The Effects of Anthropomorphism and Explanation Types on User Perception and Acceptance: Implications for Explainable AI

Jiawei Tong, Tingru Cui, Ofir Turel and Bo Du

November 3, 2023

The Effects of Anthropomorphism and Explanation Types on User Perception and Acceptance: Implications for Explainable AI

Research-in-progress

Jiawei Tong
School of Computing and Information Systems
University of Melbourne
Melbourne, Australia
Email: totj1@student.unimelb.edu.au

Tingru Cui
School of Computing and Information Systems
University of Melbourne
Melbourne, Australia
Email: tingru.cui@unimelb.edu.au

Ofir Turel
School of Computing and Information Systems
University of Melbourne
Melbourne, Australia
Email: oturel@unimelb.edu.au

Bo Du
Faculty of Engineering and Information Sciences
University of Wollongong
Wollongong, Australia
Email: bdu@uow.edu.au

Abstract

Explainable Artificial Intelligence (XAI) applications are widely used in interactions with end users. However, there remains a lack of understanding of how the different characteristics of these systems, particularly the anthropomorphic design and the type of explanations provided interact to affect user reactions to AI. We address this research gap by building on social response theory (SRT), prior XAI and anthropomorphic design literature, to investigate how anthropomorphic design (human-like vs. machine-like) and types of explanations (consensual, expert, internal, empirical validation-based explanations) affect user reactions to AI (perceived trust and persuasiveness) and acceptance of AI systems. We will evaluate the proposed research model by conducting a 2 × 4 between-subjects experiment. This study will enrich the theoretical landscape of anthropomorphic design and human-AI interaction (HAII), offering actionable insights into user perception and acceptance for XAI practitioners.

Keywords: XAI, Human-AI Interaction, Conceptualisation Validation Types, Anthropomorphic Design

1 Introduction

With the rapid development of Artificial Intelligence (AI), AI-based applications are widely used in interactions with end users, from online education transformations (Seo et al., 2020; VanLehn, 2011) to e-commerce revolutions (Bawack et al., 2022; Enholm et al., 2021). This pervasive impact emphasises the importance of understanding how users perceive and interact with these AI systems. A significant area of attention is the explainability of AI (Li and Suh, 2022; Xu et al., 2019). XAI refers to AI systems that elaborate their decision-making processes with details or rationale to their users, making their functions transparent and understandable (Barredo Arrieta et al., 2020). XAI has gained momentum given the need of humans to have explanations that resonate with their logic for trusting information sources. As such, humans will be more willing to trust and accept recommendations from AI that explain the logic behind its recommendations compared to one that does not (He et al., 2022).

Despite the growth of XAI, there are still several unexplored areas (Miller, 2019). One such area is how the interaction of different XAI characteristics, especially anthropomorphic design and the explanation type AI systems provided, shapes user reactions to AI decisions (Erlei et al., 2020; Troshani et al., 2020). These characteristics are important because first, people respond differently to computer entities with human characteristics (i.e., anthropomorphised) than others (Glikson and Woolley, 2020). Anthropomorphic design refers to the design characteristics that make a system or interface appear or behave in human-like ways (Yang et al., 2021). Prior literature advocates that anthropomorphic design should be used in AI-enabled applications that interact with end-users (Adam et al., 2020; Zhang et al., 2012), such as AI-enabled chatbots, and focuses on its impact on user perceptions and acceptance (Elkins et al., 2012; Qiu and Benbasat, 2009; Zhang et al., 2021). This not only makes interactions more intuitive but also bridges the gap between complex technical details and human understanding, which can evoke a more relatable and trustable perception from users (Benlian et al., 2019; Hong et al., 2017; Li and Suh, 2022; Zhang et al., 2021). On the other hand, Yang et al. (2021) point out that previous studies have provided inconsistent evidence regarding the impact of anthropomorphism on user perception and acceptance. Their research further emphasises that the effects of anthropomorphic design often depend on the specific application context. Therefore, as AI adoption continues to integrate into more sectors, it becomes essential to further explore the impact of anthropomorphism, particularly its interaction impact with other XAI design features. Second, the nature of the explanation matters; for example, the basis for the claim or the source of the data can affect the persuasiveness of the message (Eagly and Chaiken, 1984). Nevertheless, how these aspects work and interact in the context of AI is yet to be examined. Such interactions are particularly important because the way we depict the source of the message can have different effects on message persuasiveness under different message content and argumentation conditions (Gunaratne et al., 2018).

