Putting Humans at the Center:
Human–AI Collaboration for Unstructured Data Analysis

Jie Gao Research Statement

Analyzing unstructured data, such as social media posts, customer reviews, and clinical interviews, is critical for informing policy, business decisions, and ensuring healthcare quality. Much of today's data, including over 90% generated by organizations, is unstructured, and its volume is growing rapidly. Traditionally, interpreting such data relies on human experts, which is time-consuming and labor-intensive. Recent AI advancements offer greater efficiency, yet they often fail to capture the depth and nuance of data as humans do. How can we analyze this data efficiently while keeping human-level nuance? Can AI produce results that are both scalable and trustworthy?

I argue that, to truly analyze and interpret unstructured data in ways that meet our needs and earn our trust, **we must move beyond relying solely on AI automation**. We must keep humans at the center of analysis, ensuring they remain actively engaged with the data. We must build AI systems to collaborate with and support humans at the cognitive level, such as sensemaking, reasoning, and reflection. We must ground these systems in well-established scientific methods to ensure rigor and reliability.

My vision is to achieve real-world societal impact by producing trustworthy AI outcomes through human-AI collaboration. Toward this goal, **my core approach** is to use Human-Computer Interaction (HCI) methods, such as interviews and user studies, to deeply understand expert practices and values. Guided by these insights, I design AI systems that act as collaborative partners to support key cognitive processes like sensemaking, reasoning, and reflection. Finally, I validate these systems through rigorous controlled experiments and mixed-methods analysis. To date, I have developed human-AI partnership systems for analyzing text [1, 2, 3] and code [4], grounded in real-world domain practices and theoretical frameworks from the social sciences and software engineering. The key motivating questions include:

- **Bringing AI to Humans: Integrating AI into Human Workflows** [1, 2, 4]: How do humans (e.g., data analysts, engineers, and decision-makers) envision AI in their workflows? What human-AI collaborations enhance efficiency without sacrificing depth and rigor? Ultimately, how should we evaluate the outcomes of human-AI collaboration?
- **Bring Humans to AI: Incorporating Humans into AI Generation** [3]: When should humans intervene in automation? Can this control yield more reliable results than pure AI? Ultimately, what forms of human involvement are truly meaningful?

Impact. Together, my work has laid the foundation for next-generation human–AI collaborative systems that improves human analytical capability with unstructured data. As one of the earliest researchers in this space, I have been working at a critical moment of AI-powered qualitative analysis. My publications at TOCHI [1] and CHI [2] have since become highly cited within HCI and have attracted researchers from software engineering, healthcare and education, demonstrating that the problems I aim to address are both timely and widely felt. My early exploration of general human–LLM collaboration [5] has also inspired diverse discussion. Meanwhile, our CHI 2024 workshop, "LLMs as Research Tools" [6], has brought together researchers across domains, highlighting the growing interest and impact of this line of work. In addition, I prioritize open-source development, public reuse, and reproducibility. **To date, my work has produced two human–LLM collaborative systems: a publicly accessible platform for trustworthy QDA, MindCoder.ai, and an open-source platform for collaborative QDA, CollabCoder.**

Agenda. As a faculty member, I intend to lead a research agenda in human–AI collaboration that produces AI results that are trustworthy by placing humans at the center of AI automation. I believe that achieving this agenda requires a multidisciplinary effort, which bridges **AI-, data-, system-centered fields** (e.g., NLP, Software Engineering) with **human-centered domains** (e.g., Psychology, Social Sciences). My experiences at Johns Hopkins University (JHU) and the Singapore–MIT Alliance (SMART), as well as my visiting experiences at NUS and Notre Dame have equipped me with the mindset and skills needed to collaborate effectively across these disciplines. **The long-term goal of my lab is to develop new forms of critical human–AI collaboration that drive society's AI transformation by advancing efficiency, fostering trust, and generating meaningful social impact.**

Future Research Agenda

As a faculty member I plan to lead a research lab that advances the vision of **"placing humans at the center"** for AI-supported work so that human trust, agency, and user preferences remain central as automation grows. I will pursue this through three directions.

System: Next-Generation Human–AI Collaborative Systems. Recent AI-powered systems are expanding human-AI interaction across core cognitive activities such as reasoning and decision making. Tools like Cursor show how AI can generate intermediate plans that humans can inspect and refine, revealing a shift toward shared cognitive processes. Despite this progress, most systems still position AI as either a passive helper or a fully autonomous actor, and we lack a scientific account of how humans and AI should share cognitive responsibilities. This gap limits our ability to design reliable and interpretable collaboration at scale. My first research direction, grounded in my recent work [5, 7], builds a theoretical and empirical foundation for the architecture of human-AI joint activity. I will study three questions. First, what collaboration patterns reliably appear when humans and AI work together on complex tasks. Second, how cognitive responsibilities can be allocated to preserve human autonomy and ownership while drawing on AI strengths. Third, how to characterize capabilities that are uniquely human or uniquely AI and design workflows that allow both to excel. To advance these questions, I will integrate literature review, agent-based simulation, system building, and controlled studies. Literature review will surface stable collaboration patterns and cognitive bottlenecks. Agent simulations and system prototypes will make it possible to explore design tradeoffs and coordination strategies that are difficult to observe in human only settings. Controlled studies will validate the new tools. Together this direction will produce design principles and empirical evidence that guide the next generation of novel AI systems.

