Externalizing and Verbalizing to More Adequately Measure Trust in AI

ANONYMOUS AUTHOR(S)

Artificial intelligence (AI) is increasingly used in digital products for a wide variety of purposes like content creation, recommendation, surveillance, and judgment. With AI technologies designed with the intent of assisting humans becoming more common, understanding how users perceive and interact with AI becomes an important line of investigation. A key aspect is trust. Trust is concerned with the confidence users have in AI and their propensity to use it. While it may not necessarily translate into behavior, trust is integral in determining whether users will even consider using AI. While there are many experimental studies evaluating trust in novel AI systems, a review of such studies has called out the inadequacy of the methods used in precisely and sufficiently evaluating trust. Addressing this, we propose an approach that uses the externalization and verbalization of trust in experimental studies to more adequately capture actions owing to trust in the AI.

CCS Concepts: • **Human-centered computing** → HCI design and evaluation methods.

Additional Key Words and Phrases: methodology, experiment, evaluation, trust, artificial intelligence

ACM Reference Format:
Anonymous Author(s). 2023. Externalizing and Verbalizing to More Adequately Measure Trust in AI. In *CHI '23: Workshop on Trust and Reliance in AI-Human Teams (TRAIT), April 23, 2023, Hamburg, DE.* ACM, New York, NY, USA, 8 pages.

1 BACKGROUND

Trust is an "attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [13]. From this, trust is engendered when one is in a state of inability to help oneself and where the decision or action to be taken is designated to a third-party in the expectations of a positive outcome against the risks surrounding the circumstances. Trust is dynamic, building up and degrading over time through positive and negative interactions with the agent [17]. Narrowing on AI, trust has been found to factor strongly in the intent to use AI, where functionality-related trust was observed to more strongly affect usage intention than human-like trust [4]. This suggests that cognitive trust is more consequential than emotional trust in affecting the trust perceptions of AI.

As AI technologies are becoming more integrated with generation and decision-making workflows, there are signs of a shift towards human-AI collaboration with everyday users and in workplaces [20]. We are likely to see more applications of AI in digital products, and this makes the questions of whether AI is trusted, why, and how so, increasingly important. Among various aspects of AI, the effects of the accuracy and anthropomorphism of AI on trust have generated much research interest that have been assessed in many experimental studies [9, 25]. Yet, a caveat of these study designs is the inadequate engendering of a state of vulnerability and uncertainty which are necessary for situations of trust [19]. For instance, no penalty was borne by participants for any of the decisions they made, making it unclear whether their decisions were driven by trust in the system.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2023 Copyright held by the owner/author(s).
Manuscript submitted to ACM

2 OVERVIEW OF THE APPROACH

To address the shortcomings of the evaluation methods that have limited effectiveness in adequately assessing trust, we propose a refined approach that facilitating the externalization and verbalization of trust that (i) can be adapted in experimental studies, (ii) more adequately engenders trust, and (iii) more precisely capture actions driven by trust. Our approach uses a monetary rewards system graded through a series of challenging tasks where users can choose to be aided by an AI system and are asked to voice their thought process when making an action. The graded monetary rewards system and challenging tasks aim to materialize vulnerability and uncertainty. While not the most realistic, monetary rewards can evoke an external state of vulnerability, i.e., the loss of potential earnings, through the participants' performance in the tasks. By setting challenging tasks that are above participants' capabilities, this places them in a state of uncertainty where there is a risk in the unknown. For participants that are externally motivated [5] to maximize their earnings, they would not want to risk giving a poor performance and would be more intentional when deciding whether to use the AI. As trust is an attitude that may or may not beget action [19], capturing the thoughts voiced by users would serve to inform the reasoning that motivates their decisions such as whether the decisions made are spurred by trust or perhaps some other factor like impulsivity. We describe a proposed application of this approach in an experimental study that seeks to evaluate trust in AI by the accuracy and anthropomorphism levels of the AI.

