaiken (1999) / Evans (1989, 2006), Hammond (1995), Nisbett et al. (2001)	Dual Process Theories in Social Psychology	Heuristic /Intuitive / Holistic	Systematic / Analytic
Reber (1993), Evans & Over (1996)	Implicit Learning / Dual Process Theory of Thinking	Implicit	Explicit
Lieberman (2003)	Dual-process models In Neuroscience	Reflexive	Reflective

anthropomorphism, particularly how initial perceptions and deeper cognitive evaluations work together to shape how we perceive and relate to AI and humanoid robots.

Building on this understanding of cognitive processes, we propose that anthropomorphism is built based on two types: reflexive and reflective. As the terms suggest, we define **reflexive anthropomorphism** as an *automatic cognitive response triggered by the immediate impression of an AI or robotic agent, leading to the instant attribution of human-like qualities to the agent based on its surface-level behaviors and appearance*. Conversely, we define **reflective anthropomorphism** as a *deliberate and thoughtful cognitive process in which a person consciously attributes human-like qualities to an AI or robotic agent, based on a deeper analysis of its appearance, actions, and interactions*. We propose that the perception of anthropomorphism results from a combination of reflexive and reflective anthropomorphism.

We emphasize that dual types of anthropomorphism are not associated with temporal changes in perception, such as acclimatization to usage, familiarity, or habituation. Instead, they represent distinct cognitive processes that influence how individuals perceive and interact with AI. Consider evaluating a virtual AI agent's intelligence, especially when the agent lacks a physical form. Initial perceptions are influenced by elements like speech style, use of emoticons, and punctuation in greetings. As the conversation progresses, users might notice additional traits, revealing the AI's capacity for greater intelligence or silliness than initially assumed. The initial perception during the greeting exemplifies reflexive anthropomorphism, while the subsequent perception following interaction reflects reflective anthropomorphism. In other words, we argue that reflexive and reflective anthropomorphism are not mutually exclusive or sequentially replaced. Rather,

they coexist and can dynamically interact, each influencing the other in shaping the overall perception of AI over time.

3. Hypotheses Development

Factors that are immediately obvious and recognizable at first glance trigger an automatic response, aligning closely with Type 1 processing, which is fast, intuitive, and heuristic-based. When we first encounter a humanoid robot or virtual agent, our initial judgments are often formed quickly, driven by its explicit appearance or the greeting language it uses. This reflexive reaction is guided by visual cues or statements that our brains process quickly without deliberate thought (Evan, 2008; Stanovich and West, 2000). Therefore, we posit that *H1: perceived humanlike appearance is positively associated with reflexive anthropomorphism, and H2: perceived humanlike intelligence is positively associated with reflexive anthropomorphism*.

When users interact with an AI or a humanoid, they do not immediately form a complete judgment about its appearance and intelligence. Instead, their understanding evolves through a series of interactions where they observe and analyze the AI's responses, movements, problem-solving abilities, and overall performance in various tasks. For example, in a conversation with an AI, users might initially note the AI's use of grammar, vocabulary, and contextual relevance, which is associated with reflexive (type 1 processing). As the interaction progresses, they may test the AI's ability to handle complex questions, understand nuanced topics, or adapt to different conversational styles, which involves a reflective thinking process. Therefore, we posit that *H3: perceived humanlike appearance is positively associated with reflective anthropomorphism, and H4: perceived humanlike*

intelligence is positively associated with reflective anthropomorphism.

In the context of the dual attitudes model (Wilson et al., 2000), reflective attitudes can gradually replace reflexive attitudes through cognitive restructuring and intentional reflection. When individuals become consciously aware of a discrepancy between their reflexive and reflective attitudes—often prompted by introspection, education, or social influences—they can actively work to modify their reflexive attitudes. This process requires consistent engagement in behaviors and thought patterns that align with the reflective attitude, thereby reshaping the automatic responses that form the foundation of reflexive attitudes. Over time, these repeated efforts weaken the initial automatic reflexive associations, allowing the reflective attitude to take precedence. Consequently, reflective anthropomorphism is expected to foster greater alignment with reflexive anthropomorphism, suggesting that deliberate, conscious perceptions of humanlike qualities in AI agents will eventually shape and influence automatic, instinctive responses. Therefore, we posit that *H5: the user's reflective anthropomorphism in humanoid robots is positively associated with reflexive anthropomorphism.*

