AI-Enabled Fraudulent Participation in Human-Subjects Research

WEIJIA HE, University of Southampton, UK
EUNJEONG CHEON, Syracuse University, USA
CHIA-FANG CHUNG, University of California, Santa Cruz, USA
DIPTO DAS, University of Toronto, Canada
HARMANPREET KAUR, University of Minnesota, USA
TOBY JIA-JUN LI, University of Notre Dame, USA
JESSICA VITAK, University of Maryland, College Park, USA
YAXING YAO, Johns Hopkins University, USA

Fraudulent use of AI is increasingly prevalent in human-subjects research, rendering many traditional bot-detection methods ineffective and, consequently, raising significant concerns about data integrity among researchers who work with human-subjects data. Tools are being developed to address concerns about fraudulent participation; however, it remains unclear which best practices to follow for prevention or mitigation, how effective the adopted preventative measures are, and to what extent one can trust the collected data. In this workshop, we will bring together human-computer interaction and social computing researchers to exchange strategies for addressing fraudulent participation, assess the effectiveness of existing detection methods, discuss the potential risks and opportunities of AI in human-centered research, and develop method-specific guidelines for fraud prevention/detection and the research dissemination/review process. The workshop involves full-group activities, breakout discussions divided by research methods, and the production of a draft document summarizing method-specific guidelines that will be shared with the wider CSCW community for feedback. We hope that the conversation started during this workshop will spur deeper engagement from CSCW and the wider SIGCHI research community to implement approaches to minimize the negative impact of AI-enabled fraudulent participation in research.

CCS Concepts: • **Social and professional topics** → *Computing profession.*

Additional Key Words and Phrases: Artificial Intelligence, Fraud, Bots, Human-Subjects Research, Human-Centered Research Methods

ACM Reference Format:
Weijia He, EunJeong Cheon, Chia-Fang Chung, Dipto Das, Harmanpreet Kaur, Toby Jia-Jun Li, Jessica Vitak, and Yaxing Yao. 2026. AI-Enabled Fraudulent Participation in Human-Subjects Research. In *Companion of the Computer-Supported Cooperative Work and Social Computing (CSCW Companion '26), October 10–14, 2026, Salt Lake City, UT, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3785651.3816095

Authors' Contact Information: Weijia He, University of Southampton, UK, Weijia.He@soton.ac.uk; EunJeong Cheon, Syracuse University, USA, echeon@syr.edu; Chia-Fang Chung, University of California, Santa Cruz, USA, cfchung@ucsc.edu; Dipto Das, University of Toronto, Canada, dipto.das@utoronto.ca; Harmanpreet Kaur, University of Minnesota, USA, harmank@umn.edu; Toby Jia-Jun Li, University of Notre Dame, USA, toby.j.li@nd.edu; Jessica Vitak, University of Maryland, College Park, USA, jvitak@umd.edu; Yaxing Yao, Johns Hopkins University, USA, yaxing@jhu.edu.

Manuscript submitted to ACM

1 Introduction

The rise of generative AI poses new challenges to human-subjects research. Today's automated AI agents can successfully pass attention checks while maintaining a coherent demographic persona, with consistent 'memories' and persona-calibrated sophistication [2, 13]. Recent research has demonstrated the prevalence of AI use and fraudulent participation on online survey platforms [4, 12], while an analysis of human-written responses and LLM-aided ones also suggests that the latter is more homogeneous than the former [12]. Likewise, AI-generated qualitative responses (e.g., synthetic data and simulated users) is gaining increasing attention across various fields of user research [6, 10, 11, 14].

This topic is of keen interest to the HCI and social computing research community. A 2025 CSCW article identified AI and LLMs as a major ethical concern for HCI research [5], while a 2024 CHI article highlighted increases in fraudulent participants in qualitative HCI studies [9]. Over time, we expect AI to create signifiant challenges in identifying authentic research data in most forms of human-subjects research, raising concerns about data integrity and reliability. If the usual checks for non-human participants (e.g., attention checks, CAPTCHA) all fail in the face of generative AI, then how can we trust any data collected through online methods?