This paper aims to fill this research gap by investigating how anthropomorphic design and explanation types interact to impact user trust, message persuasiveness, and acceptance of AI systems. Based on prior literature, we designed two distinct AI agent interfaces to compare the effects of anthropomorphic design. Simultaneously, our study adapts the four conceptualisation validation types proposed by Jaccard and Jacoby (2009) as the main explanation types. Drawing from social response theory (SRT), prior XAI and anthropomorphic design literature, we propose an experimental design to compare the differences in user perception caused by different explanation features. Therefore, our study aims to answer the following two questions:

RQ1: *Which explanation types in AI systems most positively influence user trust and persuasiveness?*
RQ2: *How does the interplay between anthropomorphic design and different explanation types impact user trust, message persuasiveness, and subsequent acceptance of AI systems?*

In this research-in-progress paper, we use a 2x4 between-subjects experimental design to test the proposed hypotheses, aiming to examine the impact of anthropomorphic design and explanation types on user perception of AI systems. By making AI systems more understandable and trustworthy, we can accelerate their adoption in critical domains beyond healthcare, such as transportation, environmental protection, and public policy, which have wide-ranging impacts on the common good of society.

2 Theoretical Foundation

2.1 Explainable AI (XAI) and Explanation Types

Explainable Artificial Intelligence (XAI) has emerged as a crucial research domain within AI, stimulating numerous advancements as both academia and industry strive for algorithmic transparency (Gohel et al., 2021; Miller, 2019). Broadly, XAI refers to AI systems that demystify their decision-making processes with details or rationale to their users, making their decisions clear or easy to understand (Miller, 2019; Minh et al., 2021). Existing literature on XAI highlights the necessity of well-designed explanatory mechanisms in AI systems and the principles governing them (Chromik and Butz, 2021). A crucial aspect of XAI's comprehensibility is the type and pattern of system explanations, which aligns well with the research aim of this paper (Brdnik, 2023; Vilone and Longo, 2020), and several studies have elucidated the effectiveness of different types of explanations provided by AI on user perceptions, including fairness perceptions and trust intentions (Dodge et al., 2019; Song et al., 2023). This study adapts the four conceptualisation validation types proposed by Jaccard and Jacoby (2009) as the primary explanation types for AI systems. These four types, namely consensual, expert, internal, and empirical validation, are considered to govern human understanding and guide the formation of beliefs. These conceptualisation validation types provide a structured methodology for effectively evaluating the reliability and credibility of theories, concepts, or information. This makes them highly relevant in the context of XAI applications and aligns closely with the perceptual metrics we have set (perceived trust and persuasiveness). This consistency leads us to adapt these validation types as AI explanation types in our experiment, aiming to explore the actual impact of explanations rooted in different validation types on users' perceived trust and understanding. *Consensual validation* operates on the principle of collective agreement or shared perceptions of facts (Scarr, 1985). This validation type assumes if most users widely accept an explanation or if it aligns with their intuitive reasoning, then that explanation is more likely to be considered correct and trustworthy. *Expert validation* relies on endorsements from individuals with extensive knowledge and experience in the relevant domain; when system explanations align with expert knowledge or are validated by experts, it increases the system's perceived truth, potentially increasing user reliance on such systems. *Internal validation* sets rigorous logic principles to ensure the coherence and integrity of concepts. Logically consistent and contradiction-free AI explanations are likely to enhance user trust in the system's reasoning. *Empirical validation* involves meticulous empirical testing to determine if a concept accurately represents the phenomenon under study. For XAI, this implies that if an explanation aligns well with empirical data or mirrors real-world outcomes, users are more likely to perceive it as reliable and accurate, consequently facilitating users' AI acceptance.

2.2 Social Response Theory and Anthropomorphic Design

Anthropomorphic design, as defined by Yang et al. (2021), refers to the extent to which an AI service agent emulates human characteristics such as appearance, voice, and behaviour. This concept is deeply intertwined with SRT's foundational premise that humans instinctively perceive computers and digital agents as social entities, treating them as social actors (Nass et al., 1994). When a system displays specific social cues, such as a female avatar's voice or visual representation, users mindlessly categorise the system within the respective social framework (e.g., viewing the system as female) and consequently apply related social norms to it (Nass and Moon, 2000). Such interpretations and reactions are not limited solely to gender attributions but encompass a broad range of anthropomorphic attributes. Prior studies have shown that AI systems can increase user trust and effectively ameliorate HAII barriers through their anthropomorphic features (Li and Suh, 2022; Zhang et al., 2021). AI systems with human-like features are perceived as more relatable, with users often treating these systems as social actors, as supported by SRT. Some studies have already validated the relevance of SRT in the domains of AI and anthropomorphic design. For example, Qiu and Benbasat (2009) used SRT to explain the effects of different social cues in human-like technologies. Furthermore, advances in AI systems with enhanced perceptual and cognitive capabilities have further emphasised the profound impact of anthropomorphic features designed based on SRT on user perception and intent, including improving the overall acceptance of the system by users (Benlian et al., 2019; Schuetzler et al., 2018; Yang et al., 2021; Ying et al., 2013). Recognising the key role of SRT in navigating these perceptions can provide deeper insights to inform the design and development of AI systems that are not only technically proficient but also socially consistent and user-friendly.