Humans tend to view humanlike humanoids as more useful than machine-like humanoids, as they are perceived as more relatable and intuitive, making interactions more seamless and natural (MacDorman, 2006). This is because human-likeness activates schema congruence, meaning the product fits well within the existing cognitive frameworks related to humanlike objects (Schneider et al., 2019; Guingrich and Graziano, 2024). This makes the product easier to understand and relate to, as users can utilize their extensive experience and understanding of human behavior and traits (Goudey and Bonnin, 2016). Also, humanlike products are processed more smoothly and quickly, leading to favorable evaluations and a higher perception of usefulness (Kim and McGill, 2011). Thus, integrating

humanlike characteristics into products can lead to a higher anthropomorphic response by making them more relatable, understandable, and engaging. Therefore, we posit that *H6: Users' perception of reflexive anthropomorphism in humanoid robots is positively associated with their anthropomorphic response.*

Higher perceived humanlike intelligence in humanoid robots is positively associated with perceived utility because it enhances the robot's ability to perform complex tasks, adapt to user needs, and facilitate intuitive interactions. When users perceive a humanoid robot as possessing humanlike intelligence, they believe it can understand and respond to their commands more effectively, solve problems autonomously, and learn from interactions to improve its performance over time. This adaptability makes the robot seem more competent and reliable, increasing its overall usefulness in various scenarios. Additionally, humanlike intelligence facilitates smoother and more natural communication, as users can interact with the robot in ways like human interactions, reducing the learning curve and increasing satisfaction (Bartneck et al., 2009). Based on the studies, we posit that *H7: Users' higher perception of reflexive anthropomorphism in humanoid robots is positively associated with their anthropomorphic response.*

Most notably, humanlike humanoids, with their close resemblance to humans, can trigger the "uncanny valley" effect, where slight imperfections in appearance or behavior cause feelings of eeriness and discomfort (Mori et al., 2012; Kätsyri et al., 2015). This discomfort is rooted in our cognitive and emotional responses. Human brains are finely tuned to detect and interpret human faces and behaviors. When a humanoid robot closely mimics these but falls short in subtle ways, it creates a mismatch between what we expect and what we perceive, leading to cognitive dissonance (Stein & Ohler, 2017). This dissonance can trigger a fear response. Based on previous studies, we posit that *H8: Users' perception of reflexive anthropomorphism in*

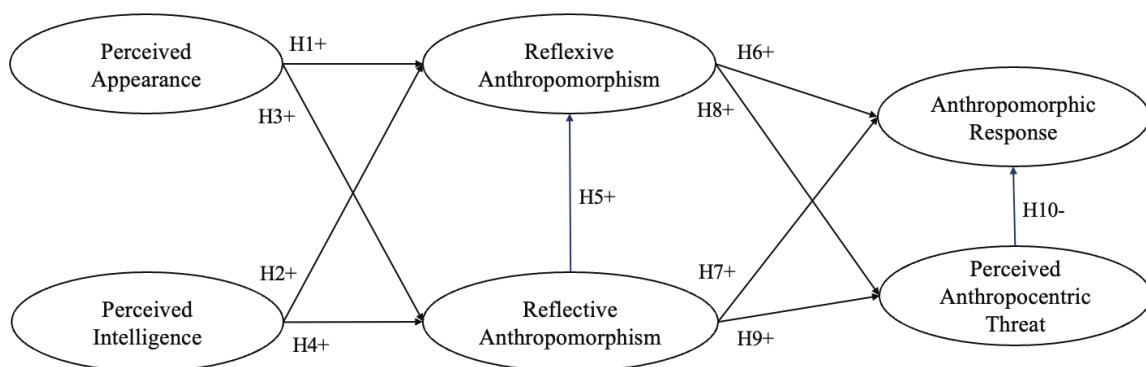


Figure 2. Research Model

humanoid robots is positively associated with their perceived anthropocentric threat.

