

Balancing Human and AI Agency in Conversational AR for Collaborative Learning

Xiuqi Tommy Zhu

zhu.xiu@northeastern.edu

College of Arts, Media and Design, Northeastern University
Boston, MA, USA

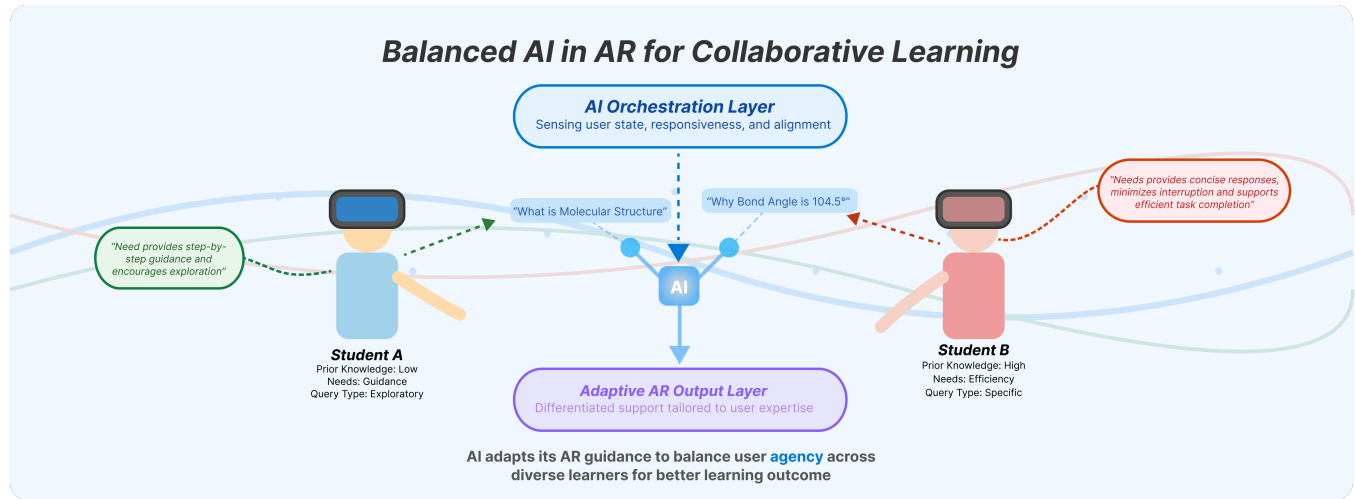


Figure 1: This teaser presents the research vision of my dissertation, which investigates how AI-driven AR systems adaptively balance user agency by tailoring guidance to individual expertise, enabling collaboration among learners with diverse knowledge levels.

Abstract

As AI-powered augmented reality (AR) smart glasses become increasingly prevalent, new opportunities emerge for supporting human collaboration through conversational user interfaces. However, existing systems are primarily designed for individual use, with a limited understanding of how AI should participate in multi-user interactions. My dissertation research investigates how AI agency can be designed and perceived through the context of collaborative AR learning. Specifically, it examines how different levels of AI initiative influence human–human collaboration, and how AI can support interaction without disrupting natural social dynamics. By exploring the balance between human and AI agency, this work aims to inform the design of conversational AR systems that enhance collaborative activities in real-world settings.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CUI '26, Bremen, Germany

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/2018/06

CCS Concepts

• Human-centered computing → Ubiquitous and mobile computing.

Keywords

AR Smart Glasses, Large Multimodal Model (LMM), Collaborative Learning, Agency

ACM Reference Format:

Xiuqi Tommy Zhu. 2026. Balancing Human and AI Agency in Conversational AR for Collaborative Learning. In *ACM Conference on Conversational User Interfaces (CUI '26)*, June 22–24, 2026, Bremen, Germany. ACM, New York, NY, USA, 3 pages.

1 Research Overview

The rise of wearable computing, especially Augmented Reality (AR)¹ smart glasses powered by Large Multimodal Models (LMMs), is transforming conversational AI into a form that can perceive users' first person visual perspective, location, and spoken commands. Combined with the ability to generate visually structured responses, such devices show strong potential to support people's everyday activities in a flexible and hands-free way [8]. However,

¹AR refers to the real-time integration of digital information into a user's environment [9]

most existing systems are designed for dyadic human–AI interaction, focusing on individual assistance, even though many real-world activities are inherently collaborative [5, 6, 10].

