

1 Research interests

My research interests broadly lie in the influence of artificial intelligence (AI) agents on human decision-making. Specifically, I aim to develop **applications for conversational agents in decision-making support**. As part of this focus, during my master's program, I focused on supporting review writing and developed a system that assists in writing user reviews using an interview dialogue system. In this approach, the conversational agent first gathers product information such as users' impressions and opinions during the interview, to create reviews, facilitating the review writing process. Additionally, I conducted a **comprehensive evaluation** from the perspectives of system users and review readers. Although experimental results have shown that the system is capable of generating helpful reviews, the quality of the reviews still depends on how effectively the agent elicits the information from users. Therefore, I believe that personalizing **the agent's interview strategy** to users' preferences regarding the review writing process can further enhance both the user experience and the helpfulness of the review.

1.1 AI-assisted decision-making

In our daily lives, other people's persuasion or suggestions can affect how we make decisions. While these words can encourage beneficial decision-making, they can also lead to poor choices. For example, a friend's recommendation of a movie may prompt you to watch it, but your experience of the movie might be either satisfying or disappointing.

Decision-making support in AI applications aims to assist humans in making optimal choices by providing information and suggestions. The tasks it targets are wide-ranging, including recommendations (He et al., 2017), reservations (Budzianowski et al., 2018), e-mail writing (Fu et al., 2023), review writing (Bhat et al., 2023), and creative support for screenplays (Mirowski et al., 2023).

1.2 Evaluation method for AI agents

Decision-making tasks often have no absolute correct answers. Therefore, offline evaluations using datasets with ground-truth labels have limitations, making actual user studies crucial. In recent years, user studies have been

conducted actively via crowdsourcing platforms such as Amazon Mechanical Turk,¹ where participants typically use the system and provide feedback in exchange for compensation. Post-surveys such as questionnaires and interviews in the user study, allow us to measure user satisfaction.

When designing user study experiments, researchers must appropriately determine the demographics of the subjects and the amount of feedback to collect. For example, my experiment targeted 100 Turkers per condition. Similarly, Mirowski et al. (2023) focused on 15 theatre and film industry professionals who have worked in TV, while Fu et al. (2023) targeted 40 participants per condition. User studies can provide direct feedback on participant satisfaction, but it remains challenging to collect as much data as in offline evaluations. A small sample size can lead to reproducibility issues. I argue that a careful design is necessary to accurately measure effectiveness when the amount of collected data is limited. In addition, it is important to properly analyze the collected data and clearly indicate the scope to which the resulting conclusions can be applied.

1.3 Constructing an interview dialogue system

In interview dialogue, the interviewer aims to elicit information from the interviewee. Existing research has shown that conducting surveys using a chatbot platform can yield higher-quality responses than using a Web survey platform (Kim et al., 2019). In cases where information collection is crucial, such as in my research, interview dialogue systems can be a promising option.

An interviewer skillfully eliciting information from the interviewee is desirable, and researchers have used human evaluation to measure this ability (Zeng et al., 2018; Okahisa et al., 2022). One effective way to develop an interview dialogue system with such capabilities is to incorporate dialogue strategies. For example, follow-up questions are a skill possessed by competent interviewers. This skill is effective in eliciting more detailed information when the interviewee's responses are ambiguous or too concise. Additionally, changing the topic during the interview dialogue allows for the collection of a wide range of information.

¹<https://www.mturk.com>

Another way to conduct effective interviews is by adapting the interview dialogue system to the interviewee. This personalization allows the system to generate more relevant questions by considering the user's background, preferences, and behaviors. The interview dialogue system I developed elicits feedback and opinions from the interviewee about their experiences using a product. In this case, for example, personalizing the system to the interviewee enables it to elicit evaluations compared to products the interviewee has used in the past. An interesting question is how to extract user information and whether such a dialogue strategy can be generalized to other tasks and situations.

2 Spoken dialogue system (SDS) research

I expect that in the future, Spoken Dialogue Systems (SDSs) will have a more significant impact on human decision-making. Although text-based dialogue systems can control vocabulary, SDSs also can control intonation and tone. This controllability expands the range of expression in system responses, enhancing both approachability and persuasiveness. I believe this will also help improve task success rates in task-oriented dialogue (ToD). Achieving these advancements requires the development of speech synthesis technology that can incorporate intonation and tone, as well as technology that can accurately understand users' emotions and intentions.

3 Suggested topics for discussion

Here are some topics to discuss:

- What new possibilities can be enabled by extending current ToD systems to be multimodal?
- How can we ensure data privacy when personalizing SDSs?
- What methods are available for personalizing non-textual information in multimodal dialogue agents?

References

Advait Bhat, Saaket Agashe, Parth Oberoi, Niharika Mohile, Ravi Jangir, and Anirudha Joshi. 2023. Interacting with next-phrase suggestions: How suggestion systems aid and influence the cognitive processes of writing. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, IUI '23, page 436–452.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. In *Proceedings of the*

2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, pages 5016–5026.

Liye Fu, Benjamin Newman, Maurice Jakesch, and Sarah Kreps. 2023. Comparing sentence-level suggestions to message-level suggestions in ai-mediated communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, CHI '23.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, WWW '17, page 173–182.

Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, CHI '19, page 1–12.

Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, CHI '23.

Taro Okahisa, Ribeka Tanaka, Takashi Kodama, Yin Jou Huang, and Sadao Kurohashi. 2022. Constructing a culinary interview dialogue corpus with video conferencing tool. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pages 3131–3139.

Jie Zeng, Yukiko I. Nakano, Takeshi Morita, Ichiro Kobayashi, and Takahira Yamaguchi. 2018. Eliciting user food preferences in terms of taste and texture in spoken dialogue systems. In *Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-Food Interaction*. Association for Computing Machinery, New York, NY, USA, MHFT'18.

Biographical sketch



Yoshiki Tanaka is a first-year PhD student at the Graduate School of Informatics, The University of Electro-Communications. He is interested in supporting human decision-making using AI. He has participated in the AIWolf-Dial2024jp competition. He is supervised by Assoc. Prof. Michimasa Inaba.