

Human-AI Interaction in the Age of Large Language Models

Diyi Yang

Stanford University
353 Jane Stanford Way
Stanford, California, USA, 94305
diyy@stanford.edu

Abstract

Large language models (LLMs) have revolutionized the way humans interact with AI systems, transforming a wide range of fields and disciplines. In this talk, I share two distinct approaches to empowering human-AI interaction using LLMs. The first one explores how LLMs transform computational social science, and how human-AI collaboration can reduce costs and improve the efficiency of social science research. The second part looks at social skill learning via LLMs by empowering therapists and learners with LLM-empowered feedback and deliberative practices. These two works demonstrate how human-AI collaboration via LLMs can empower individuals and foster positive change. We conclude by discussing how LLMs enable collaborative intelligence by redefining the interactions between humans and AI systems.

Introduction

The development of Large Language Models (LLMs) has changed the way how humans interact with AI systems. As LLMs are trained on vast amounts of Internet data, they can comprehend and generate human-like text, enabling users to engage with AI through natural language in a conversational manner, as well as facilitating a variety of creative writing and customer service applications. We are witnessing a profound transformation in how humans interact with AI systems, a subfield known as **Human-AI Interaction**, which has been established even before the introduction of LLMs (Wu, Yang, and Santy 2023).

Human-AI Interaction covers multiple forms of interaction between human and AI systems, depending on the objective of the interaction. There are two major forms: *human-AI collaboration* where humans and AI work together to achieve a goal, and *humans utilizing AI-infused systems* where AI is used to facilitate human interaction. Related to this, *mixed-initiative interaction* (Horvitz 1999) emphasizes that humans and AI systems can take on the leading roles interchangeably in an interaction strategy.

As LLMs become more competent in their capabilities (OpenAI 2023), human-AI interactions are becoming more diverse with more diverse characters (e.g., tutors, AI-enabled chat) and functionalities (e.g., math, coding,

chess). As a result, it becomes more important to think about how insights from human-human interaction can be integrated into this process to support more effective human-AI interaction workflows, especially when LLM-simulated agents (Talebirad and Nadiri 2023) are involved in simulating participants (Park et al. 2023) or other roles such as malicious characters (Ganguli et al. 2022).

To enable positive human-AI interaction, from a methodology perspective, enabling seamless interactions between humans and LLMs is not trivial. Different prompting techniques such as decomposition to planning, refinement, and interaction (Cai et al. 2023; Li et al. 2023b), together with approaches like retrieval augmentation (Lewis et al. 2020) have been proposed to enhance different aspects (e.g., factuality and grounding) of interaction. From an evaluation perspective, human-AI interaction also poses new challenges as conventional performance driven metrics might not be enough (Lee et al. 2022). User-centered design and evaluation become more needed to measure not only these task-level performances, but also characterize diverse interaction dimensions such as usability, satisfaction, responsibility, as well as long-term effects on users.

Cast Studies of Human-AI Interaction

This work presents two case studies¹ to illustrate how LLMs can empower human-AI interaction. The first one explores how LLMs transform computational social science, and how human-AI collaboration can reduce costs and improve the efficiency of social science research. The second work looks at social skill learning via LLMs by empowering learners with LLM-empowered feedback and deliberative practices.

Case Study 1: How human-AI collaboration helps social science research Many language processing tasks can be successfully performed by LLMs in a zero-shot manner. If LLMs can help code social phenomena such as persuasiveness and political ideology, then they would transform Computational Social Science (CSS). We present a road map for the use of LLMs as CSS tools, including prompting best practices and an evaluation pipeline for evaluating the zero-shot performance of 13 language models on 24 representative CSS benchmarks (Ziems et al. 2023).

¹This symposium talk uses two case studies based on our previously published research, which I would like to acknowledge.

We further demonstrate how to harness the complementary strengths of humans and LLMs for the CSS research pipeline, given their strong zero-shot capability on many text-annotation tasks (Kuzman, Ljubešić, and Mozetič 2023). We dive deep into how annotation work can be best allocated among humans and LLMs to achieve both quality and cost objectives via *CoAnnotating* a novel paradigm for Human-LLM co-annotation of unstructured texts (Li et al. 2023a). We find that CoAnnotating to be an effective means to allocate work, with up to 21% superior performance compared to random allocation on multiple datasets.