Drawing on prior CSCW research, conversational AI in AR may fail when designers overgeneralize from single-user assumptions and underestimate the complexity that arises when multiple people coordinate with AI [3]. In other words, Human–AI collaboration that functions well in isolation can falter when embedded in human–human collaboration, where agency continually and dynamically shifts and where the foundations of coordination, negotiation, and sustained cooperative activity become more complex. Thus, there is a need to build on ongoing AI+AR research on 1) multi-modal human input interactions (e.g., gaze + pinch) [7]; 2) proactive and passive AI interaction [4]; and 3) AR interface generation [1, 10], while extending these directions into collaborative contexts rather than applying them directly without adaptation.

These challenges point to a fundamental question: how should AI participate in collaborative interaction, rather than merely assist individuals? In multi-user settings, conversational AI is no longer a passive tool, but an active contributor that can initiate actions, shape interaction flow, and influence group dynamics. This raises critical design tensions, especially in AR, around when AI should intervene, how it should take from AR, and how its behavior affects human–human coordination.

My research adopts the concept of *agency* as a central lens. Here, agency refers not to human-like intentionality, but to the system’s capacity to influence collaborative activity through timing, initiative, and representation. Framing conversational AI in AR through agency allows HCI research to examine how action and control are dynamically distributed across humans and AI in collaborative settings. To ground this investigation, my dissertation focuses on collaborative STEM learning as a situated case, where coordination, shared understanding, and joint problem-solving are essential and common. This setting could examine how different configurations of AI agency will influence learning out, while generating insights that may extend to broader multi-user scenarios.

In general, my dissertation aims to answer the following core research question: ***How should AI agency be designed to support collaborative STEM learning in AR environments?***

This work is expected to contribute to the CUI and HCI communities in three ways: 1) Conceptual contribution: a framework for designing conversational AI in AR as part of AI support for human collaboration rather than dyadic human AI interaction; 2) Empirical contribution: an understanding of how different levels of AI agency affect collaborative learning in AR contexts; 3) Design contribution: the development of a collaborative AR system with balanced AI agency for learning contexts, along with derived design implications.

2 Progress to Date

I am currently in the third year of my PhD. I have completed a series of research projects that establish the foundation for my dissertation. My prior work, published at venues such as CHI, DIS, and ISMAR, has explored conversational agents in immersive environments, smart-glasses, and AI-mediated collaboration [2, 11–13].

These projects have provided both methodological and technical grounding for my current research. In particular, they have helped me develop experience in qualitative and quantitative methods, as well as in building end-to-end AR systems that integrate conversational AI. More importantly, they have revealed key challenges in real-world use, such as conversational breakdowns, misaligned AI intervention timing, and difficulties in maintaining shared understanding in collaborative contexts.

Building on this foundation, I am now transitioning into the core dissertation phase. My dissertation follows a three-phase research plan, and I am currently in the middle of Study 1, focusing on the design and early evaluation of an AI-supported collaborative AR learning system:

- **Study 1 (Design & Acceptance):** Co-design and iterative development of an AI-supported collaborative AR learning system with teachers and students, focusing on usability and classroom integration
- **Study 2 (Interaction):** Investigating how different levels of AI agency (e.g., proactive vs. passive) influence collaborative interaction, participation, and group dynamics
- **Study 3 (Outcomes):** Evaluating how AI agency impacts learning outcomes, including conceptual understanding and engagement

3 Doctoral Consortium Goals

My motivation for attending the Doctoral Consortium is to receive targeted feedback on my dissertation framing while my research is still evolving. As I transition into the core dissertation phase, I aim to refine both the conceptual positioning and methodological approach of my work within the CUI and HCI communities. Through this consortium, I hope to strengthen the theoretical grounding, methodological rigor, and overall research trajectory of my dissertation. I currently have several open questions that I hope to discuss with mentors and peers:

- (1) How should AI support for human collaboration be conceptualized within the CUI community? While my work frames conversational agents as embedded collaborators in collaborative AR settings, I seek guidance on how to distinguish this framing from existing work on dyadic human AI interaction and AI assistance.
- (2) How can AI agency be evaluated in collaborative contexts? My research explores different levels of AI agency (e.g., proactive vs. passive support), but I am still developing principled ways to measure their impact on the context.
- (3) What are appropriate methodological approaches for studying real world collaborative interactions with conversational agents? In particular, I am interested in discussing the challenges of conducting longitudinal and in situ studies with AR systems in real world educational settings.
- (4) How can the contributions of this work be positioned to resonate across CUI, CSCW and AR research communities? Given the interdisciplinary nature of my dissertation, I seek advice on how to communicate its contributions clearly to different audiences as my future career path.

