

From Human-Human Collaboration to Human-Agent Collaboration: A Vision, Design Philosophy, and an Empirical Framework for Achieving Successful Partnerships Between Humans and LLM Agents

Bingsheng Yao

b.yao@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Sherry Tongshuang Wu
sherryw@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Chaoran Chen

cchen25@nd.edu
University of Notre Dame
Notre Dame, Indiana, USA

Toby Jia-jun Li

toby.j.li@nd.edu
University of Notre Dame
Notre Dame, Indiana, USA

April Yi Wang

april.wang@inf.ethz.ch
ETH Zurich
Zurich, Switzerland

Dakuo Wang

d.wang@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

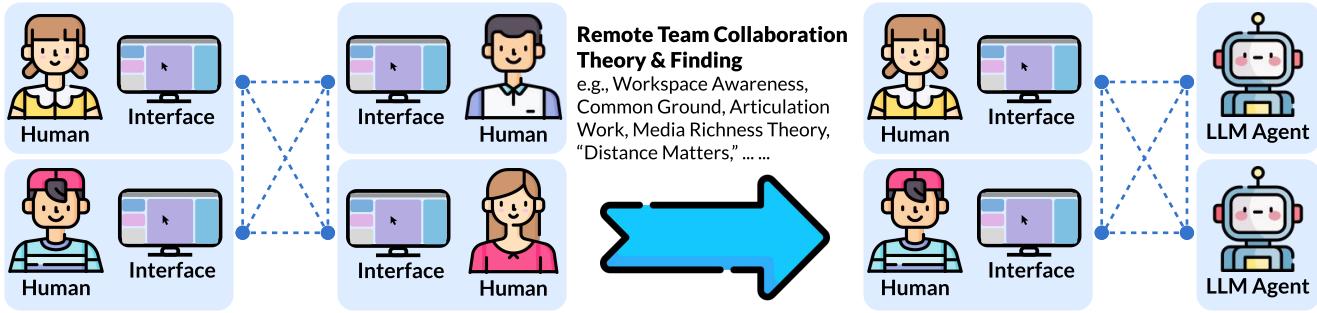


Figure 1: Our vision, design philosophy, and empirical framework: a novel research agenda from human-human collaboration to human-agent collaboration.

Abstract

The emergence of Large Language Model (LLM) agents enables us to build agent-based intelligent systems that move beyond the role of a “tool” to become genuine collaborators with humans, thereby realizing a novel human-agent collaboration paradigm. Our vision is that LLM agents should resemble remote human collaborators, which allows HCI researchers to ground the future exploration in decades of research on trust, awareness, and common ground in remote human collaboration, while also revealing the unique opportunities and challenges that emerge when one or more partners are AI agents. This workshop¹ establishes a foundational research agenda for the new era by posing the question: How can the rich understanding of remote human collaboration inspire and inform the design and study of human-agent collaboration? We

will bring together an interdisciplinary group from HCI, CSCW, and AI to explore this critical transition. The 180-minute workshop will be highly interactive, featuring a keynote speaker, a series of invited lightning talks, and an exploratory group design session where participants will storyboard novel paradigms of human-agent partnership. Our goal is to enlighten the research community by cultivating a shared vocabulary and producing a research agenda that charts the future of collaborative agents.

CCS Concepts

- Human-centered computing → Human computer interaction (HCI); Interaction paradigms; Collaborative interaction.

Keywords

human-agent collaboration, human-AI collaboration, remote collaboration, large language model, LLM agent

ACM Reference Format:

Bingsheng Yao, Chaoran Chen, April Yi Wang, Sherry Tongshuang Wu, Toby Jia-jun Li, and Dakuo Wang. 2026. From Human-Human Collaboration to Human-Agent Collaboration: A Vision, Design Philosophy, and an Empirical Framework for Achieving Successful Partnerships Between Humans and LLM Agents. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*, April 13–17, 2026, Barcelona,

¹Workshop website: <https://chi26workshop-human-agent-collaboration.hailab.io/>

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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI EA '26, Barcelona, Spain