2.3 User Perception of AI Systems

The way users perceive and interact with digital technologies, including AI systems, is a significant focus in contemporary literature (Araujo et al., 2020; Daneji et al., 2019; Denaputri and Usman, 2019; Wilson et al., 2021). Key metrics such as perceived trust and perceived persuasiveness have been identified as determinant factors influencing the adoption, and users' overall AI acceptance. The perceived trust of users in AI systems is critical, especially in scenarios where AI is assisting in decision-making, and people will avoid using systems they do not trust (Bartneck et al., 2020). Users are more likely to rely on systems they perceive as trustworthy, which in turn increases their acceptance and continued use of such systems (Lee and See, 2004). Perceived persuasiveness, on the other hand, refers to the ability of a Human-computer Interaction (HCI) system to persuade people to change their behaviour through its

UX design elements and features, which can also be regarded as a key determinant in user acceptance and evaluation (Lehto et al., 2012; O'Keefe, 2018). In our study, AI acceptance is defined as users' willingness to use, rely on, and integrate an AI system into their daily activities, which stems from their evaluation of its utility. Our investigation seeks to explore how explanation types combined with anthropomorphic design features influence these user perception metrics. By understanding these dynamics, we aim to provide insights beneficial for designing AI systems that meet user expectations, improving their perception and acceptance.

3 Research Model and Hypotheses Development

This study posits that user perception and overall acceptance of the AI system are influenced by its anthropomorphic design and explanation pattern, which consistently aligns with the theoretical foundations. Thus, our study aims to reveal the interplay between these XAI designs and further explore how these characteristics can affect the user's perceived trust, perceived persuasiveness, and overall acceptance. The proposed research model for this study is shown in Figure 1.

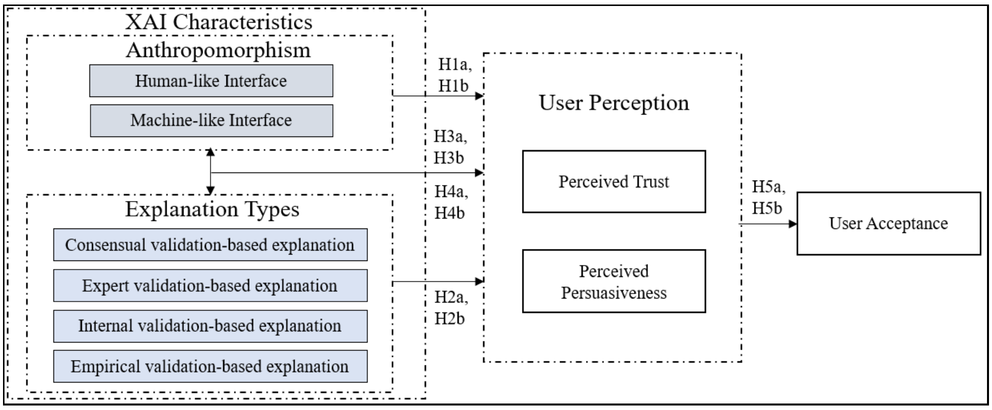

Figure 1. Research Model

3.1 Effects of Anthropomorphic Design

The presence of anthropomorphic design in computer-mediated environments creates a sense of 'being with others', which reduces the psychological distance between the user and the AI system (Lim et al., 2021). This is consistent with the SRT that posits humans inherently invoke emotional and social responses to non-human entities. Simultaneously, extensive research has shown the fact that users tend to interact more effectively with AI systems that embody human-like characteristics, leading to a positive impact on user perceptions (Ying et al., 2013; Benlian et al., 2019; Zhang et al., 2021; Li and Suh, 2022). Hence, we propose the following hypothesis:

H1a-b: *The presence of anthropomorphic design in AI systems positively influences user perceptions, i.e., (a) perceived trust and (b) perceived persuasiveness.*

3.2 Effects of Explanation Types on User Perceptions

Previous studies have already pointed out the strong connection between the explainability of AI systems, the types of explanations they provide and subsequent user perception and understanding (Brdnik, 2023; Chromik and Butz, 2021; Kulesza et al. 2013; Miller, 2019). For example, Miller (2019) emphasises the importance of choosing the right type of explanation for different users and environments, in particular the apparent impact of contrasting explanations and selective explanations on user perception. Kulesza et al. (2013) investigated the impact of the explanation types on user perception, concluding that low-fidelity explanations make it difficult for users to understand the information, thus undermining their trust and persuasion in the intelligent agent providing such explanations. Building on previous literature, our study focuses on the role of explanation types in AI systems and their impact on user perception. We decided to adapt the four types of validation proposed by Jaccard and Jacoby (2009): consensual validation, expert validation, internal validation, and empirical validation as the cornerstones for investigating the types of explanations in AI systems and examining their impact on users' perceived trust, perceived persuasiveness, and acceptance. We propose the second hypothesis as follows:

H2a-b: *Explanations provided by AI systems positively influence user perceptions, i.e., (a) perceived trust and (b) perceived persuasiveness.*

3.3 The Combined Effect of Anthropomorphic Design and Explanation Types

Currently, augmenting anthropomorphic designs with additional features to achieve a more optimal combined effect has been discussed in the HCI design literature. Xiao and He (2020) suggest that HCI products, such as AI agents based on anthropomorphic design, can be enhanced by diversifying interaction forms and expression styles to improve user experience. These enhancements also include the types of explanations output by AI agents. Drawing insights from the SRT, human tends to anthropomorphise technology, especially when it presents human-like characteristics. According to Graaf and Malle (2017), AI systems employ anthropomorphic designs and provide coherent types of explanations to optimise user interaction. When AI systems exhibit anthropomorphic characteristics and provide explanations, it is not just the design or explanation alone that impacts user perception. Instead, it is the social and emotional responses invoked in the user by the combined effect of the design and explanation. The efficacy of different explanation types may vary depending on whether the XAI interface is human-like or machine-like. In our upcoming experiment, inspired by the 2 × 2 between-subjects design used by Kim and Song (2021), we aim to identify optimal combinations of anthropomorphic design and explanation types that maximise users' perceived trust, persuasiveness, and acceptance. We hypothesise that consensual and expert validations fit well with anthropomorphic interfaces because they exploit the social and emotional aspects of human interaction. Consensual validation is based on social agreements that complement the social nature of anthropomorphic design posited by SRT. Expert validation, on the other hand, exploits the emotional trust that users tend to place in 'authoritative' or 'knowledgeable' entities (Bem, 1970), which are features commonly attributed to anthropomorphic design. In contrast, our study posits that machine-like interfaces may align better with internal and empirical validations. It is a logical extension given that users often expect machine-like interfaces to operate based on data and internal logic. Therefore, explanations rooted in empirical evidence or the system's internal logic are more likely to resonate with users, enhancing their trust and persuasiveness. Hence, we propose the following hypotheses:

H3a-d: *AI systems that use <u>human-like interfaces with a) consensual</u> and <u>b) expert-based explanations</u> will be more effective than other interface-explanation combinations in positively influencing the <u>users' c) perceived trust</u> and <u>d) perceived persuasiveness.</u>*

H4a-d: *AI systems that use <u>machine-like interfaces with a) internal</u> and <u>b) empirical-based explanations</u> will be more effective than other interface-explanation combinations in positively influencing the <u>users' c) perceived trust</u> and <u>d) perceived persuasiveness.</u>*

3.4 Perceived Trust, Perceived Persuasiveness, and Overall Acceptance

In the context of examining the users' perception, we believe that perceived trust and persuasiveness become crucial, affecting the user's relationship and acceptance of the technology. This opinion can be supported by Shin (2021), who states that AI acceptance increases when trust is established between AI and humans. Drawing from the theory and literature, we argue that perceived trust and persuasion are indicators of overall user acceptance. Therefore, we propose the following hypothesis:

H5a-b: *a) <u>perceived trust,</u> and b) <u>perceived persuasiveness</u>, will positively influence <u>overall user acceptance</u> of the AI system.*

4 Research Method

4.1 Experimental Design

This study aims to investigate the impact of anthropomorphic design (including two levels: human-like interface design and machine-like interface design) and type of explanation (including four variations: consensual, expert, internal, and empirical validation-based explanations) on user perception, measured primarily in terms of perceived trust and perceived persuasiveness. We will conduct a 2 (anthropomorphic design) × 4 (types of explanation) between-subjects design, using these XAI system characteristics as independent variables, and user perceptual metrics as dependent variables. Data collection will be conducted through an online experiment. To ensure the robustness and reliability of our findings, we plan to recruit at least 50 participants per condition, summing to a total of 400 participants. We plan to recruit these participants from Amazon Mechanical Turk (MTurk), an online crowdsourcing platform established by Amazon. This sample size was determined based on a power analysis that suggests a minimum total sample size of 179 for an effect size of $f = 0.25$, an alpha error of $\alpha = 0.05$ and a power of $(1 - \beta) = 0.8$ (Faul et al., 2007). Each participant will be randomly assigned to one of eight conditions and interact with a simulated AI system, and then take a questionnaire to assess their perception of the system. Once data collection is complete, we will test the hypotheses using a two-way analysis of variance (ANOVA), which compares significant mean differences between two factors and assesses whether there is an interaction between the two independent variables and the dependent variable (Laerd Statistics, 2018). Therefore, we can test hypotheses H1 to H4 by respectively testing the main and interaction effect of anthropomorphic design and type of explanation on the user perception. Subsequently, we will test H5 using multiple linear regression analyses, a statistical method that can effectively help predict and articulate correlations between independent and dependent variables (Ali and Younas, 2021). This approach allows us to understand how perceived trust and perceived persuasiveness affect overall acceptance. Through our experimental design, this study will comprehensively examine how different XAI characteristics affect user perception and overall acceptance, thus providing references and new insights into the further interaction design of the AI system from a human-centred perspective.