Evaluation: Evaluation Frameworks for Human–AI Systems. While existing research has explored how humans and AI can work together by designing new forms of collaboration, evaluating analytical outcomes remains an open challenge. Human judgments can be inconsistent, existing quality assurance frameworks are difficult to operationalize, and emerging approaches such as "LLM-as-a-judge" often lack grounding in human expectations. We still do not know which dimensions people should use to evaluate analytical outcomes, how to ensure quality at scale, or how to formalize interpretability and analytical rigor in ways that both humans and AI can apply. My third research direction, grounded in my ongoing work [8], addresses this gap by developing frameworks for outcome quality evaluation. I will collaborate with experts in social science and psychology to construct theoretical frameworks rooted in qualitative research theory. I will build systems that instantiate these frameworks as scalable AI-based evaluators trained to approximate human criteria, and I will design hybrid human-AI evaluation workflows that preserve interpretive accountability while scaling to large datasets.

Data: Broader Sensemaking Domains. Many high value tasks in software engineering and AI research rely on cognitive processes similar to qualitative analysis. These include codebase understanding, analysis of AI agent trajectories, and more general multimodal information interpretation. Such tasks require people to read complex data, identify information of interest, and construct coherent understanding across distributed sources. Despite these shared cognitive processes, existing work rarely applies qualitative analysis methods or theories to these domains. This leaves several questions open. How can collaboration mechanisms developed for qualitative analysis be adapted to support sensemaking in code, agent behavior, or multimodal content. Which cognitive bottlenecks in these domains can be systematically supported by AI. How can AI transform raw complex information into representations that align with human reasoning strategies. My second research direction builds on my recent work on designing human AI collaboration for complex information extraction and presentation to developers [4]. I will work closely with experts in software engineering and AI to study real workflows in these settings. I will combine interviews, controlled laboratory studies, quantitative analysis, and system building to examine how people understand complex information and how AI can support this process. The goal is to design collaborative scaffolds that generalize human AI collaboration beyond qualitative analysis and enable people to reason more effectively across diverse forms of complex data.

References

[1] **Jie Gao**, Kenny Tsu Wei Choo, Junming Cao, Roy Ka-Wei Lee, and Simon Perrault. Coaicoder: Examining the effectiveness of ai-assisted human-to-human collaboration in qualitative analysis. *ACM Transaction on Computer Human Interaction (TOCHI)*, 2023. URL https://doi.org/10.1145/3617362.

[2] **Jie Gao**, Yuchen Guo, Gionnieve Lim, Tianqin Zhang, Zheng Zhang, Toby Jia-Jun Li, and Simon Tangi Perrault. Collabcoder: A lower-barrier, rigorous workflow for inductive collaborative qualitative analysis with large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI24)*, 2024. URL https://doi.org/10.1145/3613904.3642002.

[3] **Jie Gao**, Zhiyao Shu, Shun Yi Yeo, Alok Prakash, Chien-Ming Huang, Mark Dredze, and Ziang Xiao. Efficiency with rigor! a trustworthy llm-powered workflow for qualitative data analysis, 2025. URL https://arxiv.org/abs/2501.00775. **Under review**.

[4] **Jie Gao**, Yue Xue, Xiaofei Xie, SoeMin Thant, Erika Lee, and Bowen Xu. Understanding codebase like a professional! human-ai collaboration for code comprehension, 2026. URL https://arxiv.org/abs/2504.04553. Accepted at the 34th IEEE/ACM International Conference on Program Comprehension (ICPC 2026).

[5] **Jie Gao**, Simret Araya Gebreegziabher, Kenny Tsu Wei Choo, Toby Jia-Jun Li, Simon Tangi Perrault, and Thomas W Malone. A taxonomy for human-llm interaction modes: An initial exploration. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24. URL https://doi.org/10.1145/3613905.3650786.

[6] Marianne Aubin Le Quéré, Hope Schroeder, Casey Randazzo, **Jie Gao**, Ziv Epstein, Simon Tangi Perrault, David Mimno, Louise Barkhuus, and Hanlin Li. Llms as research tools: Applications and evaluations in hci data work. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24. URL https://doi.org/10.1145/3613905.3636301.

[7] **Jie Gao**, Thomas Malone, et al. Not just prompt engineering a taxonomy for human llm interaction modes, 2025. **In preparation**, manuscript.

[8] **Jie Gao**, Mian Zhong, Jonathan Ivey, Chien Ming Huang, Ziang Xiao, and Mark Dredze. Evaluation framework for llm assisted qualitative data analysis, 2025. **In preparation**.

[9] **Jie Gao**, Xiayin Ying, Junming Cao, Yifan Yang, Pin Sym Foong, and Simon Perrault. Differences of challenges of working from home (wfh) between weibo and twitter users during covid-19. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery. URL https://doi.org/10.1145/3491101.3519790.

[10] ShunYi Yeo, Gionnieve Lim, **Jie Gao**, Weiyu Zhang, and Simon Tangi Perrault. Help me reflect: Leveraging self-reflection interface nudges to enhance deliberativeness on online deliberation platforms. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. URL https://doi.org/10.1145/3613904.3642530.

[11] Junming Cao, Bihuan Chen, Longjie Hu, **Jie Gao**, Kaifeng Huang, Xuezhi Song, and Xin Peng. Characterizing the complexity and its impact on testing in ml-enabled systems. In *2023 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. URL https://doi.org/10.1109/ICSME58846.2023.00034.

[12] Zheng Zhang, **Jie Gao**, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. Visar: A human-ai argumentative writing assistant with visual programming and rapid draft prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery. URL https://doi.org/10.1145/3586183.3606800.

[13] Ruyuan Wan, Haonan Wang, Ting-Hao Kenneth Huang, and **Jie Gao**. From noise to nuance: Enriching subjective data annotation through qualitative analysis. In *Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+NLP)*. Association for Computational Linguistics, 2025. URL https://aclanthology.org/2025.hcinlp-1.20/.