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2281-3/26/04

<https://doi.org/10.1145/3772363.3778744>

Spain. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3772363.3778744>

1 Motivation and Goals

Computing has long been shaped by paradigm shifts that redefine the relationship between humans and intelligent systems [16, 17]. A foundational challenge in Human-Computer Interaction (HCI) and Computer-Supported Cooperative Work (CSCW) is moving beyond the “tool” paradigm, where systems act as passive instruments awaiting human instructions, to interactions that resemble genuine “teammate” [1, 16, 17, 21]. Systems that function as tools often lack the critical characteristics of effective human partnerships, such as mutual awareness [13, 18, 19] (a shared understanding of who is doing what), adaptivity [21] (the ability to dynamically adjust to a partner’s changing goals), and shared accountability [23] (a joint responsibility for outcomes). Without these qualities, interactions between humans and systems can be brittle and error-prone, making trust calibration difficult, and the potential for synergistic collaboration remains largely unrealized [1, 24].

Recent advances in Large Language Models (LLMs) mark a new inflection point. **LLM agents** can sustain open-ended natural language dialogue, exhibit reasoning and planning, and even evoke social or moral responses from users through role-playing [9, 25, 26, 28, 33, 35, 36]. At the same time, users often project emotions, empathy, and even moral expectations onto these systems [4, 14, 29], suggesting that interactions with LLM agents increasingly resemble partnerships rather than tool use. These dynamics suggest a shift toward **mixed-initiative collaboration** [21], where agents proactively coordinate, maintain common ground, and negotiate labor [3]. Yet, everyday failures already reveal the risks: code assistants that conflict with ongoing edits, writing tools that summarize without provenance, and hallucinations that silently undermine reliability [1, 22]. Without careful design, such failures could entrench tool-like brittleness and introduce new harms [6, 15, 32]. Our workshop is thus motivated by the urgent need to establish a principled, research-driven foundation for navigating this dual landscape of promise and risk.

We propose a **novel human-AI interaction design philosophy**: to design and study **LLM agents as remote human collaborators** [34]. This analogy grounds human-agent collaboration in decades of CSCW research on distributed teamwork [10, 11, 27]. In remote collaboration, humans communicate and coordinate through lean channels, lack shared physical context, and must rely on explicit signals to maintain common ground [10, 12], which is analogous to how humans interact with LLM agents. Adopting this vision raises various design questions: How can humans establish **common ground** with agents that lack lived experience Clark and Brennan [10]? What representations support **workspace awareness** of an agent’s focus and intent [18]? When should agents take on **articulation work**, and how should they expose it for oversight [30]? What forms of **media richness** enable effective negotiation despite lean communication [12]?

This vision structures our workshop around a dual inquiry: **foundations** and **frontiers**. Foundations involve re-examining CSCW theories of trust, awareness, and coordination to provide theoretical rigor for human–agent interaction [18, 24, 27]. Frontiers involve

identifying where the existing principles no longer stand and new paradigms arise, such as asynchronous collaboration (agents working in a user’s absence) and multi-human, multi-agent teamwork with complex organizational structure [28, 33].

We also acknowledge the limitations and boundaries of our “remote collaborator” vision and design philosophy. Agents lack embodiment, lived experience, and social accountability, and their errors manifest in unfamiliar forms such as fabricated citations, privacy leaks, value misalignments, or cascading failures [2, 6, 8, 15, 22, 32]. Rather than shortcomings, we view them as design challenges: how might we make failures legible, safeguard privacy, and build accountability when collaborators are not human [5, 7, 23, 31]? We also draw on the “beyond being there” argument to focus on benefits that are hard in human-only teams, such as persistent memory with provenance, machine-readable task contracts, and programmable social protocols [20]. The “remote collaborator” lens is therefore not a perfect mapping but a generative heuristic: it highlights where CSCW theories offer guidance while also signaling where entirely new frameworks are needed.