4.2 Manipulation of Anthropomorphic Design

This study manipulates multiple factors of the AI system, particularly its anthropomorphic design, which includes elements such as name, instructions, visual cues, and conversational styles. These manipulations serve to establish two experimental conditions: human-like versus machine-like AI agents. In human-like conditions, AI systems use friendly human avatars as visual cues combined with advanced artificial empathy algorithms (Krakovsky, 2018; McDuff and Czerwinski, 2018), which enable AI to discern the users' emotional state and generate emotionally intelligent responses. For example, AI may respond with "I'm sorry to hear that you've been feeling uncomfortable recently." when informed of users' health concerns.

Figure 2. Experimental Manipulation of Anthropomorphic Design

At the same time, some human-like expressions are also reflected in the conversational style, including some oral expressions like "Don't worry" and "You're not alone" to calm users' potential concerns. In addition, we added emojis to the dialogue in the human-like condition, which enables it to express emotions more like a human. In contrast, the machine-like interface has been represented by using a robot icon. It communicates more matter-of-factly, tending to convey valid information with a mechanical, objective style. There are no first-person pronouns or emotional interactions but directly addresses the user's symptoms by performing an algorithmic analysis and proposing an objective solution. This distinction aims to showcase how anthropomorphic design affects users' perceptions.

4.3 Manipulation of Explanation Types

In this study, we utilise four distinct explanation types based on the four conceptualisation validation types proposed by Jaccard and Jacoby (2009) to manipulate the experiment, each explanation type has been based on a different form of validation: consensual validation, expert validation, internal validation, and empirical validation. Table 1 showcases the four types of explanations in the context of a diabetes diagnosis. *Consensual* validation leverages the collective wisdom of healthcare practitioners and people with diabetes to interpret AI system diagnoses. An example would be, "95% agreed that diabetes is typically associated with symptoms such as excessive thirst, increased diet, frequent urination and unexplained weight loss." *Expert* Validation uses authoritative sources such as medical books or the views of well-known medical experts to support AI conclusions. It might state, "The symptoms you describe align with the medical community's understanding of Type 2 Diabetes, as detailed in the latest clinical studies." *Internal* validation employs logical reasoning, such as, "Based on vast medical record data, symptoms like frequent thirst, increased eating, frequent urination, and unexplained weight loss are strongly associated with having type 2 diabetes." *Empirical* validation verifies AI diagnoses by referencing empirical studies or previous cases with similar symptoms, like "Based on data from other cases with similar symptoms, there is a significant likelihood that you have

Type 2 Diabetes." These four types of explanations will be applied separately to both human-like and machine-like AI designs. This approach enables us to investigate how interpretive styles interact with anthropomorphic design and subsequently influence user perceptions and acceptance.

Explanation Types	
Consensual Validation	Based on the surveys conducted among medical professionals and people with similar symptoms, 95% agreed that the symptoms described fit the profile of having diabetes.
Expert Validation	Medical textbooks and expert physicians generally agree that these symptoms are signs of diabetes. Dr Vincent Tong, a leading endocrinologist, has identified these symptoms as major indicators of the diagnosis of diabetes in his latest publication.
Internal Validation	The algorithm uses a logical rule-based system that links symptoms such as increased thirst, increased eating, frequent urination, and weight loss to diabetes. This logical relationship is derived from a large dataset of medical records.
Empirical Validation	In a dataset of 10,000 patients with similar symptoms, 97.2% were diagnosed with diabetes. This empirical evidence sophisticatedly supports the algorithm's suggestion that the symptoms described could likely be due to diabetes.