When robots or AI systems exhibit high levels of humanlike intelligence, they evoke a stronger sense of competition with humans, triggering fears related to job displacement and redundancy. Employees, for instance, may feel that these highly intelligent robots can perform tasks more efficiently and reliably, which leads to concerns about their own job security and potential unemployment (Huang, 2021). Furthermore, the psychological response to highly intelligent anthropomorphic robots often includes hypervigilance and anxious arousal, as these robots are seen as realistic threats not only to jobs but also to safety and resources (Huang, 2021; Kim and Im, 2024). Therefore, we also posit that *H9: The user's higher perception of reflective anthropomorphism in humanoid robots is positively associated with their perceived anthropocentric threat.*

When users perceive a humanoid robot as a threat, it triggers a fear response that overshadows any potential benefits the robot may offer. This threat perception can stem from concerns about job displacement, loss of control, privacy invasion, and physical safety (Kim and Im, 2024). Moreover, perceived threats to personal safety and privacy, especially with advanced AI capable of autonomous decision-making, can make users wary of deploying such robots in everyday environments. This wariness reduces the perceived utility because the robot is seen as more of a liability than an asset. Thus, we posit that *H10: Higher perceived anthropocentric threat is negatively associated with anthropomorphic response.*

4. Study Design

4.1. Measurement and Scenario Validation

Most of the measurements in this study are borrowed from earlier studies, except for reflexive and reflective anthropomorphism. To develop measurement items for reflexive and reflective anthropomorphism, we followed established guidelines for developing, refining, and validating measurement scales as described by Mackenzie et al. (2011) and Straub (1989). A professional expert reviewed the measurement wording and ensured all necessary aspects were covered. Subsequently, three Information Systems (IS) PhD students engaged in a card sorting exercise to categorize and refine the items. This process successfully assessed and confirmed both face and content validity.

We recruited 14 Master's and PhD students with experience in generative AI to validate scenarios for our empirical study. The purpose was to assess whether the provided scenarios varied sufficiently (high vs. low

intelligence) and if the accompanying images appropriately represented the intended variance (very humanlike vs. less humanlike). As an incentive, participants were offered a \$2.50 coffee coupon. The results indicated that 92.3% of participants recognized a clear difference between the two scenarios, with the high intelligence scenario being perceived as significantly more humanlike. On a 10-point scale, the high intelligence scenario received an average score of 8.14 (median = 8) for human-likeness, while the low intelligence scenario scored an average of 3.57 (median = 3). Additionally, regarding the AI agent's images, all 14 participants perceived a significant variance between the two types of AI agents, one with a humanlike face and one without (Figure 3).

5. Preliminary Study and Result

5.1. Data collection and Demographics

For the preliminary study, we designed a scenario-based survey to evaluate our research model. The survey was developed using Qualtrics and featured videos illustrating interactions between a humanoid and a user. **To ensure the quality of our sample, we recruited approximately 130 participants through Prolific,** leveraging the platform's pre-screening function to select individuals with an approval rate of 99% to 100%. The survey comprised eight distinct scenarios, and participants were randomly assigned to one of four scenarios (Appearance High vs. Low x Intelligence High vs. Low) involving a humanoid without a humanlike face or to one of four parallel scenarios featuring a humanoid with a humanlike face (Figure 4).

These scenarios covered a broad range of topics, starting with basic inquiries and progressing to more complex issues, such as exploring solutions for global poverty and the associated emotions. The primary goal of using such varied scenarios was to generate sufficient variance in participant responses, rather than to directly compare differences between scenario groups. Data analysis was conducted on the responses of 120 participants who successfully passed the attention check, out of the initial 130 recruited. The demographics of these participants are as follows: The final sample comprised 52.7% males, with participants ranging in age from 18 to 64. A notable 95% of respondents reported having experience with generative AI, and over 64.8% held an associate degree or higher.