This workshop provides the first dedicated forum for the HCI community to collectively engage with this dual inquiry. We aim to create a shared vocabulary and a research agenda that is both grounded in foundational knowledge and visionary about future possibilities [1]. Our objectives are:

- (1) Establish the “agent as a remote collaborator” analogy as a bounded but powerful heuristic for research.
- (2) Assess the applicability of the core CSCW theories to emerging human-agent contexts.
- (3) Identify and ideate novel interaction paradigms that extend beyond traditional human-human models.
- (4) Integrate considerations of risks (security, privacy, trust mis-calibration) into the design space of collaboration.
- (5) Develop a community-driven research agenda that is suitable for the future of human-agent collaboration studies.

2 Organizers

The organizers of this workshop are interdisciplinary researchers in HCI, CSCW, and NLP with work on agent-mediated collaboration, mixed-initiative systems, privacy, and accountability. Our record includes leading CHI and CSCW publications, prior workshops organizations, the release of open resources for agent research, and empirical studies in software engineering, writing support, healthcare, and education. The workshop agenda grows directly from this experience and from recent HCI literature on human–agent collaboration, which together motivate the scope, methods, and expected outcomes of the event.

Bingsheng Yao is an Associate Research Scientist at Northeastern University’s Khoury College of Computer Sciences. His research aims to facilitate effective human-AI collaboration in domain-specific scenarios through the design and development of human-centered NLP systems, and to investigate emerging human-agent collaboration paradigms in which humans collaborate with LLM agents in mutual partnerships. He has recently co-organized the ACM CHI 2024 Special Interest Group on LLM privacy and a workshop at CSCW 2024 on LLM-based synthetic personae.

Chaoran Chen is a fourth-year Ph.D. Candidate in the Department of Computer Science and Engineering at the University of Notre Dame. His research uses human-centered design to investigate both the risks introduced by LLM agents and their potential as tools for improving user privacy and security. His work emphasizes the role of novel interaction and interface designs in empowering end users to navigate the evolving privacy and security landscape shaped by generative AI. He has served on the program committee of the 1st Workshop on Human-Centered AI Privacy and Security at ACM CCS 2025 and actively contributes to the HCI and usable security communities, with first-author publications at top-tier HCI and Security conferences.

April Wang is an Assistant Professor in the Department of Computer Science at ETH Zurich, where she leads the Programming, Education, and Computer-Human Interaction (PEACH) Lab. Her research in HCI and educational technology designs intelligent, literate programming and notebook-based environments that improve communication, collaboration, and learning in programming, including human-AI collaboration. She has served on the organizing committee and program committee of various venues, including VL/HCC, CHI, UIST, and L@S.

Sherry Tongshuang Wu is an Assistant Professor at the Human-Computer Interaction Institute, Carnegie Mellon University. Her primary research investigates how humans (AI experts, domain experts, and lay users) interact with (debug, audit, and collaborate) AI systems. Sherry has organized three workshops at NLP and HCI conferences: Shared Stories and Lessons Learned Workshop at EMNLP 2022 and Trust and Reliance in AI-Human Teams at CHI 2023-2024. She has also given three tutorials related to Human-AI Interaction at EMNLP 2023, NAACL 2024, and ACL 2025.

Toby Jia-Jun Li is an Assistant Professor in the Department of Computer Science and Engineering at the University of Notre Dame, where he leads the SaNDwich Lab. He also serves as the Director of the Human-Centered Responsible AI Lab at the Lucy Family Institute for Data & Society. He and his group use human-centered methods to design, build, and study human-AI collaborative systems. The group's projects have been generously supported by NSF, Google, AnalytiXIN, and other agencies. He has successfully organized workshops at multiple conferences, including CHI, CSCW, and IUI. He has also served on various organizing committees, including as the General Chair for ACM IUI 2025.

Dakuo Wang is an Associate Professor at Northeastern University, jointly appointed at Khoury College of Computer Sciences and the College of Arts, Media, and Design. His research lies at the intersection of human-computer interaction (HCI) and artificial intelligence (AI), with a focus on the exploration, development, and evaluation of human-centered AI (HCAI) systems. In addition to his research interest in human-AI collaboration and collaborative editing, he has been active in multiple steering committees for the HCI community, such as CHI 2019 Conference Co-Chair in Social Media, ACM CHI 2022 Conference Co-Chair in Special Interest Group, and CHI 2023 Conference Co-Chair in Workshop.

