Investigating How People Perceive AI-Generated Responses in Social QA

HYUNWOO KIM, School of Computing, KAIST, Korea

YOONSEO CHOI, School of Computing, KAIST, Korea

MINSU PARK, Division of Social Science, New York University Abu Dhabi, United Arab Emirates

With the recent advances in large language models (LLMs), LLMs are now replacing humans as an information source to ask questions, even in social Q&A forums. Despite the current trend, it would still be essential to cooperate between humans and AI to co-produce knowledge in social forums while leveraging the strengths of both parties. It is widely known that human favoritism and AI aversion, where people are known to be more favorable to human responses rather than those of AI, is prevalent. However, it is still underexplored how people's expectations towards humans and AI differ in terms of diverse attributes of the response and the topics. In this work, we present our approach to understanding the differences in people's expectations of the responses from humans or AI. To do so, we first collected various advice-seeking questions and people's preferences for the responder for each question. Based on the collected questions, we present our preliminary design of an online controlled experiment to quantitatively measure the difference between people's expectations of human responses and AI responses across various topics. With the experimental findings, we envision providing design considerations to support human-AI collaboration in knowledge co-production.

Additional Key Words and Phrases: Do, Not, Us, This, Code, Put, the, Correct, Terms, for, Your, Paper

ACM Reference Format:

Hyunwoo Kim, Yoonseo Choi, and Minsu Park. 2018. Investigating How People Perceive AI-Generated Responses in Social QA. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Social Q&A platforms have long been playing a crucial role in the accumulation and dissemination of knowledge and information on the web by enabling individuals from diverse backgrounds and knowledge levels to ask questions and share answers to build a shared knowledge base [9]. In the early days of the internet, these platforms were invaluable, providing a space where people could seek answers directly from people rather than searching for only sparsely available information present on the web [13]. Over time, the scope of questions on these platforms has evolved, encompassing inquiries seeking personal experiences, professional knowledge, or nuanced perspectives that only people can provide. As a result, social Q&A sites have played a crucial role in the accumulation and dissemination of knowledge and information across the web.

With the recent advances in large language models (LLMs) capable of understanding complex questions of humans and generating appropriate responses, AI models are now replacing humans as an information source to ask questions.

Authors' Contact Information: Hyunwoo Kim, khw0726@kaist.ac.kr, School of Computing, KAIST, Daejeon, Korea; Yoonseo Choi, yoonseo.choi@kaist.ac.kr, School of Computing, KAIST, Daejeon, Korea; Minsu Park, mp5500@nyu.edu, Division of Social Science, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates.

Manuscript submitted to ACM

For example, there has been a noticeable decrease in activity on StackOverflow, one of the most popular platforms for asking programming questions, after the release of ChatGPT [2]. Quora introduced the bot to quickly respond to people's questions [1]. In online forums, like Reddit, there has been a rise in bot activity, highlighting the growing presence of AI in these spaces.

Despite these changes, it is still crucial that the rise of AI should not marginalize humans as information providers to keep enough presence and identity of humans in online spaces and prevent model collapse [15]. To foster cooperation between humans and AI in social forums for co-producing knowledge while leveraging the strengths of both parties, it is essential to understand how people's perceptions of the capabilities of AIs and humans for providing appropriate responses to the users' questions. It is widely known that people tend to underrate AI's performance and favor human performance, known as AI aversion and human favoritism [6]. However, we still lack a comprehensive understanding of how users' expectations differ between AI and human contributions. What do users expect from AI's performance, and what do they value in human input? By exploring these questions, we envision better facilitating human-AI cooperation and creating positive, technology-enhanced knowledge-sharing environments.

To this end, we collected examples of advice-seeking questions from crowd workers, along with their preferences on whether they'd ask the question to another person on the web or the AI. We analyzed the topics and themes of the questions to investigate the differences in preferences by the topics. Then, with a controlled experiment, we plan to investigate how people's evaluation of the responses depends on the authors' identity regarding diverse quality factors. We further plan to explore how the authors' identity affects the importance of each quality factor in evaluating the overall quality of the response. Based on the results, we aim to devise a set of design considerations for leveraging AI for producing high-quality knowledge in social Q&A.