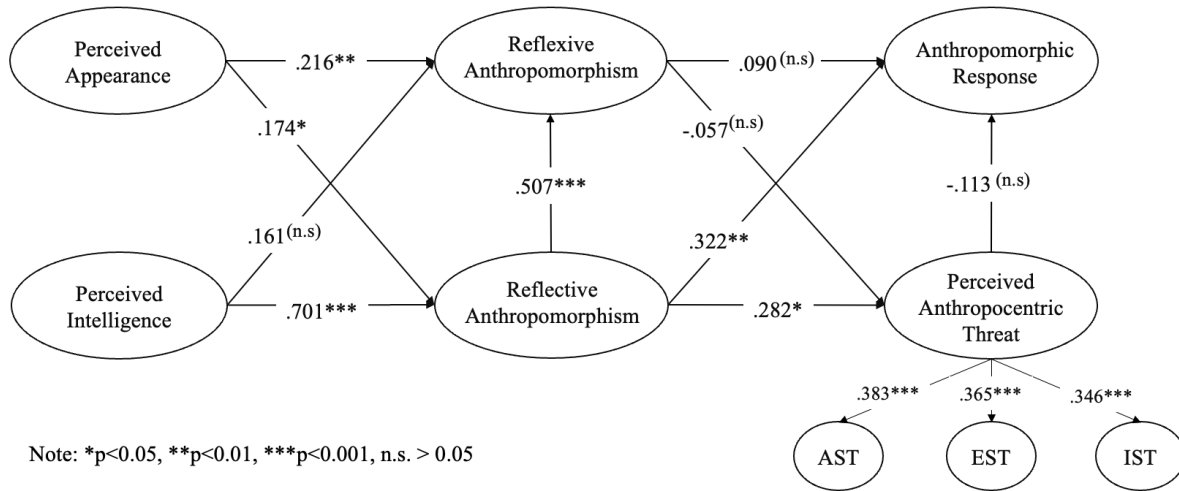


Figure 3. Research model with result

5.2. Measurement Model Testing

An exploratory factor analysis (EFA) was conducted using SPSS Version 26 with varimax rotation to assess the construct validity of the measurement items. Initially, items with low factor loadings (below 0.4) or high cross-loadings (greater than or equal to 0.6) were removed from further analysis. Specifically, the items REC 1, 3, 4, 12 and REX 4, 5, 8, 10, 11, 12 were excluded based on these criteria. This refinement resulted in a total of 14 items, which effectively measured two distinct factors. These two factors accounted for 71.4% of the total variance, demonstrating the model's robust explanatory power. Notably, the first factor alone explained 40.4% of the variance, suggesting that common method bias is not a significant concern in this analysis.

The rotated component matrix, utilizing varimax rotation, clearly revealed the presence of two factors, each with eigenvalues greater than 1, supporting the dimensionality of the constructs. In a subsequent principal component factor analysis, additional items with outer loadings of less than 0.7 were also removed to ensure the robustness of the factor structure. The items removed at this stage included AR1 (0.680) and PCI3 (0.661). Barlett's test of sphericity was significant ($p < 0.001$), confirming the appropriateness of the data for factor analysis. Additionally, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.902, which is well above the recommended threshold of 0.800, further indicating that the dataset was suitable for the factor analysis.

To evaluate our measurement model, we applied a series of reliability and validity tests to ensure the robustness of our constructs. Convergent validity was

first assessed by examining the factor loadings of each item, with a loading of 0.60 or higher considered acceptable. Additionally, we calculated the average variance extracted (AVE) for each construct, adhering to Fornell and Larcker's (1981) guideline that AVE should exceed 0.50 to confirm adequate convergent validity. Our results showed that all constructs met this criterion. Discriminant validity was evaluated by comparing the square root of the AVE for each construct to the correlations between constructs and by examining the cross-loadings of items. The AVE for each construct surpassed the minimum threshold of 0.50, indicating strong discriminant validity across the model. For reliability assessment, we measured both Cronbach's alpha and composite reliability. Both metrics are required to meet or exceed the 0.70 threshold for acceptance, and as presented in Table 2, all scale measures demonstrated reliability well above the recommended values.

Table 2. Construct reliability and validity

	Cronbach's alpha	(rho_a)	(rho_c)	(AVE)
AR	0.934	0.935	0.944	0.655
PA	0.860	0.860	0.899	0.641
PI	0.911	0.912	0.927	0.584
REC	0.933	0.934	0.947	0.749
REX	0.927	0.929	0.940	0.663

Table 3. Discriminant validity

	AR	PA	PI	REC	REX
AR					
PA	0.068				
PI	0.635	0.143			
REC	0.516	0.171	0.745		
REX	0.493	0.299	0.536	0.691	

Table 4. Reflexive Anthropomorphism Measurement Items

REX2	My first impression is that the AI/robot thinks in a human-like manner.
REX5	My quick observation is that the AI/robot experiences emotions like a human.
REX6	My rapid perception is that the AI/robot sounds like a human.
REX7	My first perception is that the AI/robot interacts like a human.
REX8	My prompt observation is that the AI/robot expresses itself like a human.
REX9	My swift impression is that the AI/robot communicates like a human.
REX10	My initial observation is that the AI/robot evaluates situations like a human.
REX11	My quick perception is that the AI/robot interprets things like a human.