3 Workshop Structure and Activities

Our workshop will be a **Long (180-minute)**, fully in-person event with a target audience population of **40-50 participants**. It is

designed to provide theoretical framing, showcase diverse perspectives, and facilitate interactive design activities.

3.1 Pre-Workshop Plan

Participants will submit a short paper (2-4 pages, ACM single-column format). We welcome ongoing research, survey papers, and position papers that stimulate discussion on human-agent collaboration. Submissions will be reviewed by two program committee members and one organizer. Five papers will be selected as spotlights, with authors giving 3-minute spotlight talks during the workshop. The program committee will include experts across HCI, NLP, and social sciences, who will also help promote the workshop. We are committed to diversity, equity, and inclusion, and particularly encourage participation from underrepresented groups and those studying their needs.

3.2 Communication and Advertising Channels

We will maintain multiple communication channels, including a **website**, a **mailing list**, and a dedicated **Slack workspace**, allowing participants to be aware of the latest updates before, during, and after the workshop. The website will provide all critical information for our workshop, including the Call for Participation, submission instructions, program schedule, and contact information for the program committee. The website will also be updated regularly to advertise and disseminate workshop contributions, including accepted papers, featured research, and a post-workshop plan for a white paper (details in Section 3.4). In addition, we will advertise our workshop through the **mailing list of research institutions and related conferences** in interdisciplinary research (e.g., IUI, CHI, CSCW, ACL), as well as **direct communication** with collaborators and colleagues through social media and in-person meetings.

3.3 Workshop Schedules

Table 1 presents the schedule of our proposed workshop, which spans two sessions with a total of 180 minutes. We will adjust the acceptance rate and structure accordingly, depending on the submission amount. For instance, if many more applicants than expected show interest in participating after the advertisement. Below is a detailed breakdown of the planned activities.

3.3.1 Introduction (5 min). The organizers will frame the workshop's theme and introduce the "remote collaborator" vision for the new paradigms of partnership between humans and LLM agents.

3.3.2 Keynote (30 min). We will feature a keynote by a leading scholar whose work bridges HCI, CSCW, and AI in the context of human-agent collaboration. The talk will provide a theoretical anchor for our discussions, reflecting on foundational principles of collaboration and how they might be challenged or extended in an era of AI agents as equal partners. The talk is expected to span 25 minutes with 5 minutes for Q&A.

We have reached out to three keynote speakers, who are supportive of joining our workshop, but pending the acceptance of this proposal and their availability. They are Diyi Yang from Stanford University, who would be able to share insights on the intersection of technical development and computational social science; Michael Bernstein from Stanford University, who would offer a

Table 1: Tentative schedule for the 180-minute workshop. The workshop combines knowledge-sharing (keynote, panel discussion, lightning talks, spotlight papers, poster session) with interactive group activities (Session 2).

Time Slot	Activities
Session 1	Introduction
	Keynote with Q&A (Diyi Yang, Stanford University, <i>tentative</i>)
	Panel Discussion
	Panelist (<i>tentative</i>): Dakuo Wang, Toby Jiajun Li, April Wang
	Featured Lightning Talks:
	1. <i>Through the Lens of Human-Human Collaboration: A Configurable Research Platform for Exploring Human-Agent Collaboration</i> (Bingsheng Yao, Northeastern University) 2. <i>Dark Patterns Meet GUI Agents: LLM Agent Susceptibility to Manipulative Interfaces and the Role of Human Oversight</i> (Chaoran Chen, University of Notre Dame) 3. ... 4-6. ... (<i>Authors of three to four selected workshop submissions will present their work.</i>)
3:00 pm-3:30 pm	Coffee Break (Concurrent with poster session)
Session 2	Exploratory Design Sessions: Consists of two 30-minute sessions. During each session, each group receives a scenario probe centered on a collaborative challenge re-imagined for human-agent collaboration. Storyboard a critical moment and focus on the interaction designs and research considerations.
4:30 pm-5:00 pm	Group Presentations, Synthesis, and Closing

perspective focusing on LLM-based social simulacra and the role of HCI in social science research; and Marshall Van Alstyne from Boston University, who could provide a unique perspective from economics and information systems on the practical implications of human-agent collaboration.