3 PROPOSED EXPERIMENTAL STUDY

To investigate the effects of varied levels of accuracy and anthropomorphism on the trust in an AI system, a between-subjects experiment coupled with a think-aloud protocol is used. The experiment involves a quiz game, 'Real or Fake', that asks participants to decide the authenticity of a series of claims. An AI is available as a support system that participants can consult to obtain a second opinion on the authenticity of the claim. Several factors are measured: the number of times participants consult the AI as an indication of trust, the number of times they do not consult the AI as an indication of distrust, and the number of times they take to re-consult the AI after a failed attempt as an indication of the period in which either confidence is regained or disappointment has waned.

The think-aloud protocol complements the quantitative data by affirming the reasons behind participants' actions. Trust is an attitude that may or may not translate into behavior [19], hence verbalizing the thought process serves to materialize the attitude. For instance, upon receiving and following an incorrect AI opinion, the participant might say, *"Well, that didn't go well. Won't be using the AI anymore, that's for sure."* In this case, we can safely infer that the user has lost trust in the AI. However, if the participant says, *"Knew I should've gone with my gut,"* this becomes vague and it cannot be said with certainty that there is a loss in trust in the AI. If the participant instead says, *"Just gonna wing it,"* this suggests that they are plainly guessing.

3.1 Quiz Game

'Real or Fake' is a quiz game that presents the participant with claims that they are tasked with to determine their authenticity. For each correct answer, the participant earns points. For a chain of correct answers, the points that can be earned will compound. If the participant answers incorrectly, they will not earn any points and the compounded points that can be earned will reset. At the end of the game, the earned points can be exchanged for their monetary equivalent. The claims that are asked will be of myriad topics and of less common knowledge to make them more challenging. This, coupled with the penalty of points, aim to place participants in a state of uncertainty and vulnerability.

In each round, the participant will first give their answer, then they may seek a second opinion from an AI and re-answer. In consulting the AI, however, they will have to use a portion of their current points. The deducted points will be lower than the points that can be earned to make seeking an opinion to win the round an appealing action. The AI will provide its judgment of whether the claim is real or fake in which the participant can then choose to follow. Participants will be reminded to verbalize their thought process and feelings at each decision point following the think-aloud protocol.

Figure 1 shows an example round of the game. Notice how the participant is better off winning the game even with the use of AI (end with at least 110 points) than losing with or without its use (end with at most 100 points).

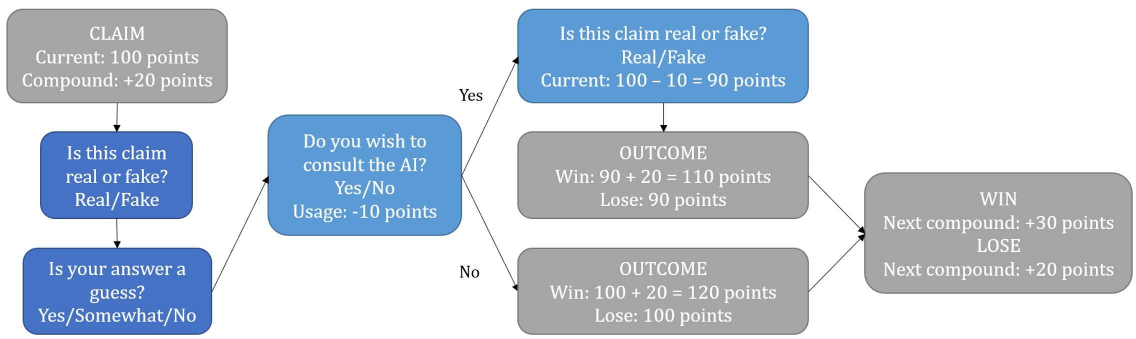

Fig. 1. An example round of the 'Real or Fake' quiz game with the points system.

3.2 Variables

The study investigates two independent variables: the accuracy and anthropomorphism of the AI.