Table 1. Experimental Manipulation of Explanation Types

4.4 Measurements

Constructs	**Item Description** (1-7 Likert scale, 1=Strongly disagree, 7=Strongly agree)	**Source**
Perceived Trust	TRU1: The AI system is dependable in making medical diagnoses. TRU2: The AI system is reliable for making medical diagnoses. TRU3: Overall, I can trust this AI system in medical diagnosis.	Choi & Ji (2015)
Perceived Persuasiveness	PER1: The AI system has influenced my thinking on medical diagnoses. PER2: Messages from the AI system regarding medical diagnoses are convincing. PER3: The AI system guides my judgment in medical diagnoses reasonably.	O'Keefe (2018); Cui et al. (2020)
Overall User Acceptance	UAC1: I intend to use this AI system for diagnosis when facing a similar health issue. UAC2: I expect to use this AI diagnostic tool over seeking advice from a colleague. UAC3: I plan to rely on the message from this AI system for more of my responsibilities.	Choi & Ji (2015); Cui et al. (2020)

Table 2. Measurement Instruments

In our study, participants will complete a questionnaire designed to measure their perceptions after interacting with a simulated AI system. As shown in Table 2, our questionnaire builds on previous studies where measures of perceived trust, perceived persuasiveness, and overall user acceptance were referenced or adapted from Choi and Ji (2015), O'Keefe (2018), and Cui et al. (2020), respectively. Each construct will be assessed using a 7-point Likert scale, where 1 means strongly disagree and 7 represents strongly agree. This measure is consistent with practices established in previous research in the field and allows for a thorough examination of our research questions (Denaputri and Usman, 2019; Lopes et al., 2022; Weitz et al., 2020). Furthermore, to enhance the robustness of the model and control for potential group characterisation effects, we will also collect demographic information such as age and gender, as well as participants' prior understanding of the AI system.

5 Discussion and Conclusion

Exploring AI system characteristics is crucial for developing trustworthy and persuasive XAI. This research-in-progress addresses this gap through a 2 × 4 between-subjects experimental design, focusing on anthropomorphic design and explanation types. This study makes three contributions. First, we deepen the understanding of anthropomorphic design's impact on user perception by employing two distinct AI agent conditions within a healthcare context. Second, we adapt Jaccard and Jacoby's (2009) four conceptualization validation types as four explanation types by linking mankind's validation of information truth and the explanation's reliability and persuasiveness. Extending their framework to XAI allows us to systematically evaluate the influence of different types of explanations on user perceptions of AI systems. Lastly and most importantly, our research is one of the first studies exploring the combined effect of anthropomorphic design and explanation types upon user perception in AI systems, thus offering both theoretical insights and practical guidance for XAI design. Our research is limited by its focus on a single healthcare context and sample group. Future research should extend this study by exploring alternative perceptual metrics and contexts for broader generalisability, investigating ethical dimensions of XAI such as the potential "double-edged sword" effect where increased explainability and trustworthiness could be exploited for deceptive purposes, and incorporating qualitative methods like interviews for deeper insights into user perception.

6 References

Adam, M., Wessel, M., & Benlian, A. (2020). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *31*(2). Springer. https://doi.org/10.1007/s12525-020-00414-7

Ali, P., & Younas, A. (2021). Understanding and interpreting regression analysis. *Evidence Based Nursing*, *24*(4), 116–118. https://doi.org/10.1136/ebnurs-2021-103425

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3). https://doi.org/10.1007/s00146-019-00931-w

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58(1), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2020). Trust and Fairness in AI Systems. *An Introduction to Ethics in Robotics and AI*, 27–38. https://doi.org/10.1007/978-3-030-51110-4_4

Bawack, R. E., Wamba, S. F., Carillo, K. D. A., & Akter, S. (2022). Artificial intelligence in E-Commerce: a bibliometric study and literature review. *Electronic Markets, 32*, 297–338. Springer. https://doi.org/10.1007/s12525-022-00537-z

Bem, D. J. (1970). *Beliefs, Attitudes, and Human Affairs*. Brooks/Cole Pub.

Benlian, A., Klumpe, J., & Hinz, O. (2019). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*. https://doi.org/10.1111/isj.12243

Brdnik, S. (2023). GUI Design Patterns for Improving the HCI in Explainable Artificial Intelligence. *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, 240–242. Association for Computing Machinery. https://doi.org/10.1145/3581754.3584114

Choi, J. K., & Ji, Y. G. (2015). Investigating the Importance of Trust on Adopting an Autonomous Vehicle. *International Journal of Human-Computer Interaction*, *31*(10), 692–702. https://doi.org/10.1080/10447318.2015.1070549

Chromik, M., & Butz, A. (2021). Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces. *Human-Computer Interaction – INTERACT 2021*, 619–640. https://doi.org/10.1007/978-3-030-85616-8_36

Cui, T., Peng, X., & Wang, X. (2020). Understanding the Effect of Anthropomorphic Design: Towards More Persuasive Conversational Agents. *ICIS 2020 Proceedings*. https://aisel.aisnet.org/icis2020/user_behaviors/user_behaviors/9/