3.3.3 Panel Discussion (20 min). Following the keynote, we will host a 20-minute interactive panel to facilitate direct conversation between multi-disciplinary experts and attendees. The panel consists of four panelists and one moderator. We will invite scholars from diverse research backgrounds, such as social science, AI, and design. The moderator will prompt the discussion with a single guiding question, for example: *"What is the significant barrier to designing LLM agents that collaborate with humans in a mutual partnership?"* The floor will then immediately open to the audience to ask questions of specific panelists. This interactive format allows participants to identify shared challenges and contrast diverse perspectives early in the workshop agenda.

3.3.4 Featured Lightning Talks (35 min). After the panel discussion, we will feature a series of lightning talks. This session combines three to four **invited talks from leading researchers** with three to four **authors of selected workshop submissions**. Each presentation will last 4 minutes, followed by a joint 5-minute Q&A session at the end of the block.

Invited talks will offer a comprehensive overview of recent research progress on topics including LLM agent design, human-agent collaboration, and downstream applications. These talks are conceptual provocations designed to seed the workshop with conflicting ideas, setting an energetic and intellectually stimulating tone. Confirmed presenters include Bingsheng Yao (Northeastern University) and Chaoran Chen (University of Notre Dame).

Integrating selected submissions into this session allows the broader community to showcase a variety of novel designs and empirical findings alongside established experts. The program committee will select the spotlights with a focus on diverse applications

and innovative methodologies that extend the boundaries of current research. All remaining accepted submissions will be presented in the poster session.

3.3.5 Poster Session (30 min). During the coffee break, we will have a poster session for participants to present their latest work and discuss their insights on human-agent collaboration with other participants. We will split the posters into two groups so that half of the participants can walk around while the other half is presenting.

3.3.6 Exploratory Design Session: Storyboarding Human-Agent Interactions (60 min). At the beginning of this activity, organizers will introduce the design sessions and form sub-groups. Participants will be divided into groups of 5-7 people each, based on their identified research expertise, to ensure diversity of expertise within each group. A moderator from the organizing committee will be assigned to join each group to facilitate discussion.

Two sessions of 30 minutes each will be conducted, with each group receiving a scenario probe centered on a classic collaborative challenge but re-imagined for a human-agent team in each session. The goal is to storyboard a critical interaction moment, focusing on the design of the interaction itself. Probes will encourage thinking about both foundations and frontiers. Probe examples include:

- (1) **Trust & Accountability:** "An agent working asynchronously makes a critical mistake. Storyboard the interface that allows the human to investigate the provenance of the error and provide corrective feedback that avoids similar errors and improves the agent's future behavior."
- (2) **Workspace Awareness in Multi-Human, Multi-Agent (MHMA) Teams:** "In a team with three humans and three agents, the agents are collaborating at high speed. Storyboard a novel awareness visualization that gives humans an intuitive, ambient sense of the agents' progress and confidence without causing information overload."

- (3) **Creative Partnership:** “A human designer is stuck. The agent partner offers a completely unexpected design direction. Storyboard how the agent presents this idea and the creative negotiation between the collaborators.”

3.3.7 Group Presentations, Synthesis, and Closing (30 min). A representative from each group will deliver a brief presentation after the design sessions to showcase their most insightful storyboard concept in a 3-minute pitch. This collective presentation will serve as the foundation for a broader discussion on establishing best practices in human-agent collaboration design and evaluation. The organizers will facilitate a concluding synthesis and outline the next steps for the post-workshop white paper and community building.