References

- [1] CAI, R., JANAKA, N., KIM, H., CHEN, Y., ZHAO, S., HUANG, Y., AND HSU, D. AiGet: Transforming Everyday Moments into Hidden Knowledge Discovery with AI Assistance on Smart Glasses. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama Japan, Apr. 2025), ACM, pp. 1–26.
- [2] DING, J., ZHANG, Y., ZHU, X. T., YANG, K., WEI, Y., WANG, S., LIU, Y., AND JIAO, Y. Reshaping inclusive interpersonal dynamics through smart glasses in mixed-vision social activities, 2026.
- [3] GRUDIN, J. Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work* (New York, NY, USA, Jan. 1988), CSCW '88, Association for Computing Machinery, pp. 85–93.
- [4] LEE, G., XIA, M., NUMAN, N., QIAN, X., LI, D., CHEN, Y., KULSHRESTHA, A., CHATTERJEE, I., ZHANG, Y., MANOCHA, D., KIM, D., AND DU, R. Sensible Agent: A Framework for Unobtrusive Interaction with Proactive AR Agents, Sept. 2025. arXiv:2509.09255 [cs].
- [5] LEE, J., ALEOTTI, F., MAZALA, D., GARCIA-HERNANDO, G., VICENTE, S., JOHNSTON, O. J., KRAUS-LIANG, I., POWIERZA, J., SHIN, D., FROELICH, J. E., BROSTOW, G., AND VAN BRUMMELEN, J. ImagineAR: AI-Assisted In-Situ Authoring in Augmented Reality. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology* (Busan Republic of Korea, Sept. 2025), ACM, pp. 1–21.
- [6] LEE, J., TIAHJADI, A. D., KIM, J., YU, J., PARK, M., ZHANG, J., FROELICH, J. E., TIAN, Y., AND ZHAO, Y. CookAR: Affordance Augmentations in Wearable AR to Support Kitchen Tool Interactions for People with Low Vision. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh PA USA, Oct. 2024), ACM, pp. 1–16.
- [7] LIU, X., JIA, D., LIU, X. C., GONZALEZ-FRANCO, M., AND ZHU-TIAN, C. Reality Proxy: Fluid Interactions with Real-World Objects in MR via Abstract Representations. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, Sept. 2025), UIST '25, Association for Computing Machinery, pp. 1–16.
- [8] SUZUKI, R., GONZALEZ-FRANCO, M., SRA, M., AND LINDLBAUER, D. Everyday AR through AI-in-the-Loop. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2025), CHI EA '25, Association for Computing Machinery, pp. 1–5.
- [9] VERTUCCI, R., D'ONOFRIO, S., RICCIARDI, S., AND DE NINO, M. History of augmented reality. In *Springer Handbook of Augmented Reality*. Springer, 2023, pp. 35–50.
- [10] ZHU, C., HSIA, S.-K., HU, X., LIU, Z., SHI, J., AND RAMANI, K. agentAR: Creating Augmented Reality Applications with Tool-Augmented LLM-based Autonomous Agents. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology* (Busan Republic of Korea, Sept. 2025), ACM, pp. 1–23.
- [11] ZHU, X., WANG, C., GUO, Z., ZHAO, Y., AND JIAO, Y. "Can You Move It?": The Design and Evaluation of Moving VR Shots in Sport Broadcast, Sept. 2023. arXiv:2309.14490 [cs].
- [12] ZHU, X. T., CHEERMAN, H., CHENG, M., KIAM, S. R., CHUKOSKIE, L., AND MCGIVNEY, E. Designing VR Simulation System for Clinical Communication Training with LLMs-Based Embodied Conversational Agents. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2025), CHI EA '25, Association for Computing Machinery, pp. 1–9.
- [13] ZHU, X. T., LIU, X., HARTEVELD, C., DESAI, S., AND MCGIVNEY, E. Conversational successes and breakdowns in everyday non-display smart glasses use. *arXiv preprint arXiv:2602.22340* (2026).

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009