Kaito Nakae
The University of Electro-Communications
Tokyo, Japan
n2530097@gl.cc.uec.ac.jp

1 Research interests

My research interests lie in **human-agent collaboration** and **user adaptation**, with a particular emphasis on adaptation in **real-time collaborative environments**. The field of collaborative systems aims to support human teams in completing complex tasks efficiently while ensuring natural and adaptive interaction experiences. I investigate how AI agents can function as effective partners by adapting to their human collaborators. A central focus of my research is the personalization of agent behavior based on user proficiency. This includes methods for adapting the agent's communication strategies according to the user's skill level and task experience.

To pursue this goal, I collected and analyzed a multimodal dataset of human-human interaction using the real-time collaborative cooking game environment (Wu et al., 2021; Liu et al., 2024). The chosen environment is characterized by its complex task mechanics and strict time constraints, which necessitate seamless coordination and elicit dynamic, natural collaborative behaviors such as role negotiation and error recovery. Through this analysis, I investigated how partners with different levels of task proficiency communicate and coordinate effectively. Based on the findings, I proposed practical design guidelines for future adaptive AI agents, enabling them to adjust their level of guidance and initiative in response to the user's proficiency.

1.1 Collaborative Dialogue Games

Several dialogue-based collaborative game environments have been developed to investigate how humans and agents coordinate via language. For example, in a Minecraft-based environment (Narayan-Chen et al., 2019), two players assigned the roles of architect and builder, collaborate via text chat, with one player (the architect) responsible for instructing the other. These scenarios commonly involve fixed asymmetric roles, where one participant leads and the other follows, which structures the interaction around instruction and compliance.

More recent work has shifted toward environments without fixed roles, where participants collaborate more symmetrically (Ichikawa and Higashinaka, 2023; Jeknic et al., 2024). In these settings, the absence of strict time constraints and the reduced need for immediate decision-making often make it easier to infer a partner's intentions and coordinate effectively. This shift facilitates research into more natural and flexible forms of collaborative dialogue.

1.2 Human-AI Collaboration

Collaborative task environments such as Overcooked have become widely used benchmarks for studying human-AI coordination. Among them, Overcooked-AI (Carroll et al., 2019) has received particular attention. It has become a standard platform for reinforcement learning research on multi-agent cooperation (Charakorn et al., 2020; Sarkar et al., 2022; Wang et al., 2024), due in part to its simplicity and emphasis on coordination through actions. However, Overcooked-AI employs significantly simplified task mechanics. For instance, actions like chopping and mixing ingredients are absent, and only a single cooking method is available. I observed that, due to the low task complexity, dialogue between human players in this environment often results in shallow and repetitive interaction patterns.

To analyze richer and more dynamic collaborative behaviors that emerge under complex, time-pressured conditions, my study instead employs another Overcooked-style environment (Wu et al., 2021; Liu et al., 2024). This environment introduces more complex task mechanics and enforces strict time constraints, enabling the study of natural and dynamic coordination behaviors such as role negotiation and recovery from errors. Such an environment serves as a rich testbed for exploring human-AI collaboration challenges that arise in fluid, real-time interactions rather than strict turn-taking scenarios.

1.3 User Adaptation

User adaptation plays a key role in effective collaboration, especially in real-time settings where diverse user strategies must be handled under time constraints (Puig et al., 2021; Zhang et al., 2024). Recent approaches have begun to address this by developing agents with sophisticated adaptation mechanisms. For example, the DPT-Agent framework (Zhang et al., 2025) combines fast, intuitive actions with deliberate reasoning, using a Theory of Mind (ToM) module to infer a partner's behavioral patterns and dynamically adapt its own collaborative style.

However, while there has been some progress in building cooperative agents, prior work has often assumed a uniform level of user proficiency, overlooking how collaboration dynamics change with varying skill levels. To

The 21th Annual Meeting of the Young Researchers' Roundtable on Spoken Dialogue Systems, pages 12–14
Aug 28–29, 2025. ©2025 Association for Computational Linguistics

address this gap, my research focuses on adaptively improving human-AI collaboration by accounting for the user's task proficiency.