2 Background

A large body of research on human favoritism and AI aversion has primarily focused on scenarios where individuals make decisions based on algorithmic suggestions [7, 11]. However, with the advent of large language models (LLMs), there has been increasing interest in understanding whether these tendencies of human favoritism and AI aversion persist when LLMs are used to generate productive content. The existing research in this area has produced mixed results, suggesting that the relationship between human favoritism and AI aversion is more complex than initially thought.

Bohm et al. [5] compared human-generated and AI-generated advice on personal and societal challenges and found that while there was a negative bias against the competence of AI authors, this bias does not extend to people's evaluation of the advice quality or willingness to adopt it. Yin et al. [16] showed that AI-generated responses were generally perceived as making people feel more heard compared to human-generated ones, but the effect diminished when participants were aware that the response was generated by an AI. In the context of generating persuasive texts, Zhang and Gosline [17] found that when the author's identity was unknown, AI-generated or AI-revised content receives higher satisfaction ratings. However, when the identity was disclosed, human-created contents were rated more highly. Rae [14] reported that AI involvement in content creation across topics such as news, travel, health, and jokes led to the perception that the creator was less qualified and the content was less satisfactory compared to human-generated content. Finally, Bauer et al. [3] found that creative images generated by design students using generative AI were rated lower in creativity and quality when the use of AI was disclosed.

These studies collectively suggest that while human favoritism remains strong, AI aversion is more nuanced and context-dependent, highlighting the need for further exploration into how AI can be effectively integrated into creative and productive domains without compromising the perceived value of the content. In this research, we aim to explore the presence of human favoritism and algorithmic aversion across a broader range of topics and explore how it would guide human-AI collaborative knowledge production on the web.

3 Social Question Collection Process

To understand the differences in people's expectations of humans and AI responses, we first collected a variety of advice-seeking questions and people's preferences on the responder for each question. To do so, we used the WikiSurvey [**?**] platform. The platform collects the preferences of the participants by presenting a pairwise comparison question with two questions as options and making them choose one option that they'd prefer to ask another human. Based on the results, the platform provides a rank of questions by the degree of preferring humans as the responder. Participants can also add any other question to the platform, which enables us to collect a wide variety of questions.

Considering the context of social Q&A forums and the findings on how algorithm aversion depends on the task type [6, 12], we devised a set of 10 seed questions spanning a variety of topics.

- If you've ever planned a surprise trip for someone, what advice would you give others?
- Need a gift for mom who doesn't like anything
- What should a young first-time dog owner do before adopting a dog?
- What's the difference between being financially responsible and being "cheap"?
- How do people actually lose weight with dieting and exercise?
- I have 3 months to prep for a half-marathon as a beginner... what are the best training plans?
- What are the first steps if I want to learn about wine?
- What are some practical ways an ordinary member of society can help with conservation efforts, aside from recycling, turning off the lights when possible, etc?
- How can I increase my employee retention?
- Does email marketing really work?

With the seed questions, we first ran five batches of question collection tasks. Ten participants recruited from Prolific participated in each batch. In each batch, participants were asked to first respond to 30 pairwise comparison questions. Then, we further asked the participants to add six advice-seeking questions to expand the question set. We repeated the batches till the set of questions was saturated.

After the collection sessions, we ran the final rating session with 50 participants from Prolific. In the rating session, each participant responded to 60 pairwise comparison questions.

After the collection, we manually reviewed the whole question set to remove questions that were not advice-seeking, such as seeking facts or subjective preferences. After review, we had 173 valid questions in total. Two authors organized the question sets by conducting affinity diagramming to find the frequently used domains. Then, we grouped them into three high-level categories based on people's responder preference: 'Preferred to ask a human', 'Mixed preferences' and 'Preferred to ask AI'. The preference categories, domains and example questions are shown in Fig. 1.