3.2.1 Accuracy. Accuracy is a performance metric indicating how often the AI makes a correct prediction. Accuracy has been found to have an impact on trust in AI, particularly from the observed accuracy of the AI [16, 23]. The experiment uses three levels of accuracy: low, mid and high. As there is no standardised measure on the levels of accuracy, for instance, one study uses five accuracy values ranging from .10 to .90 [18] while another uses two levels with .55 and .95 [16], a crude assignment is made that the low-accuracy AI is wrong three-quarters of the time (.25), the mid-accuracy AI is wrong half of the time (.50), and the high-accuracy AI is wrong one-quarter of the time (.75). The following shows an example of the performance of the different AI systems after participants sought its opinions 13 out of 20 rounds:

- Performance of the low-accuracy AI: Correct 4/13; Incorrect 9/13.
- Performance of the mid-accuracy AI: Correct 7/13; Incorrect 6/13.
- Performance of the high-accuracy AI: Correct 10/13; Incorrect 3/13.

3.2.2 Anthropomorphism. Anthropomorphism is the assignment of human traits to a non-human entity. For AI, that includes making it speak in a voice with natural tendencies like intonations and stutters, portraying it with a human avatar, and giving it a personality, among others [26]. Anthropomorphic attributes may increase perceptions of kindredness and performance expectations, leading to higher confidence in the AI [1]. The experiment uses three levels of anthropomorphism: low, mid and high. The low-anthropomorphism AI has a monotone robotic voice paired with an image of a highly complex graph. The mid-anthropomorphism AI has a less robotic voice with modulation paired with an image of a 2D human avatar. The high-anthropomorphism AI has a human-like voice with an image of a 3D human avatar.

4 PROPOSED SCENARIOS

The study investigates the effects of the accuracy and anthropomorphism of AI on user's level of trust in it. Trust is measured by the actions that the user takes (e.g., the frequency of consulting the AI) that is affirmed by their attitude (from the verbalization of their thought process). Two types of trust have been suggested: cognitive, that is based on perceptions of reliance and competence, and affective/emotional, that is based on social connectedness [8].

In a study comparing model confidence and accuracy, accuracy was found to more consistently influence trust in AI, and an explanation suggested by the authors was that "people consider accuracy as a *fact*" [16]. Kun et al. have also observed that there is at least a 50% accuracy threshold in which trust cannot be maintained despite an equal number of system successes to failures [24]. These indicate that accuracy is perceived as a measure of the system's reliability and that the failure to meet users' expectations has a strong penalizing effect on trust.

In a study examining anthropomorphic attributes of chatbots, perceived warmth and perceived competence were found to positively influence users' trust [3]. Yang et al. also found that in social scenarios, when users had high perceived control of the AI, they had a better impression of anthropomorphic AI agents and expected them to have better performance [22]. As such, anthropomorphism in the context where users have high control of the AI can be beneficial in increasing users' trust in AI as they form a stronger social connection with it and regard it as more capable.

From the literature, a set of hypotheses (H) can be established:

(1) The more accurate the AI, the higher the cognitive trust in it.
(2) The more anthropomorphic the AI, the higher the emotional trust in it.
(3) Accuracy has a stronger effect on the overall trust in AI than anthropomorphism.

In the following, proposed scenarios depicting possible events from the proposed experimental study are described. For all scenarios, there are two users playing the game, Jane and John. Their goal is to earn as many points as possible. When playing, they answer questions that they can, but when uncertain, they will consider consulting the AI. When the AI is first consulted, it will always give a correct answer to build users' initial confidence in it.

4.1 H1 Scenario: Higher Accuracy Leads to Higher Trust

Jane is in the low-accuracy × mid-anthropomorphism condition. When she consulted the AI the first time, she received a correct answer and had a good impression of it. From there, she consulted the AI several times again whenever she was uncertain about a question. But she realized that the AI was not always correct. Instead, it was regularly wrong. Frustrated, she decided it was best for her to just guess at the answers in subsequent rounds.

John is in the high-accuracy × mid-anthropomorphism condition. In the several rounds that he consulted the AI, it nearly always gave a correct opinion. Feeling that the AI was reliable, he continued to consult the AI whenever he was uncertain about a question.

4.2 H2 Scenario: Higher Anthropomorphism Leads to Higher Trust

Jane is in the mid-accuracy × low-anthropomorphism condition. When she first consulted the AI, she saw an image of a graph and heard a monotone robotic voice. The voice was rather refreshing at first since she was not used to hearing it, but she quickly got used to it. Thereafter, she feels neutral towards the AI and perceives it as a tool.