Table 5. Reflective Anthropomorphism Measurement Items

REC1	Upon careful reflection, I perceive that the AI/robot resembles a human.
REC2	Upon considered analysis, I perceive that the AI/robot thinks like a human.
REC3	Upon introspective thought, I perceive that the AI/robot acts like a human.
REC6	Upon considered thought, I perceive that the AI/robot sounds like a human.
REC7	Upon evaluated assessment, I perceive that the AI/robot interacts like a human.
REC9	Upon measured consideration, I perceive that the AI/robot communicates like a human.

5.3. Hypotheses Testing

We selected PLS-SEM analysis (bootstrap = 5000) as more suitable to perform exploratory research as a basis for theory development (Hair et al., 2011). The results confirmed H1, H2, H4, H7, H8, and H9 (Table 6). Hypotheses H3, H5, and H10 were not supported, as the direction of the effects aligned with our expectations, but the p-values were not statistically significant. The three dimensions of Perceived Anthropocentric Threat—Intelligence Superiority Threat (IST), Autonomy Superiority Threat (AST), and Ethical Superiority Threat (EST)—each contributed to PAT to a similar extent.

Table 6. Path Coefficient

Path Coefficient	P values	Total Indirect Effect	P values
PA → REC	0.011		
PA → REX	0.002	PA → REX	0.011
PI → REC	0.000		
PI → REX	0.150	PI → REX	0.000
REX → AR	0.346	REX → AR	0.718
REX → PAT	0.646		
REC → AR	0.002	REC → AR	0.773
REC → PAT	0.011	REC → PAT	0.657
REC → REX	0.000		
PAT → AR	0.123		
		PA → AR	0.015
		PA → PAT	0.406
		PI → AR	0.000
		PI → PAT	0.012

6. Discussion

The study's findings underscore the dual influence of an AI agent's appearance on both reflexive and reflective anthropomorphism. In contrast, the AI agent's perceived intelligence primarily impacted reflective anthropomorphism, as indicated by a relatively high coefficient. This suggests that perceived intelligence is strongly associated with reflective anthropomorphism, while reflexive anthropomorphism is predominantly influenced by perceived appearance. These results provide partial confirmation of the dual pathways through which an AI agent's appearance and intelligence contribute to anthropomorphism. In addition, reflective anthropomorphism plays a critical role in determining users' anthropomorphic response towards the AI, as it influences key aspects such as connection, helpfulness, trust, empathy, and satisfaction (Kim & Im, 2023). However, this deeper cognitive engagement can also lead to increased perceived threats, where users become more aware of potential risks.

The study reveals that the anthropomorphic response (AR) was notably higher when robots exhibited low perceived appearance (PA) but high perceived intelligence (PI), regardless of whether the robot had a face, with AR scores of 5.19 and 5 compared to other conditions (4.88 and 4.41). Additionally, the impact on AR was most significant when both PA and PI were high, particularly for robots with a face where the average AR was higher compared to robots with only a body (5.29 vs. 4.64). Interestingly, when only the body was present, AR was highest with low PA and high PI, while for robots with a face, the highest AR was observed when both PA and PI were elevated. This implies that robots without faces can still achieve positive AR even with a non-humanlike appearance