2 Spoken dialogue system (SDS) research

I expect that future Spoken Dialogue Systems (SDSs) will evolve from simple assistants into effective collaborative partners, capable of engaging in real-time, hands-on tasks alongside humans. A key driver of this evolution will be the shift from text-based to voice-based communication. In visually demanding, high-pressure tasks such as the cooking game environment used in my research, text-based chat can increase the user's cognitive and visual load, forcing them to divert attention away from the task itself. Voice interaction frees the user's visual channel, allowing them to maintain focus on the shared workspace. This more natural communication modality can also foster a stronger sense of presence and engagement, making the AI feel more like a true partner.

However, to realize this potential, the field must address the critical challenge of real-time responsiveness. While speaking is a fast and natural input method for humans, the SDS must perform a complex pipeline of speech recognition, intention understanding, planning, and response generation, all within a fraction of a second to keep pace with the collaborative task. The inference latency inherent in today's large-scale models poses a significant obstacle to achieving this fluidity. Therefore, I believe a crucial research direction for the next generation of researchers is to develop novel SDS architectures that balance advanced reasoning capabilities with the stringent low-latency requirements of simultaneous human-AI collaboration.

3 Suggested topics for discussion

I suggest discussing the following topics:

- How should the "quality" of human-AI collaboration be evaluated?
- What kind of information is most crucial for personalizing dialogue systems in collaborative tasks?
- How can prosody, timing, or disfluency be used by dialogue systems to enhance user engagement in collaborative tasks?

References

Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems* 32.

Rujikorn Charakorn, Poramate Manoonpong, and Nat Dilokthanakul. 2020. Investigating partner diversification methods in cooperative multi-agent deep reinforcement learning. In *Neural Information Processing. ICONIP*. Springer, pages 395–402.

Takuma Ichikawa and Ryuichiro Higashinaka. 2023. Modeling collaborative dialogue in minecraft with action-utterance model. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: Student Research Workshop*. pages 75–81. https://doi.org/10.18653/v1/2023.ijcnlp-srw.10.

Isidora Jeknic, David Schlangen, and Alexander Koller. 2024. A dialogue game for eliciting balanced collaboration. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, Kyoto, Japan, pages 477–489. https://doi.org/10.18653/v1/2024.sigdial-1.41.

Jijia Liu, Chao Yu, Jiaxuan Gao, Yuqing Xie, Qingmin Liao, Yi Wu, and Yu Wang. 2024. Llm-powered hierarchical language agent for real-time human-ai coordination. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*. pages 1219–1228.

Anjali Narayan-Chen, Prashant Jayannavar, and Julia Hockenmaier. 2019. Collaborative dialogue in Minecraft. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, pages 5405–5415. https://doi.org/10.18653/v1/P19-1537.

Xavier Puig, Tianmin Shu, Shuang Li, Zilin Wang, Yuan-Hong Liao, Joshua B. Tenenbaum, Sanja Fidler, and Antonio Torralba. 2021. Watch-and-help: A challenge for social perception and human-ai collaboration. In *International Conference on Learning Representations*. https://openreview.net/forum?id=w$_7JMpGZRh0$.

Bidipta Sarkar, Aditi Talati, Andy Shih, and Dorsa Sadigh. 2022. Pantheonrl: A marl library for dynamic training interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence*. volume 36, pages 13221–13223.

Xihuai Wang, Shao Zhang, Wenhao Zhang, Wentao Dong, Jingxiao Chen, Ying Wen, and Weinan Zhang. 2024. Zsc-eval: An evaluation toolkit and benchmark for multi-agent zero-shot coordination. *Advances in Neural Information Processing Systems* 37:47344–47377.

Sarah A. Wu, Rose E. Wang, James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner. 2021. Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *Topics in Cognitive Science* 13(2):414–432.

Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B Tenenbaum, Tianmin Shu, and Chuang Gan. 2024. Building cooperative embodied agents modularly with large language models. In *The Twelfth International Conference on Learning Representations*.

Shao Zhang, Xihuai Wang, Wenhao Zhang, Chaoran Li, Junru Song, Tingyu Li, Lin Qiu, Xuezhi Cao, Xunliang Cai, Wen Yao, et al. 2025. Leveraging dual process theory in language agent framework for real-time simultaneous human-ai collaboration. *arXiv preprint arXiv:2502.11882* .

Biographical sketch

Kaito Nakae is a master's student at the Graduate School of Informatics and Engineering, University of Electro-Communications. He is interested in human-agent collaboration and user adaptation systems. He is supervised by Assoc. Prof. Michimasa Inaba.