

Impact Statement

Human-AI Collaboration in Qualitative Data Analysis

Jie Gao

I explain the my research impact along three dimensions: (1) the fundamental importance of the research problem itself, (2) the contributions and impact of my prior work, and (3) my future research impact and societal potential.

(1) Research Topic and Significance. My research aims to enhance and generalize qualitative data analysis (QDA)—a foundational analytical methodology widely used across disciplines and domains. QDA can be used in critical sensemaking and analysis tasks, such as analyzing clinical conversations between clinicians and patients, examining the reasoning trajectories of LLM agents to understand their successes and failures, and interpreting large-scale social media data to capture public perspectives and social dynamics. Consequently, QDA plays a pivotal role in healthcare, public policy, AI system development, and business decision-making, where qualitative insights often guide decisions with significant real-world consequences. Despite its importance, qualitative analysis remains labor-intensive, time-consuming, and difficult to scale, creating substantial barriers for researchers, practitioners, and broader user communities. **I argue that domain experts should be able to conduct qualitative analysis with substantially less effort and time; AI automated analysis must be reliable, transparent, and responsible; and general users should be able to access qualitative analytical capabilities with a significantly lower learning curve.** Achieving these goals requires moving beyond full automation toward the design of human-AI collaborative systems that preserve human judgment, interpretation, and agency while leveraging AI's strengths in scalability and pattern discovery. By positioning AI as a collaborator rather than a replacement, my work seeks to make qualitative analysis more efficient, trustworthy, and accessible—without compromising the rigor, accountability, and contextual understanding that this methodology demands.

(2) My Past Achievements and Impact. My prior work has pioneered AI-assisted QDA through human-AI collaboration, including [CoAIcoder](#) and [CollabCoder](#). Together, these systems advance a unified vision: enabling humans to collaborate with AI in ways that are intuitive, low-effort, and aligned with existing qualitative workflows, so that AI reduces manual and cognitive burden rather than increasing it. My first major work, [CoAIcoder](#) (TOCHI 2023), was the first to identify and formally define the emerging research area of AI-assisted human-to-human collaboration in collaborative QDA. By articulating this nascent space and empirically revealing fundamental trade-offs between efficiency and interpretive richness, the paper **reshaped how researchers conceptualize multi-human-AI collaboration in qualitative work**. As of December 15, 2025, CoAIcoder has received 95 citations, establishing it as a foundational and highly influential contribution that has inspired follow-up research across HCI, CSCW, software engineering, and education. My second major work, [CollabCoder](#) (CHI 2024), represents a significant methodological and systems advance. It presents the **first end-to-end, LLM-powered workflow explicitly designed to support all stages of inductive collaborative qualitative analysis**, including open coding, code merging, discussion, and code-group generation. Central to this system is a **transparent common ground that effectively supports diverse collaboration styles and mitigates power imbalances within teams**. As of December 15, 2025, CollabCoder has received 172 citations, making it a highly cited, field-advancing contribution that has motivated subsequent research across HCI, CSCW, healthcare, software engineering, and education. **Together, CoAIcoder and CollabCoder are now widely regarded as foundational works in AI-assisted qualitative data analysis.** Moreover, my initial investigation provided one of the [earliest systematic taxonomy](#) of human-LLM interaction modes, offering a shared vocabulary and conceptual structure for understanding how humans and LLMs collaborate across tasks and systems. As of December 15, 2025, this work has received 78 citations, positioning it as an early but influential foundation for the growing research area of human-LLM interaction design. Beyond individual works and systems, I have also co-organized the “[LLM as research tools](#)” workshop that helped establish shared foundations for this emerging area, articulating early research questions around LLM integration into data work, particularly concerning artifacts, evaluation practices, and ethical considerations. This effort has contributed to shaping the broader research agenda and fostering a growing interdisciplinary community around responsible, human-centered AI usage in data work.

(3) My Future Impact and Societal Potential. Building on this foundation, my future research agenda advances a cohesive vision of human-AI collaboration by investigating more complex architectures of human-LLM interaction modes and translating these insights into practical system designs. A central goal of my ongoing and future work is to ensure that AI tools are not only easy to use, but also responsible, transparent, and human-controllable, particularly in scenarios where data analysis quality is critical. This goal is exemplified by my ongoing project [MindCoder](#) and its publicly available system [mindcoder.ai](#). Methodologically, I will integrate literature reviews, interviews, participatory design, controlled experiments, and user studies to both design and rigorously evaluate human-AI collaborative systems. Through this agenda, I aim to shape the next generation of AI applications, systems, and frameworks that strengthen human agency while responsibly amplifying analytical capacity, with direct implications for healthcare, policy-making, AI development, and other societally critical domains where trustworthy understanding and interpretation of complex data are essential.