In response to the crises arising from the use of AI in human-subjects research, AI-preventative measures have been proposed, such as LLM-targeted command injections and cognitive traps [7], as well as keystroke tracking to detect generative AI use in study tasks [1]. Platforms like Prolific have also developed their own AI bot detection, but they have also suggested that researchers take additional measures to prevent the collection of fraudulent responses [3]. How reliable these detection methods are in the wild and in response to the evolving AI technology remains uncertain. Importantly, this creates an adversarial ecosystem for human-subjects research where researchers cannot confidently rely on the collected data while participants do not see the value of providing good data. Taken to its extreme, this trend can undermine the scientific rigor of our field. Therefore, it is critical that researchers quickly identify ways to prevent or mitigate fraudulent data and ensure data quality remains high.

2 Workshop Overview

In this workshop, we invite researchers who work with human-subjects data to discuss the use of AI in online data collection and to identify specific steps the HCI/social computing community should take moving forward to ensure the value of human-subjects research while coexisting with the prevalence of AI.

For this workshop, we use the term *"fraudulent participation"* broadly to refer to any instance where the data collected in a human-subjects study does not authentically represent the experiences, perspectives, or behaviors of the intended participant population. This includes fully automated AI agents posing as human participants; human participants who delegate their responses to AI tools (e.g., using ChatGPT to generate survey or interview answers); and other forms of misrepresentation such as identity fabrication or repeated participation under different identities. While AI-enabled non-human participation has received the most attention recently, we intentionally adopt this broader framing to capture the full spectrum of threats to data integrity in human-subject research.

We are especially interested in addressing the following topics during the workshop:

- Strategies to detect or prevent AI-enabled non-human participants.
- Enhancement of trust and confidence in human-subjects data quality and integrity in the age of AI.
- Effectiveness of existing AI-enabled fraudulent participation detection methods.
- Contexts where human-simulated AI can be beneficial.

- Methodological vulnerabilities across research methods; e.g., how do the risks of fraudulent participation differ across surveys, remote interviews, and in-person/ethnographic studies?
- Cross-cultural and multilingual dimensions of fraudulent participation detection; e.g., how do language, cultural norms, and platform ecosystems outside English-speaking contexts complicate both fraud and its detection?
- The role of AI mediation in blurring the boundaries of "genuine" human participation; e.g., participants using AI tools to assist their responses vs. fully automated non-human agents.
- Ethical and epistemological implications; e.g., what constitutes authentic human data when AI increasingly mediates everyday communication and decision-making?
- Developing method-specific guidelines for fraud prevention and detection tailored to different study designs.
- Developing community guidelines for paper submission and review processes to help authors and reviews assess data quality.

The workshop aims to provide a space for researchers to exchange ideas on ensuring genuine human responses, foster conversations on the risks and benefits of AI in human-subject research, and inspire ideas for potential new standards for human-subject research. We plan to use this workshop to quickly generate a draft document providing guidance to HCI/social computing researchers and then share with conference attendees for feedback.

3 Workshop Logistics

3.1 Workshop Goals

The primary goal of this one-day workshop is to produce a draft guide for HCI and social computing researchers that (1) identifies the most common forms of fraudulent participation in different types of research and (2) provides clear guidance on how to design studies to minimize the likelihood of fraudulent participation and identify such behaviors at the time of occurrence. Our secondary goal is to engage as many CSCW (and SIGCHI) members as possible in conversations about this topic to ensure diverse perspectives and experiences are reflected in the guide and to maximize the utility of such a document for human-subjects researchers.

3.2 Workshop Activities

Because the workshop is focused on producing a tangible deliverable, activities will largely be focused on brainstorming and drafting documentation. This will include full-group activities, breakout discussions, report-outs to the group, and production of a draft document to be finalized and shared with the larger HCI/social computing community for feedback.

In structuring these activities, we recognize that the risks and mitigation strategies for fraudulent participation vary significantly across research methods. To address this, we propose organizing breakout groups around major methodological categories, such as: (1) online surveys and questionnaires, (2) remote interviews and focus groups, (3) diary studies and longitudinal data collection, and (4) in-person and ethnographic studies. Each group will identify the most common forms of fraudulent participation within their method, assess the effectiveness of existing detection and prevention strategies, and develop targeted recommendations. This structure acknowledges that while online surveys face the most acute threat from AI-generated responses, other methods—including remote interviews—are increasingly vulnerable as AI capabilities expand. At the same time, methods involving sustained in-person engagement (e.g., ethnographic fieldwork) offer a useful contrast, revealing which aspects of human participation are most resistant to automation and why.

In parallel with the creation of the document, we will also co-create a poster to share the core discussion points from this workshop. With the permission of the poster chairs, we will ask workshop participants to volunteer to attend the poster session and gather feedback from attendees, both by adding comments and suggestions and by voting on proposed solutions (e.g., using stickers). Likewise, we will share our findings with other conference attendees through social media channels (Bluesky, Discord) to provide opportunities for the wider CSCW/HCI community to contribute to the development of the final resource.