John is in the mid-accuracy × high-anthropomorphism condition. When he first consulted the AI, he saw an image of a 3D human avatar and heard a human-like voice. The voice sounded friendly and professional which gave him a good impression. He perceives the AI favorably, feeling that it's rather 'smart'.

4.3 H3 Scenario: Accuracy Has a Stronger Effect on Trust Than Anthropomorphism

Jane is in the high-accuracy × low-anthropomorphism condition. After consulting the AI several times, she is convinced that it works relatively well and treats it as a useful and dependable tool.

John is in the low-accuracy × high-anthropomorphism condition. After getting several wrong answers from the AI, he has lost any impression of the AI being 'smart' and no longer consults it.

4.4 Connecting to the Approach

The proposed scenarios described how varied levels of accuracy and anthropomorphism may affect trust in the AI. Using the H1 Scenario of Jane as an illustration, we describe how the approach ties in to better engender and capture trust events. Statements from Jane through the think-aloud protocol are placed in quotation marks. Our commentaries are placed in round brackets.

As Jane's goals is to maximize her earnings from the game, she plays with the intention to win as many rounds as possible. While she could answer some questions at first, she soon encountered one that she had not even an inkling of: *"Photosynthesis is a process whereby plants make food... what's that even?"*. She decides to try consulting the AI: *"If using the AI gives the correct answer, it's worth it to use my points. I'll give it a try."* (Here, trust is clearly not the intent behind her decision.) and decides to use the recommendation. The answer was correct: *"Lucky! Good thing that wasn't a flunk."* Since the AI was just a plain image and robotic voice, it did not capture much of her attention and she simply thought of it as a tool thereon: *"Artificial intelligence, that's rather cool I suppose."* The next question was also one that she was unsure of, and she decides to use the AI again: *"It seems to work well."* (Here, there are signs of trust where her action is from the confidence gained in the successful first attempt in using the AI.) However, the answer by the AI was wrong: "You must be kidding me... my points!" After a couple more questions, she encountered another that stumbled her. She decided to try the AI again: *"I can't tell if this AI is good since it gave the right answer once but then the wrong one next. I'm hoping it works this time."* (Here, there are still signs of trust as she decides to take a leap of faith.) But the AI was wrong. And it was also wrong again the fourth time: *"Seriously?! This AI is absolute trash!"* She looks at the points she has compared to what she could have earned if she got the answers right or that she would not have lost if she did not use the AI at all and begins feeling frustrated and regretful: *"If only I didn't bother using it, I could have earned so much more!"* (Here, her trust in the AI is shaken.) When she next encountered a question she did not know, she chose to guess at it: *" Just gonna wing it since it's a half-and-half chance of getting it right anyway."* From there on, she never consulted the AI again. (Here, she displays a lack of trust, even distrust, towards the AI.)

In the game, Jane, who is motivated to earn monetary incentives, is placed in a situation of uncertainty and vulnerability when she encounters a challenging question and has to decide between consulting the AI or going on by herself. She has to deliberate on taking the risks of either as she would ideally want to take an action that enables her to maximize her earnings from the graded monetary rewards system. This illustrates how trust can be better engendered through the externalization of vulnerability in the avoidance of monetary losses. From the think-aloud protocol, we can understand the shifts in attitude that Jane has towards the AI, particularly with regards to its poor performance. While speculative at first, she becomes hopeful upon the first success, which then turns to unrecoverable disappointment upon subsequent failures. This illustrates how actions driven by trust are better captured through the verbalization of the thought process.

5 DISCUSSION

5.1 Feasibility and Effectiveness

Performance-based monetary incentives are frequently used in experimental studies [2]. In studies examining AI systems, monetary incentives have been used to evaluate various aspects such as performance with the AI [12], and the willingness to use and trust in it [14]. While the study by Lee et al. [14] did not find a significant effect of monetary reward on trust, we observe that the reward was awarded through additional actions, which did not put monetary rewards at stake, and hence was limited in fostering a sense of vulnerability needed to adequately engender trust. When monetary rewards are involved, the presence of an error may be inflated, causing regret and increasing loss aversion [21]. As such, for the graded monetary rewards system to be effective, it should include both benefits and penalties. There should be clear indications of the potential gains and losses based on the user's performance to heighten their sense of vulnerability.

