AI for Qualitative User Research: LLM-Mediated Collaborative Sensemaking

Junze Li
The Hong Kong University of Science and Technology
Hong Kong, China
jlijj@connect.ust.hk

Abstract

Qualitative user research in HCI involves methods to provide the deep, empathetic insights necessary to design technologies that align with human needs and values. However, the practice of qualitative inquiry is constrained by inherent cognitive and communicative limitations, such as memory decay, empathy gap, and interpretation opacity. To overcome these constraints, we explore and construct a new paradigm of AI-Mediated Collaborative Sensemaking for qualitative user research. First, we utilize generative inference of contextual information to anchor memory in diary studies. Second, we leverage real-time information synthesis and visual communication to align mental models during synchronous interviews. Third, we employ multimodal sensor fusion to reconstruct affective experiences during retrospective think-aloud protocols. The role of AI is shifted from a passive post-hoc analysis tool to a proactive, in-situ cognitive scaffold. Lastly, we propose a roadmap for future work, ensuring that AI-augmented research remains rigorously grounded while expanding the boundaries of human cognition.

CCS Concepts

• **Human-centered computing → User interface toolkits**; • **Computing methodologies → Artificial intelligence**.

Keywords

Human-AI collaboration, HCI research methods, Collaborative sensemaking

ACM Reference Format:
Junze Li. 2026. AI for Qualitative User Research: LLM-Mediated Collaborative Sensemaking. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3772363.3799210

1 Introduction

The challenges of qualitative user research in Human-Computer Interaction (HCI) lie in the inherent difficulty in accurately capturing, reconstructing, and interpreting the subjective lived experience of the user [2, 5, 6, 14]. Existing work has established and applied a set of research methods, such as diary studies, semi-structured interviews, and think-aloud protocols, aiming at bridging the gap between the researcher's understanding and the user's intent [1, 3, 15]. However, despite their popularity and validity, these methods rely heavily on fragmentary and misaligned communication channels and media, which may cause the gap between participants' ability to articulate their internal state and researchers' ability to interpret that articulation [4, 12, 13]. Verbalizing a past experience with complex in-situ emotion or an implicit cognitive process often forces participants to simplify or rationalize their experience, stripping away the nuance that is crucial for user-centered design procedures. Moreover, researchers are required to overcome the cognitive overload and bias during the data collection and analysis processes, making it challenging to capture every point of interest, connect related themes, and formulate proper follow-up questions simultaneously [11].

The integration of cutting-edge Artificial Intelligence (AI) techniques, such as Large Language Models (LLMs), into the user research paradigm offers a transformative potential to bridge these gaps, based on LLMs' information synthesizing, cognitive reasoning, and multimodal data fusion abilities [7, 9, 10]. In our research paradigm, the LLM-empowered system is positioned as a sensemaking mediator and facilitator. To achieve our ultimate goal of mitigating the sensemaking gaps between researchers and participants and explore potential of AI to facilitate the user research methods, we plan our research in the following dimensions:

- Temporal: DiaryHelper for the memory-experience gap in elicitation diary studies (CHI 24) [9]
- Interpersonal: InsightBridge for the empathy gap in real-time user interviews (CHI 25) [10]
- Multisensory: SenseFusion for the inference gap in retrospective think-aloud protocols (Ongoing)

In this paper, we introduce a theoretical and technical analysis of how LLM-empowered systems collectively facilitate collaborative sensemaking of user research. It organizes these contributions under a unified theoretical framework, evaluates their effectiveness through rigorous empirical studies, and critically examines the ethical implications of relying on algorithmic intermediaries to interpret human experience. Lastly, we discuss some future research directions of leveraging AI for qualitative user research.

2 DiaryHelper for Memory-Experience Gap through Contextualization (CHI24)

Diary studies are utilized for capturing behaviors, habits, and experiences over time in naturalistic settings. Unlike lab studies, they offer high ecological validity. However, they suffer from a critical trade-off between participant burden and data quality. To minimize