4 Approach

Based on the collected questions, we plan to run an online controlled experiment to understand how people's expectation of human responses and AI responses differs. We first expect that people would have preferences for human responses compared to AI and tendencies to evaluate the quality of the human responses higher than that of AI, based on the

Preference	Domain	Example Questions
Preferred to ask a human	Family	- How do you decide to have another child? - How to cope with grief after family death?
	Relationship	- How do I break up with someone in a kind way? - What should I say to someone that I just got in a fight with to make them feel better?
	Self-improvement	- How do you stay motivated when working on long term projects? - How can I effectively balance work and personal life?
Mixed preferences	Health	- How did you keep the weight off after weight loss? - How to improve the quality of sleep? - What is the best workout plan for a beginner in the gym?
	Career	- What is the best strategy to be promoted quickly? - What jobs have the best salary?
	Pet	- What should a young first-time dog owner do before adopting a dog? - How can I keep my dog entertained during the summer? - How long should a new puppy be sleeping in average?
	Starter Tips	- Which infant products would you advise new parents to skip purchasing? - What are the first steps I need to take before learning to drive? - Please suggest some reading matter if I want a basic introduction to classic literature
Preferred to ask AI	Business	- How can I increase my employee retention? - Is digital business profitable?
	Finance	- How should I start investing with £300 a month? - How to split my money between accounts to get the most interest?

Fig. 1. Question preference collections after thematic analysis.

existing work on AI aversion. We also expect that people would have different expectations of the capabilities of humans and AI, such as humans would be better at providing emotional empathy while AIs would be better at providing objective and neutral responses [12, 16]. Furthermore, existing work suggests that people have different perceptions of the capabilities of AI depending on the task [6, 12], which affects the degree of human preference. Similarly, we expect that people will have less human favoritism when they perceive the topic of the question as something that could be dealt with well by AI. Hence, we propose the following set of research questions:

- RQ1. How would people evaluate the quality of answers differently based on the perceived identity of the answerer?
- RQ2. How would people have different preferences on the answers based on the perceived identity of the answerer?
- RQ3. How would people perceive the importance of each quality factor for the answer differently based on the perceived identity of the answerer?
- RQ4. How would the effect size of perceived identity differ by the perception on the topic of the question?

In the experiment, we plan to have three experimental conditions: Human, AI, and Neutral. In the Human and AI conditions, the response would be presented as either a human-written response or an AI-generated response. In the Neutral condition, the response would not contain any information on the author's identity. To decide the questions and the answers presented to the participants, we plan to randomly sample three questions from each high-level category

— ‘Prefer to ask a human’, ‘Mixed feelings’, and ‘Prefer to ask AI’ — and collect one response for each question from existing online forums like Reddit.

During the experiment, the participant will be randomly assigned to one of the conditions and the questions. They would respond with their preference for the response in terms of willingness to upvote the post. Furthermore, they would evaluate the quality of the response in terms of a predefined set of quality factors. To devise a set of quality factors for the responses, we plan to review measures from a wide variety of existing literature including AI aversion [5, 14, 17], social Q&A analysis [8], and AI alignment [10]. Finally, for each quality factor, they would weigh the importance of each factor on their willingness to upvote the post.

5 Position Statement around Positech

In this study, we explored the potential for human-AI collaboration in knowledge co-production, focusing on how such interactions can be strategically leveraged to enhance knowledge creation dissemination. Here, we would like to discuss the potential human-AI collaboration in a knowledge production setting and how to leverage or mitigate human biases toward AI aversion and human favoritism.

In various social computing systems, such as social Q&A platforms, the involvement of AI is becoming inevitable with the advancing capabilities of LLMs. As human favoritism or AI aversion has been observed in diverse settings, such as content generation and advice provision, it is crucial to actively discuss and consider how to effectively collaborate with AI while considering such biases when designing these platforms. For instance, exploring which types or topics of knowledge can be effectively outsourced to AI could provide valuable insights when designing social Q&A platforms with AI involvement. We will further investigate whether aligning information resources with these human biases — such as providing human-generated content to those who prefer human authors — can create synergies that enhance the effectiveness of knowledge outsourcing. Understanding how to optimize human-AI collaboration within these systems aligns with Positech’s vision of creating positive, human-centered technological environments.

An interesting question to ask for designing future platforms for question answering is whether the bias (e.g., aversions) people hold towards AI are firm and solid or they can be mitigated through design interventions. By collecting the preferences of the question responders, we found that biases towards AI were strong and consistent. However, existing work also suggests some of the potential approaches to mitigate the bias, such as limiting the role of AI and framing the AI as having more expertise or experience [4]. An interesting discussion can be whether it is more effective to break biases or to work within them. On the one hand, breaking biases might open up new possibilities for human-AI collaboration, allowing AI to be integrated or complemented more deeply into domains traditionally dominated by humans. On the other hand, working within existing biases may yield immediate benefits by providing users with content that aligns with their expectations and preferences. In the future, we further envision investigating the effect of breaking or complying with the biases in different contexts to understand the impact of bias toward AI aversion in future studies.