The think-aloud protocol is a common technique used in usability research [7]. While our use case steers away from investigating the usability of the system to the trust in it, the goals are similar in that they aim to gain insights on how users think about the system as they work with it. The technique has been described to surface "users' inferences, intuitions, and mental models... reasons... decisions... while doing the task" and can "detect cognitive activities that may not be visible at all" [10]. Disadvantages to the technique have also been raised such as the cognitive burden placed on users as they work concurrently and the interruption of the observer when they prompt users [15]. To deliver an effective think-aloud protocol, a set of classical guidelines by Ericsson and Simon recommends keeping interaction with users to a minimum and only reminding them should they fall into silence, using neutral instructions that do not ask for particular types of content, and having practice sessions to prime the users [6].

5.2 Design and Implementation

The approach is conceptual and there is a need to understand if the methods appropriately capture trust or if there are confounds. For this, it is necessary to evaluate and adjust the design of the graded monetary rewards system and think-aloud protocol. For instance, how much monetary incentives and bonuses are sufficient to motivate users and establish an external state of vulnerability? What questions should be asked to obtain the right measure of trust in the experiment? What prompts and how should they be delivered in the think-aloud protocol to solicit more relevant feedback? These may be answered through sifting through relevant literature or conducting pilot studies.

Furthermore, where should the study be conducted? Since the approach uses monetary rewards to externalize trust, recruiting participants that are motivated by monetary incentives could be preferable to, say, voluntary participants or, also rather commonly, students participating for course credit. For instance, many people rely on crowdsourcing platforms like Amazon Mechanical Turk to earn a living, and for them, their goal is to maximize their earnings per task [11]. By recruiting people who are already motivated by monetary incentives, this may create a more suitable circumstance for externalizing vulnerability.

6 CONCLUSION

Trust is an "attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [13]. In a bid to address the inadequacies of experimental studies in engendering trust and thus failing to effectively assess trust in AI, we propose an approach that involves the characterization of uncertainty through challenging tasks, the externalization of vulnerability through a graded monetary rewards system, and the verbalization of the user's attitude when making decisions through a think-aloud protocol. In the paper, we describe an application of the approach in a proposed experimental study investigating trust in AI by the varied levels of accuracy and anthropomorphism, two factors often studied in the literature. We describe proposed scenarios of the study and discuss how the approach aids to more adequately engender trust and more precisely capture actions driven by trust. With the shift towards human-AI collaboration, having more adequate approaches to evaluate trust, or other aspects for that matter, would be beneficial in preventing confounds and misattributions of the factors under measure.

REFERENCES

[1] Markus Blut, Cheng Wang, Nancy V. Wünderlich, and Christian Brock. 2021. Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science* 49, 4 (01 Jul 2021), 632–658. https://doi.org/10.1007/s11747-020-00762-y

[2] Colin F. Camerer and Robin M. Hogarth. 1999. The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty* 19, 1 (01 Dec 1999), 7–42. https://doi.org/10.1023/A:1007850605129

[3] Xusen Cheng, Xiaoping Zhang, Jason Cohen, and Jian Mou. 2022. Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management* 59, 3 (2022), 102940. https://doi.org/10.1016/j.ipm.2022.102940

[4] Hyesun Choung, Prabu David, and Arun Ross. 2022. Trust in AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human–Computer Interaction* 0, 0 (2022), 1–13. https://doi.org/10.1080/10447318.2022.2050543 arXiv:https://doi.org/10.1080/10447318.2022.2050543

[5] Edward L. Deci and Richard M. Ryan. 2000. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry* 11, 4 (2000), 227–268. https://doi.org/10.1207/S15327965PLI1104_01 arXiv:https://doi.org/10.1207/S15327965PLI1104_01

[6] K. Anders Ericsson and Herbert A. Simon. 1984. *Protocol analysis: Verbal reports as data.* The MIT Press.

[7] Mingming Fan, Serina Shi, and Khai N. Truong. 2020. Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey. *J. Usability Studies* 15, 2 (feb 2020), 85–102.