Daneji, A. A., Ayub, A. F. M., & Khambari, M. N. Md. (2019). The effects of perceived usefulness, confirmation and satisfaction on continuance intention in using massive open online course (MOOC). *Knowledge Management & E-Learning: An International Journal*, 201–214. https://doi.org/10.34105/j.kmel.2019.11.010

Denaputri, A., & Usman, O. (2019). Effect of Perceived Trust, Perceived Security, Perceived Usefulness and Perceived Ease of use on Customers' Intention to Use Mobile Payment. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3511632

Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K. E., & Dugan, C. (2019). Explaining models. *Proceedings of the 24th International Conference on Intelligent User Interfaces*. https://doi.org/10.1145/3301275.3302310

Eagly, A. H., & Chaiken, S. (1984). Cognitive Theories of Persuasion. *Advances in Experimental Social Psychology, 17*(C), 267–359. Academic Press. https://doi.org/10.1016/S0065-2601(08)60122-7

Elkins, A. C., Derrick, D. C., Burgoon, J. K., & Nunamaker Jr., J. F. (2012). Predicting Users' Perceived Trust in Embodied Conversational Agents Using Vocal Dynamics. *2012 45th Hawaii International Conference on System Sciences*. https://doi.org/10.1109/hicss.2012.483

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial Intelligence and Business Value: a Literature Review. *Information Systems Frontiers*, *24*(5), 1709–1734. springer. https://link.springer.com/article/10.1007/s10796-021-10186-w

Erlei, A., Nekdem, F., Meub, L., Anand, A., & Gadiraju, U. (2020). Impact of Algorithmic Decision Making on Human Behavior: Evidence from Ultimatum Bargaining. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *8*, 43–52. https://doi.org/10.1609/hcomp.v8i1.7462

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals, 14*(2)*,* 627–660. https://doi.org/10.5465/annals.2018.0057

Gohel, P., Singh, P., & Mohanty, M. (2021). Explainable AI: current status and future directions. *ArXiv:2107.07045 [Cs]*. https://arxiv.org/abs/2107.07045

Graaf, M. D., & Malle, B. F. (2017). How People Explain Action (and Autonomous Intelligent Systems Should Too). *AAAI Fall Symposia*., 19–26.

Gunaratne, J., Zalmanson, L., & Nov, O. (2018). The Persuasive Power of Algorithmic and Crowdsourced Advice. *Journal of Management Information Systems, 35*(4), 1092–1120. https://doi.org/10.1080/07421222.2018.1523534

He, X., Hong, Y., Zheng, X., & Zhang, Y. (2022). What Are the Users' Needs? Design of a User-Centered Explainable Artificial Intelligence Diagnostic System. *International Journal of Human–Computer Interaction*, *39*(7), 1519–1542. https://doi.org/10.1080/10447318.2022.2095093

Hong, E., Cho, K., & Choi, J. (2017). Effects of Anthropomorphic Conversational Interface for Smart Home : An Experimental Study on the Voice and Chatting Interactions. *Journal of the HCI Society of Korea*, *12*(1), 15. https://doi.org/10.17210/jhsk.2017.02.12.1.15

Jaccard, J., & Jacoby, J. (2009). *THEORY CONSTRUCTION AND MODEL-BUILDING SKILLS : a practical guide for social scientists.* Guilford.

Kim, T., & Song, H. (2021). How should intelligent agents apologize to restore trust? Interaction effects between anthropomorphism and apology attribution on trust repair. *Telematics and Informatics*, *61*, 101595. https://doi.org/10.1016/j.tele.2021.101595

Krakovsky, M. (2018). Artificial (emotional) intelligence. *Communications of the ACM*, *61*(4), 18–19. https://doi.org/10.1145/3185521

Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I., & Wong, W.-K. (2013, September 1). *Too much, too little, or just right? Ways explanations impact end users' mental models*. IEEE Xplore. https://doi.org/10.1109/VLHCC.2013.6645235

Laerd Statistics. (2018). *Two-way ANOVA in SPSS statistics - step-by-step procedure including testing of assumptions | laerd statistics*. Laerd.com. https://statistics.laerd.com/spss-tutorials/two-way-anova-using-spss-statistics.php

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lehto, T., Harri Oinas-Kukkonen, & Drozd, F. (2012). Factors Affecting Perceived Persuasiveness of a Behavior Change Support System. *International Conference on Information Systems, ICIS 2012, 3*.