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3799210

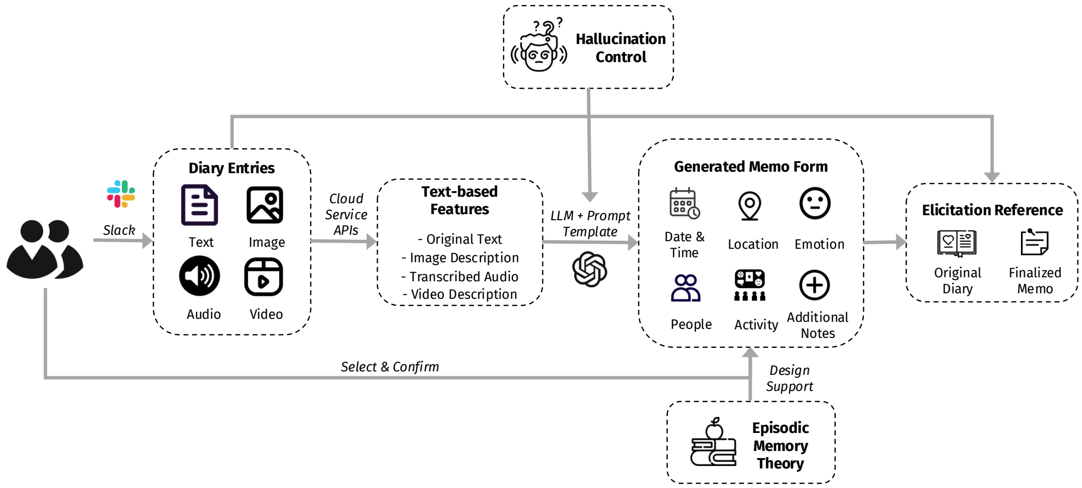

Figure 1: The system framework of DiaryHelper. The memory anchors are designed based on the episodic memory theory, and automatically inferred by LLMs.

disruption to their daily lives, participants often log events succinctly without capturing the rich contextual details. When the researcher interviews the participant a period of time later (*i.e.,* elicitation interview), the participant often cannot recall the specific circumstances and details of the entry. The participant may reconstruct the memory based on incomplete records rather than recalling the past reality, filling in gaps with generic assumptions rather than specific facts.

Therefore, we develop DiaryHelper (the system framework shown in Figure 1), an LLM-empowered agent for participants to capture multidimensional contextual information efficiently and accurately. This agent integrates into a standard diary logging platform. When a user inputs a brief log, the underlying generative AI model analyzes the entry and the user's historical patterns to predict five specific dimensions of contextual information: time, location, emotion, people, and activity. The selection of such dimensions is based on the episodic memory theory. Instead of asking the user to fill out a long form, DiaryHelper presents these predictions as a set of selectable tags or a conversational confirmation. The user can simply tap to confirm or make minor edits. This interaction significantly lowers the interaction cost of logging contextual information.

The results of this work provide evidence for the efficacy of AI scaffolding in longitudinal data collection:

- Abundance and accuracy: DiaryHelper assists participants in capturing abundant and accurate contextual information without significant burden. The contextual data acts as nuanced details that remind users of details they might have otherwise ignored.
- Reduced retrospection bias: During the post-study interviews, the entries created with DiaryHelper served as significantly stronger memory cues. Participants could look at the tags and immediately reconstruct the scene. And they could connect related scenarios integratively.
- Sensemaking perspective (Data-Frame Theory) [8]: From a sensemaking perspective, DiaryHelper acts as a pre-computation engine for sensemaking. In a traditional diary study, the participant provides the log, and the framing happens weeks later. DiaryHelper forces the participant to apply a framing of the specific context tags in the moment of experience. This prevents the frame from degrading or drifting in the participant's memory.

3 InsightBridge for the Empathy Gap through Real-time Synthesis (CHI25)

If diary studies suffer from a lack of contextual data, user interviews often suffer from an excess of it. A researcher conducting an in-depth interview must perform multiple high-demand cognitive tasks simultaneously: active listening to the user's narrative, transcribing key points, synthesizing information to form a mental model, and formulating relevant follow-up questions to probe deeper. Furthermore, verbal communication of real user scenarios may be ambiguous. When a user describes a process or a problem, the researcher constructs a mental image of that description. However, such an image may suffer from the researcher's own biases and assumptions. A mismatch between the user's description and the researcher's understanding creates an empathy gap.