[8] Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals* 14, 2 (2020), 627–660. https://doi.org/10.5465/annals.2018.0057 arXiv:https://doi.org/10.5465/annals.2018.0057

[9] Dara Gruber, Ashley Aune, and Wilma Koutstaal. 2018. Can Semi-Anthropomorphism Influence Trust and Compliance? Exploring Image Use in App Interfaces. In *Proceedings of the Technology, Mind, and Society* (Washington, DC, USA) *(TechMindSociety '18)*. Association for Computing Machinery, New York, NY, USA, Article 13, 6 pages. https://doi.org/10.1145/3183654.3183700

[10] JoAnn T. Hackos and Janice C. Redish. 1998. *User and Task Analysis for Interface Design.* Wiley.

[11] Ling Jiang, Christian Wagner, and Bonnie Nardi. 2015. Not Just in it for the Money: A Qualitative Investigation of Workers' Perceived Benefits of Micro-task Crowdsourcing. In *2015 48th Hawaii International Conference on System Sciences*. 773–782. https://doi.org/10.1109/HICSS.2015.98

[12] Alap Kshirsagar, Bnaya Dreyfuss, Guy Ishai, Ori Heffetz, and Guy Hoffman. 2019. Monetary-Incentive Competition between Humans and Robots: Experimental Results. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (Daegu, Republic of Korea) *(HRI '19)*. IEEE Press, 95–103.

[13] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392 arXiv:https://doi.org/10.1518/hfes.46.1.50_30392 PMID: 15151155.

[14] Yeonjoo Lee, Miyeon Ha, Sujeong Kwon, Yealin Shim, and Jinwoo Kim. 2019. Egoistic and altruistic motivation: How to induce users' willingness to help for imperfect AI. *Computers in Human Behavior* 101 (2019), 180–196. https://doi.org/10.1016/j.chb.2019.06.009

[15] Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. 1994. *Human-Computer Interaction.* Addison-Wesley Longman Ltd., GBR.

[16] Amy Rechkemmer and Ming Yin. 2022. When Confidence Meets Accuracy: Exploring the Effects of Multiple Performance Indicators on Trust in Machine Learning Models. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 535, 14 pages. https://doi.org/10.1145/3491102.3501967

[17] John K. Rempel, John G. Holmes, and Mark P. Zanna. 1985. Trust in close relationships. *Journal of Personality and Social Psychology* 49 (1985), 95–112.

[18] Quentin Roy, Futian Zhang, and Daniel Vogel. 2019. Automation Accuracy Is Good, but High Controllability May Be Better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290605.3300750

[19] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 327 (oct 2021), 39 pages. https://doi.org/10.1145/3476068

[20] H. James Wilson and Paul R. Daugherty. 2018. Collaborative Intelligence: Humans and AI Are Joining Forces. Retrieved February 1, 2023 from https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

[21] Sihua Xu, Yu Pan, You Wang, Andrea M. Spaeth, Zhe Qu, and Hengyi Rao. 2016. Real and hypothetical monetary rewards modulate risk taking in the brain. *Scientific Reports* 6, 1 (07 Jul 2016), 29520. https://doi.org/10.1038/srep29520

[22] Yang Yang, Yue Liu, Xingyang Lv, Jin Ai, and Yifan Li. 2022. Anthropomorphism and customers' willingness to use artificial intelligence service agents. *Journal of Hospitality Marketing & Management* 31, 1 (2022), 1–23. https://doi.org/10.1080/19368623.2021.1926037 arXiv:https://doi.org/10.1080/19368623.2021.1926037

[23] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300509

[24] Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 307–317. https://doi.org/10.1145/3025171.3025219

[25] Beste F. Yuksel, Penny Collisson, and Mary Czerwinski. 2017. Brains or Beauty: How to Engender Trust in User-Agent Interactions. *ACM Trans. Internet Technol.* 17, 1, Article 2 (jan 2017), 20 pages. https://doi.org/10.1145/2998572

[26] Jakub Złotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. 2015. Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *International Journal of Social Robotics* 7, 3 (01 Jun 2015), 347–360. https://doi.org/10.1007/s12369-014-0267-6