3.4 Post-Workshop Plans

We will produce a white paper and repository of workshop materials because, even if we have the maximum number of participants, this is a much broader conversation that is sorely needed. Our publishing strategy will ensure that the best practices and other results of the workshop are widely disseminated and accessible for ongoing engagement and collaboration. Our plan consists of the following components:

- (1) **An Agenda-Setting White Paper:** A white paper summarizing the workshop discussions will be drafted by the workshop organizers and participants post-workshop and submitted to a leading journal or conference, such as Communications of the ACM, Interactions, NeurIPS, or CHI. This document will synthesize the theoretical discussions, the design patterns from the sessions, and the key questions identified during the workshop. The drafting process will be collaborative; an outline will be shared with all participants, who will be invited to co-author sections based on their expertise. We will hold virtual follow-up meetings to ensure the paper represents the collective intelligence of the workshop and submit it to a high-impact venue.
- (2) **Building a Sustained Community** We see this workshop as the beginning of a sustained conversation. We will maintain the workshop website as a lasting online repository that includes all the resources used during the workshop, such as accepted papers, keynote slides (with the consent of the keynote speakers), research presentations, and the output of interactive sessions. This repository will be accessible to the public and referenced in the white paper to encourage further exploration and experimentation with human-agent collaboration. We will explore the possibility of forming a more formal Special Interest Group (SIG) and recurrently host the workshop at future HCI (e.g., CHI, UIST), NLP (e.g., ACL and NAACL), and ML (e.g., Neurips, AAAI) conferences, and use the same website to document these materials.
- (3) **Plans to Publish Proceedings** All accepted position papers will be compiled into a digital proceedings volume. This volume will be made publicly and permanently available on the workshop’s website. We will explore publishing the workshop proceedings in the ACM Digital Library. To ensure archival stability and broad, citable access for the authors, we will also publish the proceedings on an established open-access platform such as arXiv.org or CEUR-WS.org. This

guarantees that the intellectual contributions of our participants are formally recognized and can be built upon by the broader research community.

4 Accessibility

We are committed to providing an inclusive and accessible experience for all participants. To support this, we will work with the conference organizers to **ensure the physical workshop space is fully accessible** to individuals with mobility needs. We will guide all presenters to follow best practices for **content accessibility**, including using high-contrast color schemes in their slides and verbally describing key visual information. Furthermore, to allow participants to prepare and use their own assistive technologies, we will make **all digital materials**, such as accepted papers and presentation slides, **available on our workshop website**. Finally, we are prepared to **accommodate specific accessibility requests**. We will offer to arrange for services such as live captioning or sign language interpretation, and clear contact information for making these requests will be prominently displayed on our website.

5 Call for Participation – A Vision, Design Philosophy, and Empirical Framework for Achieving Successful Partnerships Between Humans and LLM Agents

Recent advances in Large Language Models (LLMs) open up possibilities for moving intelligent systems beyond the “tool” paradigm toward more equal partnerships with humans. These agents can engage in fluid natural language dialogue, exhibit role-playing behaviors, and demonstrate reasoning and planning capabilities. At the same time, they introduce risks such as misaligned autonomy, hallucinations, privacy and security vulnerabilities, and over-anthropomorphization.

This workshop calls on the HCI, CSCW, AI, and related communities to critically and creatively explore new paradigms of human-agent partnership. We invite contributions that draw on the rich history of technology-mediated human-human collaboration while also envisioning novel forms of mixed-initiative, multi-party, and asynchronous collaboration with AI agents. Topics of interest include, but are not limited to:

- Theoretical perspectives: applying and extending CSCW/HCI theories (e.g., common ground, workspace awareness, articulation work) to human-agent interaction
- Empirical studies of human-agent collaboration (e.g., trust calibration, coordination, failure recovery)
- Novel design paradigms for human-agent collaboration, including mixed-initiative workflows, coordination strategies, and asynchronous delegation
- Interface and interaction designs that make agents’ goals, actions, and limitations transparent to human partners
- Risks, challenges, and safeguards for collaboration (e.g., security, privacy, safety, misaligned autonomy)
- Critical perspectives on anthropomorphization, social accountability, and the ethics of agentic systems

Submission Format: We welcome 2–4 page submissions (ACM single-column format), including research articles, survey papers,

and position papers that present original ideas, empirical studies, or design visions.

Review Process: Submissions will undergo a lightweight, non-anonymous review, focusing on the quality of ideas and diversity of perspectives. Accepted papers will be published online on the workshop website.