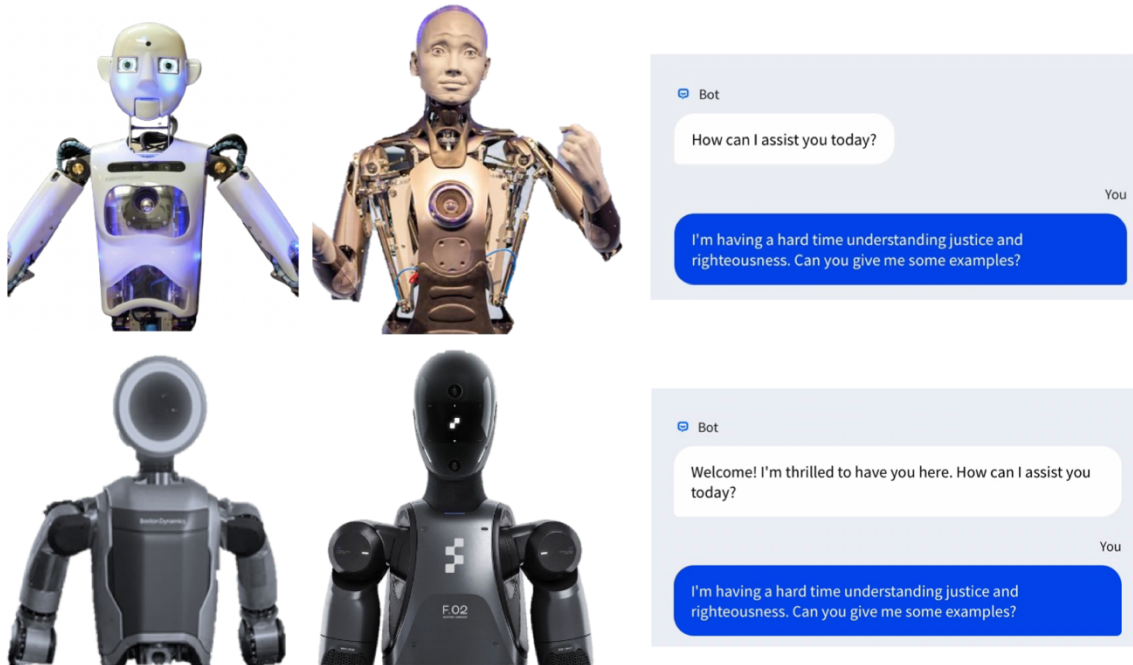


Figure 4. Visual Stimuli and Scenario Design in the Survey

whereas robots with faces must have a highly humanlike appearance to elicit positive AR. These findings underscore the complex interplay between appearance, intelligence, and the presence of facial features in shaping anthropomorphic responses to AI.

In scenarios where both PA and PI were low, robots with a human face elicited a higher average perceived anthropocentric threat (PAT) than those without a face (3.07 vs. 2.86). Moreover, perceived threat reached its peak when both perceived appearance and perceived intelligence of the AI agent were high, surpassing scenarios where only one of these factors was high or both were low. This implies that the combination of a highly capable appearance and intelligence in AI agents amplifies the sense of threat perceived by individuals. The interaction between these two attributes likely heightens concerns about the AI's potential power or control, as both a strong appearance and intelligence may signal greater competence or autonomy.

Among the control variables—age, race, education, knowledge and familiarity with generative AI, self-efficacy in using generative AI, and AI aversion—only AI aversion demonstrated a statistically significant effect. This suggests that individuals' negative predispositions toward AI played a more influential role than demographic factors or levels of AI familiarity and competence in shaping the outcomes.

7. Contributions

Our study enhances the anthropomorphism framework by distinctly separating reflexive and

reflective anthropomorphism. This distinction offers a more nuanced understanding of how individuals perceive and interact with AI humanoids. Notably, this study is the first to empirically examine both reflexive and reflective anthropomorphism, underscoring the critical importance of reflective anthropomorphism in these interactions. By elucidating the differences in how appearance and intelligence influence anthropomorphism, we can better understand the underlying mechanisms, such as the threat perception associated with highly intelligent AI-humanoids. This insight explains why certain AI or humanoid characteristics may elicit varying degrees of acceptance or resistance from users.

For its practical implications, our research offers significant contributions to the field of human-robot interaction by providing detailed guidelines for designing AI systems that interact seamlessly with humans. First, to optimize AI design, developers should prioritize the visual appearance of AI agents to engage users immediately and evoke humanlike perceptions. A well-crafted, visually appealing design plays a critical role in capturing users' attention and establishing an instant connection, which is vital for both initial and ongoing interactions. In scenarios where long-term engagement is crucial, such as in healthcare or customer service, enhancing the perceived intelligence of AI agents becomes more important than visual appearance. By focusing on intelligence, developers can foster deeper reflective anthropomorphism, where users thoughtfully ascribe humanlike qualities to the AI,

thereby strengthening user connection and trust over time.

8. Limitations and Future Study

A limitation of this study is the use of videos and images to assess the perceived appearance and intelligence of the AI agent. Future research should consider employing actual humanoids to obtain more accurate measurements of these perceptions. Furthermore, additional studies are necessary to examine reflexive and reflective anthropomorphism in the context of direct interactions with humanoids. Based on the limitations of this study, we are planning to conduct a main study.