To facilitate mutual understanding and empathy, we design InsightBridge, as shown in Figure 2, which processes the live audio stream of the ongoing interview and provides real-time assistance. It uses Automatic Speech Recognition (ASR) to transcribe the conversation in real-time. An LLM then analyzes this transcript to extract key informational entities and organizes them into a structured empathy map. The researcher can modify, move, and merge the sticky notes on the empathy map to organize their thoughts efficiently. Crucially, InsightBridge goes beyond text. It uses a Text-to-Image generation model to create visual abstracts that represent the user's narrative. The researcher and participant can negotiate and discuss visual abstracts to align their understanding during the interview. This creates a feedback loop where the user and

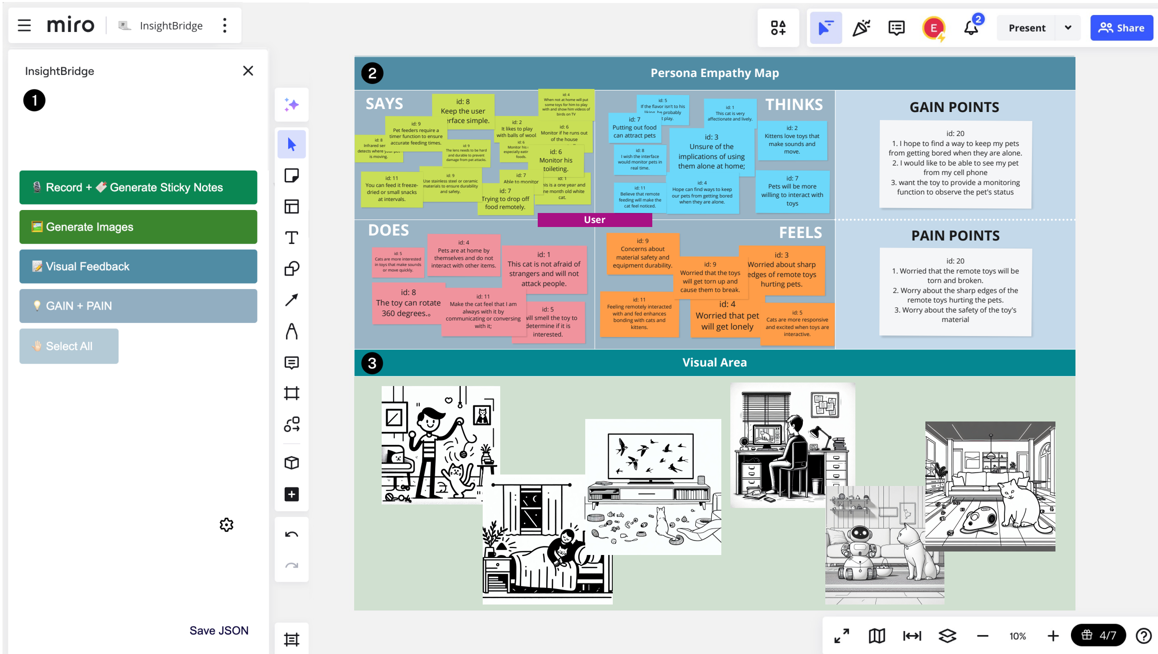

Figure 2: The user interface of InsightBridge system. It comprises an empathy map for real-time information synthesis and a visual sketch area for concrete visual-based sensemaking.

researcher can validate or correct their interpretations mutually in real-time.

The findings demonstrate that InsightBridge significantly enhances the quality of empathetic connection and data validity:

- Assistance in information synthesis: Researchers report significantly lower cognitive load regarding note-taking, as the system automatically and accurately captures and organizes the empathy map. This allows them to maintain better eye contact and emotional rapport with the participant.
- Prompting recall of overlooked details: The most profound finding is the effect of the visual abstracts on the participants. Users report that seeing the visuals prompted them to recall overlooked details and scenarios.
- Collaborative reframing and sensemaking: The visual abstract serves as a boundary object. Even when the AI generates an improper visual, it is still valuable, and the user is engaging in explicit reframing to further explanation. This correction process forces the user to articulate tacit knowledge that they might not have mentioned otherwise. The visual abstract creates a shared focal point for the collaborative sensemaking.

4 SenseFusion for Inference Gap through Multimodal Reasoning (Ongoing)

Think-aloud protocols are widely used in usability testing. However, concurrent think-aloud (CTA) can interfere with cognitive processes, especially in complex tasks. Retrospective think-aloud (RTA), where users watch a replay of their task and comment, mitigates this but introduces the memory-experience gap. Users watch the video and see what they did, but often forget why they did it or how they felt. Existing RTA solutions try to cue memory using gaze overlays or system logs. However, these only show behavior (external actions), not experience (internal states). Physiological sensors (*e.g.,* heart rate, GSR, pupil dilation) can capture these internal states potentially, but raw sensor data is noisy and unclear for users' interpretation.