Case Study 2: Using LLMs to help humans learn social skills Since social skills are mostly learned through expert supervision, it is difficult to scale training, especially given the shortage of trained professionals. We argue that LLMs can provide opportunities for people to practice and develop social skills that will help them in the workplace and in their personal lives. For instance, the success of peer counseling platforms depends on effective volunteer counselors, but many volunteers lack access to individualized learning resources. One of our recent works introduces CARE: an interactive AI coach that trains peer counselors with automatic suggestions (Hsu et al. 2023). CARE diagnoses which counseling strategies are most appropriate in the given context and suggests tailored responses during the practical training stage. We find that this LLM-based system, trained on Motivational Interviewing strategies from counseling conversation data, significantly helps novice counselors respond to challenging situations. One can use similar paradigm for other skills such as conflict resolution, for which we developed the Rehearsal system (Shaikh et al. 2023). Rehearsal helps users practice conflicts with a believable simulated interlocutor, identify alternative conversational paths, and learn through feedback on how to apply specific conflict strategies. Our between-subjects evaluation showed that Rehearsal significantly helps learners navigate later unaided conflict compared to control groups.

Conclusion

Using two case studies, this work illustrates how LLMs can enable collaborative intelligence by enabling humans and AI systems to collaborate together, as well as by enabling humans to learn social skills. Note that the design and development of human-AI interaction systems must address ethical challenges related to bias, harms, risks, and privacy concerns. Human-AI interaction goes beyond what we have covered here, including but not limited to: (1) designing new forms of interaction, (2) iteratively improving LLMs to enable better interactions through learning from interactions, (3) making human-AI interactions more personalized, (4) and analyzing how human-AI interaction might influence humans and change their behavior within a broader social context.

References

Cai, Y.; Mao, S.; Wu, W.; Wang, Z.; Liang, Y.; Ge, T.; Wu, C.; You, W.; Song, T.; Xia, Y.; et al. 2023. Low-code

LLM: Visual Programming over LLMs. *arXiv preprint arXiv:2304.08103*.

Ganguli, D.; Lovitt, L.; Kernion, J.; Askell, A.; Bai, Y.; Kadavath, S.; Mann, B.; Perez, E.; Schiefer, N.; Ndousse, K.; et al. 2022. Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. *arXiv preprint arXiv:2209.07858*.

Horvitz, E. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 159–166.

Hsu, S.-L.; Shah, R. S.; Senthil, P.; Ashktorab, Z.; Dugan, C.; Geyer, W.; and Yang, D. 2023. Helping the Helper: Supporting Peer Counselors via AI-Empowered Practice and Feedback. *arXiv preprint arXiv:2305.08982*.

Kuzman, T.; Ljubešić, N.; and Mozetič, I. 2023. Chatgpt: beginning of an end of manual annotation? Use case of automatic genre identification. *arXiv preprint arXiv:2303.03953*.

Lee, M.; Srivastava, M.; Hardy, A.; Thickstun, J.; Durmus, E.; Paranjape, A.; Gerard-Ursin, I.; Li, X. L.; Ladhak, F.; Rong, F.; et al. 2022. Evaluating human-language model interaction. *arXiv preprint arXiv:2212.09746*.

Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.; et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33: 9459–9474.

Li, M.; Shi, T.; Ziem, C.; Kan, M.-Y.; Chen, N. F.; Liu, Z.; and Yang, D. 2023a. CoAnnotating: Uncertainty-guided work allocation between human and large language models for data annotation. *arXiv preprint arXiv:2310.15638*.

Li, Y.; Hu, B.; Chen, X.; Ma, L.; and Zhang, M. 2023b. LMEye: An Interactive Perception Network for Large Language Models. *arXiv preprint arXiv:2305.03701*.

OpenAI. 2023. GPT-4 Technical Report. *arXiv:2303.08774*.

Park, J. S.; O’Brien, J.; Cai, C. J.; Morris, M. R.; Liang, P.; and Bernstein, M. S. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–22.

Shaikh, O.; Chai, V.; Gelfand, M. J.; Yang, D.; and Bernstein, M. S. 2023. Rehearsal: Simulating conflict to teach conflict resolution. *arXiv preprint arXiv:2309.12309*.

Talebirad, Y.; and Nadiri, A. 2023. Multi-Agent Collaboration: Harnessing the Power of Intelligent LLM Agents. *arXiv preprint arXiv:2306.03314*.

Wu, T.; Yang, D.; and Santy, S. 2023. Designing, Evaluating, and Learning from Humans Interacting with NLP Models. In Zhang, Q.; and Sajjad, H., eds., *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, 13–18. Singapore: Association for Computational Linguistics.

Ziem, C.; Held, W.; Shaikh, O.; Chen, J.; Zhang, Z.; and Yang, D. 2023. Can Large Language Models Transform Computational Social Science? *arXiv preprint arXiv:2305.03514*.