9. References

- Airenti, Gabriella. 2015. "The Cognitive Bases of Anthropomorphism: From Relatedness to Empathy." *International Journal of Social Robotics* 7:117–27.
- Bartneck, C., Bleeker, T., Bun, J., Fens, P., & Riet, L. (2010). The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn*, 1(2), 109–115. <https://doi.org/10.2478/s13230-010-0011-3>
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), 71–81. <https://doi.org/10.1007/s12369-008-0001-3>
- Chaiken S. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *J. Personal. Soc. Psychol.* 39:752–66
- Chaiken S. 1999. The heuristic-systematic model in its broader context. See Chaiken & Trope 1999, pp. 73–96
- Chen, M. (2024, March 13). Part-1: Pavai.vocei(c3p0) re-imagines C-3PO's Tranlang III Communication Module using AI in 2024. Medium. <https://mychen76.medium.com/part-1-pavai-vocei-c3p0-1021ec966293>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Evans, J. (2007). Hypothetical Thinking: Dual Processes in Reasoning and Judgement. *Hypothetical Thinking: Dual Processes in Reasoning and Judgement*, 1–206. <https://doi.org/10.4324/9780203947487>
- Evans, J. S. B. T. (1990). *Bias in Human Reasoning: Causes and Consequences*. Psychology Press.
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Evans J. S. B. T., & Over D.E. 1996. *Rationality and Reasoning*. Hove, UK: Psychol. Press
- Fodor J. 1983. *The Modularity of Mind*. Scranton, PA: Crowell
- Fodor J. 2001. *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press
- Gong, Li. 2008. "How Social Is Social Responses to Computers? The Function of the Degree of Anthropomorphism in Computer Representations." *Computers in Human Behavior* 24 (4): 1494–1509.
- Goudey, A., & Bonnin, G. (2016). Must smart objects look human? Study of the impact of anthropomorphism on the acceptance of companion robots. *Recherche et Applications En Marketing*, 31(2), 2–20. <https://doi.org/10.1177/2051570716643961>
- Guingrich, R. E., & Graziano, M. S. A. (2024). Ascribing consciousness to artificial intelligence: human-AI interaction and its carry-over effects on human-human interaction. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1322781>
- Hammond KR. 1996. *Human Judgment and Social Policy*. New York: Oxford Univ. Press
- Hecht, E. E., Patterson, R., & Barbey, A. K. (2012). What can other animals tell us about human social cognition? An evolutionary perspective on reflexive and reflexive processing. In *Frontiers in Human Neuroscience* (Issue JULY). Frontiers Media S. A. <https://doi.org/10.3389/fnhum.2012.00224>
- Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008). Understanding social robots: A user study on anthropomorphism. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, 574–579. <https://doi.org/10.1109/ROMAN.2008.4600728>
- Kätsyri, J., Förger, K., Mäkäriäinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, 6(MAR). <https://doi.org/10.3389/fpsyg.2015.00390>
- Kim, J., & Im, I. (2023). Anthropomorphic response: Understanding interactions between humans and artificial intelligence agents. *Computers in Human Behavior*, 139. <https://doi.org/10.1016/j.chb.2022.107512>
- Kim, J., & Im, I. (2024). Do Users Really Want "Human-like" AI? The Effects of Anthropomorphism and Ego-morphism on User's Perceived Anthropocentric Threat. In T. X. Bui (Ed.), *Proceedings of the 57th Annual Hawaii International Conference on System Sciences, HICSS 2024* (pp. 477–487). (Proceedings of the Annual Hawaii International Conference on System Sciences). IEEE Computer Society.
- Kim, S., & McGill, A. L. (2011). Gaming with Mr. Slot or Gaming the Slot Machine? Power, Anthropomorphism, and Risk Perception. *Journal of*