In this work, we try to leverage multisensory data, screen context, and interaction logs collected when users execute the tasks, to facilitate the later debriefing of their in-situ experiences. Unlike previous systems that treat these multimodal data as separate streams, SenseFusion uses a Vision-Language Model (VLM) to fuse them. The system detects events of interest that users may have obvious changes in their internal cognitive and mental states, and presents such pieces of record for their in-depth debriefing. Instead of showing raw data in multiple views, the system supports natural language-based inquiry, search, and annotation. Some preliminary findings show that SenseFusion can promote RTA participants to provide more comprehensive and in-depth insights, particularly around personalized user needs, which paves the way for further research about user experience understanding.

5 Discussion

In this section, we will summarize the published and ongoing works through the human-AI collaboration perspective and the temporal relationship between data collection and sensemaking, and propose some future directions.

Table 1: The LLM-mediated collaborative sensemaking under the human-AI collaboration perspective.

Dimension	DiaryHelper (Longitudinal)	InsightBridge (Synchronous)	SenseFusion (Retrospective)
Research Phase	Data Collection (In-situ)	Data Collection (Interview)	Analysis/Debrief (Post-task)
Primary Gulf	Memory Decay	Empathy Gap	Interpretive Opacity
AI Role	Context Recorder	Synthesizer	Interpreter
Human Role	Verifier of Context	Co-creator of Meaning	Debriefing of Experience

5.1 Framework for LLM-Mediated Collaborative Sensemaking

Looking across DiaryHelper, InsightBridge, and SenseFusion, a trajectory emerges regarding the primary gulfs and roles of AI in diverse qualitative research methods as shown in Table 1. Such a temporal and modality coverage demonstrates a comprehensive framework where AI is not just a tool for one phase, but a persistent cognitive and sensemaking mediator that supports the qualitative user research lifecycle. The human-AI collaboration in user research is not only in collecting more data, but in collecting better data by using AI to preprocess, structure, and interpret the multimodal data before the human even begins the deep thought. The balance of the richness of the collected user data and the intrusiveness of users' ongoing tasks should be carefully considered in different scenarios.

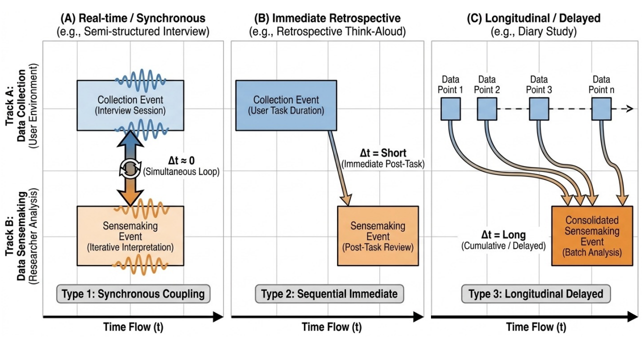

Figure 3: The temporal relationship between data collection and data sensemaking during different qualitative user research methods.

In Figure 3, we represent the temporal relationships between the user research data collection and the data sensemaking process. Under different scenarios, the amount, modality, and quality of the collected data are crucial to design the corresponding LLM-mediated tools, adapting to dynamic user-researcher interactive behaviours.

5.2 Future Directions

The involvement of LLMs in the user research may introduce potential biases, including contextual bias, agency bias, and cultural bias. Moreover, the interactions between humans and AI should not be limited to the digital and information space, but also the physical world and interpersonal social cues. The future direction of this doctoral work will focus on critical human-AI collaboration design through:

- Measuring and mitigating bias across multimodalities: A detailed ablation study could be conducted to examine the effect of each data modality and the relationship among modalities under different user groups and scenarios. We will try to distinguish between instantaneous context-specific reactions and longer-lasting affective or cognitive states of users to reconstruct a comprehensive understanding of in situ experiences, as their combined capture enables a holistic reflection of real-time human responses in naturalistic settings.
- Designing for adversarial sensemaking: we expect that humans and AI can discuss, debate, and negotiate during the collaboration process. The confidence score and alternative interpretations from AI could be provided to humans and trigger the iterative sensemaking. This preserves epistemic agency, where AI serves to widen the hypothesis space, but the user collapses it.
- Longitudinal user modeling: we will explore the application of contextual intelligent agents as personal assistants for long-term user modeling. The agents are not only for collecting user data in daily activities, but also proactively provide assistance and intervention to change the human and real world. Through the fine-grained analysis of information flow within the human-agent-real world framework, the modeling of the user is not limited to the digital space but also to the reaction from the physical space.