3.3 Workshop Schedule

The full-day workshop schedule will be structured as follows:

Time	Activity
9:30 am	Welcome, overview, and introductions
10:00 am	Full-group brainstorming of topic areas (e.g., organize by method or by category of problem)
10:45 am	Ideating and outlining in breakout groups
12:00 pm	Progress reports from each group
12:30 pm	Lunch break
2:00 pm	Check-in to confirm group goals
2:15 pm	Breakout groups begin drafting their sections
4:00 pm	Full group discussion of document
4:45 pm	Discussion of next steps (finalizing document, community feedback, dissemination plan)
5:00 pm	Workshop wraps

3.4 Recruitment

Interested participants should visit https://fraudulent-participation-workshop.github.io, which provides full details on how to apply. We will recruit participants through social media channels (e.g., CSCW Meta Facebook group and Bluesky) as well as through our personal networks. We expect this topic to be of interest to many CSCW participants given how common it is becoming for researchers to experience fraudulent participants in their own research.

Prospective participants will be asked to submit a short statement of interest via Google Forms regarding their participation in the workshop, including their specific areas of interest within and beyond the points listed above, as well as identifying their methodological background and any experiences they may have had with fraudulent participants in their own research. These statements will be reviewed by the organizing team to ensure applicants' interests align with the workshop. We will also use this information to better structure working groups to align with participants' backgrounds and interests.

Participants will be expected to attend the workshop in person.

3.5 Required Equipment

For this workshop, we will need a room that can be set up for 5-6 breakout groups with tables and chairs to seat 30 people. We will also need large poster sheets for brainstorming. We will provide attendees with additional writing materials. The meeting room should be equipped with a projector and a projection screen, along with the necessary HDMI or Type-C cables.

4 Organizers

The organizing team represents a diverse set of early and mid-career academics at eight different institutions in the U.S., Canada, and UK. The organizing team has significant experience in both organizing workshops at SIGCHI conferences and in conducting a wide range of human-subjects research. Some organizers have published papers on the topic and most have experienced fraudulent participants in their own research.

- **EunJeong Cheon** is an Assistant Professor at Syracuse University's School of Information Studies. Her research examines how emerging technologies reshape labor and social relations through ethnographic and critical design methods across international contexts, including fieldwork in South Korea, Denmark, and the U.S. She has organized 10+ workshops at SIGCHI conferences and has extensive experience designing and conducting human-subjects research across diverse methods and cultural settings.
- **Christina Chung** is an associate professor in the Department of Computational Media at the University of California, Santa Cruz. She studies personal informatics in various health contexts. She has co-led a CHI 2024 paper [9] and workshop [8] on fraudulent participation in qualitative research (with limited attention to AI-enabled participation).
- **Dipto Das** is a postdoctoral fellow in the Department of Computer Science at the University of Toronto. His mixed-methods, interdisciplinary, community-engaged research approach examines how AI systems and digital platforms shape social relations, discourse, and inequality across various domains and diverse sociopolitical and cultural contexts (e.g., South Asia, North America, and the Middle East). He has previously organized workshops at CSCW and CHI around indigeneity, ethics, migration, and HCI across borders.
- **Weijia He** is an Assistant Professor in the School of Electronics and Computer Science at the University of Southampton. Her research focuses on human-centered security and privacy in emerging technologies (e.g., the Internet of Things, human-AI interactions). She has extensive experience in human-subjects research using a wide range of methods.
- **Harmanpreet Kaur** is an Assistant Professor in the Department of Computer Science and Engineering at the University of Minnesota, where she co-leads the GroupLens research lab. She studies appropriate reliance on AI in knowledge work (e.g., data science, search) and everyday decision-making settings. Her work often relies on crowdsourced data collection and she evaluates the measurement impact of data sources. She has organized workshops and panels on topics related to human-centered AI and crowdsourcing at several conferences including CHI, CSCW, IUI, and HCOMP.
- **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 works at the intersection of HCI, ML, and NLP applications, where he uses human-centered methods to design, build, and study interactive systems to empower individuals to create, configure, and extend AI-powered computing systems. He has co-organized SIGs and workshops on relevant topics and published an article on the challenges of LLMs in HCI research [5].
- **Jessica Vitak** is a professor in the College of Information at the University of Maryland and a mixed-methods researcher with expertise in surveys, interviews, and focus groups. She has organized 10+ workshops at CSCW and CHI on issues of privacy and ethics, and she has run panels at CSCW on similar other topics important to the community (e.g., reviewing crisis).
- **Yaxing Yao** is an assistant professor in the Department of Computer Science at Johns Hopkins University. His research focuses on human-centered privacy and security, with an emphasis on different populations, such as

minors, older adults, and people with disabilities. He has organized 10+ workshops at various HCI and security and privacy venues.