- Consumer Research*, 38(1), 94–107.
<https://doi.org/10.1086/658148>
- Kim, T. W., Tong, Lu, Lee, K., Cheng, Z., Tang, Y., & Hooker, J. (2021). *When is it permissible for artificial intelligence to lie? A trust-based approach*.
<http://arxiv.org/abs/2103.05434>
- Kim, Taejin & Lee, One-Ki & Kang, Juyoung. (2024). Can We Trust an AI Agent? Interaction Effects of Its Machine Learning Performance and Digital Character.
<https://doi.org/10.1111/cogs.13032>
- Lambert, F. (2024b, June 12). Tesla claims it has 2 optimus humanoid robots working autonomously in factory. Electrek. <https://electrek.co/2024/06/12/tesla-2-optimus-humanoid-robots-working-autonomously-factory/>
- Lieberman MD. 2003. Reflective and reflexive judgment processes: a social cognitive neuro-science approach. See Forgas et al. 2003, pp. 44–67
- MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 7(3), 297–337.
- Mackenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct Measurement and Validation Procedures in MIS and Behavioral Research : Integrating New and Existing Techniques1. *MIS Quarterly*, 35(2), 293–334.
- Maleki, N., Padmanabhan, B., & Dutta, K. (2024). AI Hallucinations: A Misnomer Worth Clarifying. arXiv preprint arXiv:2401.06796.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, 19(2), 98–100.
<https://doi.org/10.1109/MRA.2012.2192811>
- Nass, C., & Moon, Y. (2000). Mindfulness Theory and Social Issues - Machines and Mindlessness - Social Responses to Computers. *Journal of Social Issues : A Journal of the Society for the Psychological Studies of Social Issues*, 56(1), 81–103. <http://www.coli.uni-saarland.de/courses/agentinteraction/contents/papers/Nass00.pdf>
- Nisbett R, Peng K, Choi I, Norenzayan A. 2001. Culture and systems of thought: holistic vs.analytic cognition. *Psychol. Rev.* 108:291–310
- Reber AS. 1993. *Implicit Learning and Tacit Knowledge*. Oxford, UK: Oxford Univ. Press
- Schneider, Sascha, Alexandra Häßler, Tanja Habermeyer, Maik Beege, and Günter Daniel Rey. 2019. “The More Human, the Higher the Performance? Examining the Effects of Anthropomorphism on Learning with Media.” *Journal of Educational Psychology* 111 (1): 57.
- Schneider W, Shiffrin RM. 1977. Controlled and automatic human information processing I: detection, search and attention. *Psychol. Rev.* 84:1–66
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1), 3–22.
<https://doi.org/10.1037/0033-2909.119.1.3>
- Stanovich KE. 1999. *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Erlbaum
- Stanovich, K. E., Sá, W. C., & West, R. F. (2004). Individual Differences in Thinking, Reasoning, and Decision Making. In *The nature of reasoning*. (pp. 375–409). Cambridge University Press.
<https://doi.org/10.1037/12822-019>
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). Intelligence and rationality. In *The Cambridge handbook of intelligence*. (pp. 784–826). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511977244.040>
- Stein, J. P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160, 43–50.
<https://doi.org/10.1016/j.cognition.2016.12.010>
- Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly: Management Information Systems*, 13(2), 147–165.
<https://doi.org/10.2307/248922>
- Su, Y. (2023). Artificial Intelligence: The Significance of Tesla Bot. Highlights in Science, Engineering and Technology, 39, 1351–1355.
<https://doi.org/10.54097/hset.v39i.6767>
- Takayama, L. (2009). Making sense of agentic objects and teleoperation: In-the-moment and reflective perspectives. 2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 239–240. <https://doi.org/10.1145/1514095.1514155>
- Tarsney, C. (2024). Deception and Manipulation in Generative AI. arXiv preprint arXiv:2401.11335.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. In *New Series* (Vol. 185, Issue 4157).
- Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219–232.
<https://doi.org/10.1177/1745691610369336>
- Willis, J., & Todorov, A. (2006). *First Impressions Making Up Your Mind After a 100-Ms Exposure to a Face*.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A Model of Dual Attitudes. *Psychological Review*, 107(1), 101–126. <https://doi.org/10.1037/0033-295X.107.1.101>
- Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The Interactive Effects of Robot Anthropomorphism and Robot Ability on Perceived Threat and Support for Robotics Research. *Journal of Human-Robot Interaction*, 5(2), 29.
<https://doi.org/10.5898/jhri.5.2.yogeeswaran>
- Złotowski, J., Sumioka, H., Eyssel, F., Nishio, S., Bartneck, C., & Ishiguro, H. (2018). Model of Dual Anthropomorphism: The Relationship Between the Media Equation Effect and Implicit Anthropomorphism. *International Journal of Social Robotics*, 10(5), 701–714.
<https://doi.org/10.1007/s12369-018-0476-5>