Acknowledgments

This work is supported by the Research Grants Council of the Hong Kong Special Administrative Region with Grant No. 16203421, No. C6044-23G, and the HKUST Sports Science and Technology Research Grant No. SSTRG24EG07. I am grateful to my supervisor, Prof. Xiaojuan Ma, and my collaborators for their generous support and valuable feedback. I also appreciate all the participants in my study for their great effort and insightful feedback.

References

[1] Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, et al. 2004. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* 37, 4 (2004), 445–456.

[2] Mohammed Alhamadi, Omar Alghamdi, Sarah Clinch, and Markel Vigo. 2022. Data Quality, Mismatched Expectations, and Moving Requirements: The Challenges of User-Centred Dashboard Design. In *Nordic Human-Computer Interaction Conference* (Aarhus, Denmark) *(NordiCHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 11, 14 pages. doi:10.1145/3546155.3546708

[3] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. 2003. Diary methods: Capturing life as it is lived. *Annual review of psychology* 54, 1 (2003), 579–616.

[4] Anders Bruun, Effie Lai-Chong Law, Thomas Dyhre Nielsen, and Matthias Heintz. 2021. Do You Feel the Same? On the Robustness of Cued-Recall Debriefing for User Experience Evaluation. *ACM Trans. Comput.-Hum. Interact.* 28, 4, Article 25 (July 2021), 45 pages. doi:10.1145/3453479

[5] Scott Carter and Jennifer Mankoff. 2005. When participants do the capturing: the role of media in diary studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 899–908. doi:10.1145/1054972.1055098

[6] Nanna Gorm and Irina Shklovski. 2017. Participant Driven Photo Elicitation for Understanding Activity Tracking: Benefits and Limitations. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1350–1361. doi:10.1145/2998181.2998214

[7] Jiaxiong Hu, Junze Li, Yuhang Zeng, Dongjie Yang, Danxuan Liang, Helen Meng, and Xiaojuan Ma. 2024. Designing Scaffolding Strategies for Conversational Agents in Dialog Task of Neurocognitive Disorders Screening. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 70, 21 pages. doi:10.1145/3613904.3642960

[8] Gary Klein, Jennifer K Phillips, Erica L Rall, and Deborah A Peluso. 2007. A data–frame theory of sensemaking. In *Expertise out of context*. Psychology Press, 118–160.

[9] Junze Li, Changyang He, Jiaxiong Hu, Boyang Jia, Alon Y Halevy, and Xiaojuan Ma. 2024. DiaryHelper: Exploring the Use of an Automatic Contextual Information Recording Agent for Elicitation Diary Study. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 818, 16 pages. doi:10.1145/3613904.3642853

[10] Junze Li, Yue Zhang, Chengbo Zheng, Dingdong Liu, Zeyu Huang, and Xiaojuan Ma. 2025. InsightBridge: Enhancing Empathizing with Users through Real-Time Information Synthesis and Visual Communication. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 167, 15 pages. doi:10.1145/3706598.3713640

[11] Ji-Ye Mao, Karel Vredenburg, Paul W. Smith, and Tom Carey. 2005. The state of user-centered design practice. *Commun. ACM* 48, 3 (March 2005), 105–109. doi:10.1145/1047671.1047677

[12] Talya Miron-Shatz, Arthur Stone, and Daniel Kahneman. 2009. Memories of yesterday's emotions: Does the valence of experience affect the memory-experience gap? *Emotion* 9, 6 (2009), 885.

[13] Mary M Omodei and Jim McLennan. 1994. Studying complex decision making in natural settings: using a head-mounted video camera to study competitive orienteering. *Perceptual and motor skills* 79, 3_suppl (1994), 1411–1425.

[14] David Swallow, Helen Petrie, Christopher Power, and Alistair D. N. Edwards. 2022. Using Photo Diaries to Elicit User Requirements from Older Adults: A Case Study on Mobility Barriers. In *Human-Computer Interaction – INTERACT 2015*. Springer-Verlag, Berlin, Heidelberg, 147–164. doi:10.1007/978-3-319-22701-6_11

[15] Maaike Van Den Haak, Menno De Jong, and Peter Jan Schellens. 2003. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology* 22, 5 (2003), 339–351.