Participation: The workshop will be a 180-minute, in-person session featuring a keynote, panel discussion, invited lightning and spotlight talks, and collaborative design activities via storyboarding. All accepted papers will be posted on the workshop website and feature a poster presentation. A number of selected submissions will be presented as community spotlight talks. More details of our workshop and paper submission can be found at <https://chi26workshop-human-agent-collaboration.hailab.io/>

Important Date:

CFP released date: December 18, 2025

Submission due date: February 15, 2026

Notification of acceptance date: February 28, 2026

Workshop date: April 13, 2026

References

- [1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collison, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
- [2] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. 2021. Extracting training data from large language models. In *30th USENIX security symposium (USENIX Security 21)*, 2633–2650.
- [3] Chaoran Chen, Leyang Li, Luke Cao, Yanfang Ye, Tianshi Li, Yaxing Yao, and Toby Jia-Jun Li. 2025. Why am I seeing this: Democratizing End User Auditing for Online Content Recommendations. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 154, 23 pages. doi:10.1145/3746059.3747798
- [4] Chaoran Chen, Weijun Li, Wenxin Song, Yanfang Ye, Yaxing Yao, and Toby Jia-Jun Li. 2024. An Empathy-Based Sandbox Approach to Bridge the Privacy Gap among Attitudes, Goals, Knowledge, and Behaviors. 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 234, 28 pages. doi:10.1145/3613904.3642363
- [5] Chaoran Chen, Bingsheng Yao, Rui Shi Zou, Wenyue Hua, Weimin Lyu, Toby Jia-Jun Li, and Dakuo Wang. 2025. Towards a Design Guideline for RPA Evaluation: A Survey of Large Language Model-Based Role-Playing Agents. In *Findings of the Association for Computational Linguistics: ACL 2025*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 18229–18268. doi:10.18653/v1/2025.findings-acl.938
- [6] Chaoran Chen, Zhiping Zhang, Bingcan Guo, Shang Ma, Ibrahim Khalilov, Simret A Gebregziabher, Yanfang Ye, Ziang Xiao, Yaxing Yao, Tianshi Li, et al. 2025. The Obvious Invisible Threat: LLM-Powered GUI Agents' Vulnerability to Fine-Print Injections. *arXiv preprint arXiv:2504.11281* (2025).
- [7] Chaoran Chen, Zhiping Zhang, Ibrahim Khalilov, Bingcan Guo, Simret A Gebregziabher, Yanfang Ye, Ziang Xiao, Yaxing Yao, Tianshi Li, and Toby Jia-Jun Li. 2025. Toward a human-centered evaluation framework for trustworthy llm-powered gui agents. *arXiv preprint arXiv:2504.17934* (2025).
- [8] Chaoran Chen, Daodao Zhou, Yanfang Ye, Toby Jia-Jun Li, and Yaxing Yao. 2025. CLEAR: Towards Contextual LLM-Empowered Privacy Policy Analysis and Risk Generation for Large Language Model Applications. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 277–297. doi:10.1145/3708359.3712156
- [9] Jiaju Chen, Yuxuan Lu, Xiaojie Wang, Huimin Zeng, Jing Huang, Jiri Gesi, Ying Xu, Bingsheng Yao, and Dakuo Wang. 2025. Multi-agent-as-judge: Aligning LLM-agent-based automated evaluation with multi-dimensional human evaluation. *arXiv preprint arXiv:2507.21028* (2025).
- [10] Herbert H Clark and Susan E Brennan. 1991. Grounding in communication. (1991).
- [11] Catherine Durnell Cramton. 2001. The mutual knowledge problem and its consequences for dispersed collaboration. *Organization science* 12, 3 (2001), 346–371.
- [12] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.
- [13] Paul Dourish and Sara Bly. 1992. Portholes: Supporting awareness in a distributed work group. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 541–547.
- [14] Nicholas Epley, Adam Waytz, and John T Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review* 114, 4 (2007), 864.