Acknowledgments

References

[1] [n. d.]. More great answers on more questions. https://productupdates.quora.com/More-great-answers-on-more-questions

[2] [n. d.]. Stack Overflow Announces AI-Powered Features. https://www.i-programmer.info/news/99-professional/16487-stack-overflow-announces-ai-powered-features.html

[3] Kevin Bauer, Ekaterina Jussupow, Rebecca Heigl, Benjamin Vogt, and Oliver Hinz. 2024. All just in your head? Unraveling the side effects of generative AI disclosure in creative task. *Unraveling the Side Effects of Generative AI Disclosure in Creative Task (February 29, 2024)* (2024).

[4] Yochanan E. Bigman and Kurt Gray. 2018. People are averse to machines making moral decisions. *Cognition* 181 (Dec. 2018), 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

[5] Robert Böhm, Moritz Jörling, Leonhard Reiter, and Christoph Fuchs. 2023. People devalue generative AI's competence but not its advice in addressing societal and personal challenges. *Communications Psychology* 1, 1 (2023), 32.

[6] Noah Castelo, Maarten W. Bos, and Donald R. Lehmann. 2019. Task-Dependent Algorithm Aversion. *Journal of Marketing Research* 56, 5 (Oct. 2019), 809–825. https://doi.org/10.1177/0022243719851788 Publisher: SAGE Publications Inc.

[7] Lingwei Cheng and Alexandra Chouldechova. 2023. Overcoming Algorithm Aversion: A Comparison between Process and Outcome Control. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–27. https://doi.org/10.1145/3544548.3581253

[8] Hengyi Fu and Sanghee Oh. 2019. Quality assessment of answers with user-identified criteria and data-driven features in social Q&A. *Information Processing & Management* 56, 1 (Jan. 2019), 14–28. https://doi.org/10.1016/j.ipm.2018.08.007

[9] F. Maxwell Harper, Joseph Weinberg, John Logie, and Joseph A. Konstan. 2010. Question types in social Q&A sites. *First Monday* (July 2010). https://doi.org/10.5210/fm.v15i7.2913

[10] Hannah Rose Kirk, Alexander Whitefield, Paul Röttger, Andrew Bean, Katerina Margatina, Juan Ciro, Rafael Mosquera, Max Bartolo, Adina Williams, He He, Bertie Vidgen, and Scott A. Hale. 2024. The PRISM Alignment Project: What Participatory, Representative and Individualised Human Feedback Reveals About the Subjective and Multicultural Alignment of Large Language Models. arXiv:2404.16019 [cs.CL] https://arxiv.org/abs/2404.16019

[11] Hasan Mahmud, A. K. M. Najmul Islam, Syed Ishtiaque Ahmed, and Kari Smolander. 2022. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change* 175 (Feb. 2022), 121390. https://doi.org/10.1016/j.techfore.2021.121390

[12] Carey K. Morewedge. 2022. Preference for human, not algorithm aversion. *Trends in Cognitive Sciences* 26, 10 (Oct. 2022), 824–826. https://doi.org/10.1016/j.tics.2022.07.007

[13] Kevin Kyung Nam, Mark S. Ackerman, and Lada A. Adamic. 2009. Questions in, knowledge in? a study of naver's question answering community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 779–788. https://doi.org/10.1145/1518701.1518821

[14] Irene Rae. 2024. The Effects of Perceived AI Use On Content Perceptions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.

[15] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. 2024. AI models collapse when trained on recursively generated data. *Nature* 631, 8022 (July 2024), 755–759. https://doi.org/10.1038/s41586-024-07566-y Publisher: Nature Publishing Group.

[16] Yidan Yin, Nan Jia, and Cheryl J Wakslak. 2024. AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences* 121, 14 (2024), e2319112121.

[17] Yunhao Zhang and Renée Gosline. 2023. Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgment and Decision Making* 18 (2023), e41.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009