Li, M., & Suh, A. (2022). Anthropomorphism in AI-enabled technology: A literature review. *Electronic Markets*. https://doi.org/10.1007/s12525-022-00591-7

Lim, X.-J., Cheah, J.-H., Ng, S. I., Basha, N. K., & Soutar, G. (2021). The effects of anthropomorphism presence and the marketing mix have on retail app continuance use intention. *Technological Forecasting and Social Change*, *168*, 120763. https://doi.org/10.1016/j.techfore.2021.120763

Lopes, P., Silva, E., Braga, C., Oliveira, T., & Rosado, L. (2022). XAI Systems Evaluation: A Review of Human and Computer-Centred Methods. *Applied Sciences*, *12*(19), 9423. https://doi.org/10.3390/app12199423

McDuff, D., & Czerwinski, M. (2018). Designing emotionally sentient agents. *Communications of the ACM*, *61*(12), 74–83. https://doi.org/10.1145/3186591

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2021). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-021-10088-y

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*. https://doi.org/10.1145/191666.191703

O'Keefe, D. J. (2018). Message Pretesting Using Assessments of Expected or Perceived Persuasiveness: Evidence About Diagnosticity of Relative Actual Persuasiveness. *Journal of Communication*, *68*(1), 120–142. https://doi.org/10.1093/joc/jqx009

Qiu, L., & Benbasat, I. (2009). Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *Journal of Management Information Systems*, *25*(4), 145–181. https://www.jstor.org/stable/40398956

Scarr, S. (1985). Constructing psychology: Making facts and fables for our times. *American Psychologist*, *40*(5), 499–512. https://doi.org/10.1037/0003-066x.40.5.499

Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems,* 114, 94–102. https://doi.org/10.1016/j.dss.2018.08.011

Seo, K., Fels, S., Yoon, D., Roll, I., Dodson, S., & Fong, M. (2020). Artificial Intelligence for Video-based Learning at Scale. *Proceedings of the Seventh ACM Conference on Learning @ Scale*, 215–217. https://doi.org/10.1145/3386527.3405937

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Song, Y., Cui, T., & Liu, F. (2023). DESIGNING FAIR AI SYSTEMS: HOW EXPLANATION SPECIFICITY INFLUENCES USERS' PERCEIVED FAIRNESS AND TRUSTING INTENTIONS. *ECIS 2023 Research-In-Progress Papers*. https://aisel.aisnet.org/ecis2023_rip/7

Troshani, I., Rao Hill, S., Sherman, C., & Arthur, D. (2020). Do We Trust in AI? Role of Anthropomorphism and Intelligence. *Journal of Computer Information Systems*, 1–11. https://doi.org/10.1080/08874417.2020.1788473

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vilone, G., & Longo, L. (2020). *Explainable Artificial Intelligence: a Systematic Review*. https://doi.org/10.48550/arxiv.2006.00093

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & André, E. (2020). "Let me explain!": exploring the potential of virtual agents in explainable AI interaction design. *Journal on Multimodal User Interfaces*. https://doi.org/10.1007/s12193-020-00332-0

Wilson, N., Keni, K., & Tan, P. H. P. (2021). The Role of Perceived Usefulness and Perceived Ease-of-Use toward Satisfaction and Trust which Influence Computer Consumers' Loyalty in China. *Gadjah Mada International Journal of Business*, *23*(3), 262. https://doi.org/10.22146/gamaijb.32106

Xiao, W., & He, C. (2020). Study on Anthropomorphism in Human–Computer Interaction Design. *Man-Machine-Environment System Engineering*, 629–635. https://doi.org/10.1007/978-981-15-6978-4_72

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. *Natural Language Processing and Chinese Computing*, 563–574. https://doi.org/10.1007/978-3-030-32236-6_51

Yang, Y., Liu, Y., Lv, X., Ai, J., & Li, Y. (2021). Anthropomorphism and customers' willingness to use artificial intelligence service agents. *Journal of Hospitality Marketing & Management*, 1–23. https://doi.org/10.1080/19368623.2021.1926037

Ying, F. T., Wang, G. Y., Tao, Y., & Cai, J. X. (2013). Integrating Controlling and Anthropomorphic Affordances on Smart Product Design. *Advanced Materials Research*, *718-720*, 1352–1355. https://doi.org/10.4028/www.scientific.net/amr.718-720.1352

Zhang, H., Lu, Y., Shi, X., Tang, Z., & Zhao, Z. (2012). Mood and social presence on consumer purchase behaviour in C2C E-commerce in Chinese culture. *Electronic Markets*, *22*(3), 143–154. https://doi.org/10.1007/s12525-012-0097-z

Zhang, M., Gursoy, D., Zhu, Z., & Shi, S. (2021). Impact of anthropomorphic features of artificially intelligent service robots on consumer acceptance: moderating role of sense of humor. *International Journal of Contemporary Hospitality Management*, *33*(11), 3883–3905. https://doi.org/10.1108/ijchm-11-2020-1256

Acknowledgements

This research was supported by the National Natural Science Foundation of China under Grant Number 72172011.