- [15] Simret Araya Gebregziabher, Yukun Yang, Charles Chiang, Hojun Yoo, Chaoran Chen, Hyo Jin Do, Zahra Ashktorab, Werner Geyer, Diego Gómez-Zará, and Toby Jia-Jun Li. 2026. The Behavioral Fabric of LLM-Powered GUI Agents: Human Values and Interaction Outcomes. In *Proceedings of the 31st International Conference on Intelligent User Interfaces (IUI '26)*.
- [16] Jonathan Grudin. 1988. 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*. 85–93.
- [17] Jonathan Grudin. 2018. From tool to partner: The evolution of human-computer interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–3.
- [18] Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW) 11*, 3 (2002), 411–446.
- [19] Steve Harrison and Paul Dourish. 1996. Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*. 67–76.
- [20] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 119–125.
- [21] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.
- [22] Zwei Ji, Tiezheng Yu, Yan Xu, Nayeon Lee, Etsuko Ishii, and Pascale Fung. 2023. Towards mitigating LLM hallucination via self reflection. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. 1827–1843.
- [23] Joshua Alexander Kroll. 2015. *Accountable algorithms*. Ph.D. Dissertation. Princeton University.
- [24] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.
- [25] Joanne Leong, John Tang, Edward Cutrell, Sasa Junuzovic, Gregory Paul Baribault, and Kori Inkpen. 2024. Dittos: Personalized, embodied agents that participate in meetings when you are unavailable. *Proceedings of the ACM on Human-Computer Interaction 8, CSCW2* (2024), 1–28.
- [26] Yuxuan Lu, Jing Huang, Yan Han, Bennet Bei, Yaochen Xie, Dakuo Wang, Jessie Wang, and Qi He. 2025. LLM Agents That Act Like Us: Accurate Human Behavior Simulation with Real-World Data. *arXiv e-prints* (2025), arXiv–2503.
- [27] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human-computer interaction* 15, 2–3 (2000), 139–178.
- [28] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*. 1–22.
- [29] Byron Reeves and Clifford Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK* 10, 10 (1996), 19–36.
- [30] Kjeld Schmidt and Liam Bannon. 1992. Taking CSCW seriously: Supporting articulation work. *Computer supported cooperative work (CSCW) 1*, 1 (1992), 7–40.
- [31] Daniel D. Slate, Chaoran Chen, Yaxing Yao, and Toby Jia-Jun Li. 2025. Iterative Contextual Consent: AI-enabled Data Privacy Contracts. In *Proceedings of the 2025 Workshop on Human-Centered AI Privacy and Security (HAIPS '25)*. Association for Computing Machinery, New York, NY, USA, 84–91. doi:10.1145/3733816.3760757
- [32] Jingyu Tang, Chaoran Chen, Jiawen Li, Zhiping Zhang, Bingcan Guo, Ibrahim Khalilov, Simret Araya Gebregziabher, Bingsheng Yao, Dakuo Wang, Yanfang Ye, et al. 2025. Dark patterns meet gui agents: Llm agent susceptibility to manipulative interfaces and the role of human oversight. *arXiv preprint arXiv:2509.10723* (2025).
- [33] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, et al. 2024. Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *First Conference on Language Modeling*.
- [34] Bingsheng Yao, Jiaju Chen, Chaoran Chen, April Wang, Toby Jia-jun Li, and Dakuo Wang. 2025. Through the Lens of Human-Human Collaboration: A Configurable Research Platform for Exploring Human-Agent Collaboration. *arXiv preprint arXiv:2509.18008* (2025).
- [35] Bingsheng Yao, Bo Sun, Yuanzhe Dong, Yuxuan Lu, and Dakuo Wang. 2025. DPRF: A Generalizable Dynamic Persona Refinement Framework for Optimizing Behavior Alignment Between Personalized LLM Role-Playing Agents and Humans. *arXiv preprint arXiv:2510.14205* (2025).

- [36] Jiayi Ye, Chaoran Chen, Yue Huang, Yanfang Ye, Toby Jia-Jun Li, and Xiangliang Zhang. 2025. My Favorite Streamer is an LLM: Discovering, Bonding, and Co-Creating in AI VTuber Fandom. *arXiv preprint arXiv:2509.10427* (2025).