References

[1] Michael W Asher, Gillian Gold, Eason Chen, and Paulo F Carvalho. [n. d.]. Chatbots Are Undermining Crowdsourced Research in the Behavioral Sciences: Detecting AI-Assisted Cheating with a Keystroke-Based Tool. ([n. d.]).

[2] Chaoran Chen, Bingsheng Yao, Ruishi 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

[3] Andrew Gordon. 2026. Authenticity Checks: How We Tested the Most Accurate Method for Identifying Agentic AI. https://www.prolific.com/resources/authenticity-checks-how-we-tested-the-most-accurate-method-for-identifying-agentic-ai.

[4] Elizabeth Hitches, Dean Dudley, Melissa Johnstone, and Stuart Woodcock. 2025. Bots and Baddies: Supporting the Integrity of Online Survey Research in the Face of a Growing Challenge. *Quality & Quantity* 59, 2 (April 2025), 1481–1506. doi:10.1007/s11135-024-02001-w

[5] Shivani Kapania, Ruiyi Wang, Toby Jia-Jun Li, Tianshi Li, and Hong Shen. 2025. 'I'm Categorizing LLM as a Productivity Tool': Examining Ethics of LLM Use in HCI Research Practices. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW102 (May 2025), 26 pages. doi:10.1145/3711000

[6] Jie Li. 2024. How Far Can We Go with Synthetic User Experience Research? *Interactions* 31, 3 (May 2024), 26–29. doi:10.1145/3653682

[7] Felipe M. Affonso. 2025. Detecting Vision-Enabled AI Respondents in Behavioral Research Through Cognitive Traps. social science research network:5766848 doi:10.2139/ssrn.5766848

[8] Aswati Panicker, Novia Nurain, Zaidat Ibrahim, Chun-Han (Ariel) Wang, Seung Wan Ha, Elizabeth Kaziunas, Maria K Wolters, and Chia-Fang Chung. 2024. Forms of Fraudulence in Human-Centered Design: Collective Strategies and Future Agenda for Qualitative HCI Research. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 469, 5 pages. doi:10.1145/3613905.3636309

[9] Aswati Panicker, Novia Nurain, Zaidat Ibrahim, Chun-Han (Ariel) Wang, Seung Wan Ha, Yuxing Wu, Kay Connelly, Katie A. Siek, and Chia-Fang Chung. 2024. Understanding fraudulence in online qualitative studies: From the researcher's perspective. 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 824, 17 pages. doi:10.1145/3613904.3642732

[10] M Rosala, K Moran, and Nielsen Norman Group. 2024. Synthetic users: If, when, and how to use AI-generated "research.". *Nielsen Norman Group* (2024).

[11] Joni Salminen, Danial Amin, Soon-Gyo Jung, and Bernard Jansen. 2025. The Use of Large Language Models in HCI: A Critical Analysis of Synthetic Users. In *Proceedings of the Augmented Humans International Conference 2025 (AHs '25)*. Association for Computing Machinery, New York, NY, USA, 413–417. doi:10.1145/3745900.3746108

[12] Veniamin Veselovsky, Manoel Horta Ribeiro, Philip J. Cozzolino, Andrew Gordon, David Rothschild, and Robert West. 2025. Prevalence and Prevention of Large Language Model Use in Crowd Work. *Commun. ACM* 68, 3 (Feb. 2025), 42–47. doi:10.1145/3685527

[13] Sean J. Westwood. 2025. The Potential Existential Threat of Large Language Models to Online Survey Research. *Proceedings of the National Academy of Sciences* 122, 47 (Nov. 2025), e2518075122. doi:10.1073/pnas.2518075122

[14] Bingsheng Yao, Jiaju Chen, Chaoran Chen, April Wang, Toby Jia jun Li, and Dakuo Wang. 2026. Through the Lens of Human-Human Collaboration: A Configurable Research Platform for Exploring Human-Agent Collaboration. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*.