Towards Responsible Use of Large Multi-Modal AI to Analyze Human Social Behaviors

Qingxiao Zheng
University of Illinois
Urbana-Champaign, USA
qzheng14@illinois.edu

Xi Lu
University at Buffalo, SUNY
Buffalo, USA
xlu30@buffalo.edu

Qiao Jin
University of Minnesota
Minneapolis, USA
jin00122@umn.edu

Jitesh Jain
Georgia Institute of Technology
Atlanta, USA
jjain45@gatech.edu

Hedda Meadan-Kaplansky
University of Illinois at
Urbana-Champaign
Champaign, USA
meadan@illinois.edu

Humphrey Shi
Georgia Institute of Technology
Atlanta, USA
shi@gatech.edu

Jinjun Xiong
University at Buffalo, SUNY
Buffalo, USA
jinjun@buffalo.edu

Yun Huang
University of Illinois at
Urbana-Champaign
Champaign, USA
yunhuang@illinois.edu

Abstract

This SIG is primarily designed for researchers and practitioners interested in exploring the utilization of multimodal artificial intelligence (AI) to understand and analyze human behaviors across various domains. It will particularly benefit professionals in UX design and research, cognitive science, education, psychology, and data science who are pursuing innovative ways to integrate multimodal data with AI to gain deeper behavioral insights. The session will focus on using AI to track, visualize, and generate insights around both verbal and non-verbal social interactions for various purposes, such as video question-answering. Discussions will cover practical applications in settings such as online learning environments, telecommunication, and social media platforms. This SIG will also provoke reflection on the ethical stewardship of AI, highlighting the importance of responsible usage to mitigate concerns such as algorithmic bias and data privacy.

CCS Concepts

• **Human-centered computing → Systems and tools for interaction design**; *Collaborative and social computing systems and tools*; **Interactive systems and tools**; **Collaborative and social computing**; • **Information systems → Multimedia information systems**; • **Social and professional topics**;

Keywords

AI, Large multimodal models (LMMs), Human social behaviors, Generative AI, Video analysis

1 Multi-modal AI for Analyzing Human Social Behaviors

Human connections thrive on complex social interactions that incorporate verbal cues like spoken language, and non-verbal cues, including but not limited to body language, gestures, facial expressions, and eye contact. The diversity and dynamism of these interactions demand a nuanced understanding for accurate interpretation and analysis.

Many works utilize behavioral insights to improve social activities in various settings [21], such as facilitating interpersonal communication in online meetings [2], understanding user-user connections on social media [13, 22], and supporting peer-peer or student-teacher interaction experiences in online learning environments [3, 9, 17]. Additionally, healthcare systems utilize behavioral analysis to accurately interpret patient-provider interaction in therapy sessions [7, 19]. Similarly, personalized marketing strategies are designed to align closely with brand-consumer interaction behavior, leveraging insights from behavioral studies [15]. Understanding human behaviors allows for more effective decision-making across these various fields [14].

Generative AI models with audio-visual capabilities, specifically Large Multi-Modal Models (LMMs) [8, 12] such as GPT-4o, LLaVA, and Gemini present a promising solution to analyze human behaviors [5, 10]. Their ability to summarize, analyze, and interpret complex verbal and non-verbal behaviors has been leveraged in many fields of application [4, 18]. By processing text, video, and audio data, recent applications are investigating ways to leverage

multimodal models for tasks such as highlighting behavior changes, tracking behavioral cues, and generating detailed and objective data [6, 20].

Despite progress in AI models that have enhanced capabilities in analyzing human behaviors, as indicated by recent studies [1, 4], several critical challenges remain. These include the complexities involved in decoding the wide spectrum of social cues, understanding contextual subtleties, addressing ambiguities, and recognizing individual differences. Ethical concerns such as data privacy, algorithmic bias, and the potential for misuse also require urgent attention. The potential implications of harnessing the power of LMMs to improve our understanding of interpersonal communication interaction patterns in this field are vast, ranging from screening and assessment to intervention, behavior modification, and even predictive analytics for preventative measures.

2 Our Goal: Advancing Responsible Use of Large Multi-Modal Models in Behavioral Research

Ethical considerations are critical when integrating LMMs into behavioral research. These powerful AI models have the potential to significantly influence our understanding of human behavior, but they also pose risks such as invasions of privacy, amplification of biases, and lack of accountability [1, 5]. Researchers must engage thoughtfully with these issues to harness the benefits of LMMs while preventing potential harm. Ethical oversight is essential to ensure that the deployment of LMMs respects individual rights and adheres to high standards of fairness and transparency. Addressing these ethical challenges proactively helps build users' trust in AI technology, safeguards vulnerable populations, and ensures that the advancements in behavioral research by LMMs contribute positively to society.

Many past SIGs at CSCW and other conferences have discussed AI-related themes, such as AI-powered Social Computing Systems and Purposeful AI at CSCW'23 [11, 16]. As more advanced multimodal models emerge, this SIG offers a timely platform for social computing researchers to explore the impact of LMMs. The focus will be on understanding how LMMs can be effectively used in human behavioral research. This SIG aims to dive into current practices, perspectives, and challenges faced by researchers in integrating multi-modal generative AI into social computing. Researchers will brainstorm responsible methodologies to enhance understanding and improve user interaction design. We aim to facilitate in-depth discussions within the CSCW community about the challenges and opportunities of LMMs, engaging diverse perspectives to enhance understanding of generative AI's societal and technological effects. Our discussions will build a dedicated community and develop research agendas to be publicly shared, driving future innovations in this vital area.

3 Intended Community

Attendees of this SIG are expected to be interested in using generative AI (e.g., LMMs) to conduct behavioral research. Ideally, participants will come from a variety of disciplines such as psychology, education, behavioral science, computer science, data science, and human-computer interaction. Deep technical expertise in AI

is not a prerequisite. A basic grasp of how AI technologies can be applied to analyze human behavior will allow attendees to fully engage with the discussions and contribute meaningfully to the dialogue.

4 Research Agenda

The SIG will be an opportunity to discuss a variety of issues with respect to the SIG's topic. Discussion topics range from ethics and risks in using LMMs for human behavior analysis, challenges in analyzing human social interactions, to applications of LMMs and their potential long-term impacts. Included are several key questions, although these are not exhaustive:

- How effectively can LMMs detect and analyze the full spectrum of verbal and non-verbal cues in diverse social interactions? What are the limitations of their current capabilities?
- What are the primary technical and methodological challenges in integrating LMMs with traditional behavioral research methods? How can these challenges be addressed to enhance the robustness of research outcomes?
- How can researchers ensure the accuracy and reliability of the insights derived from LMMs? What validation techniques or assessments are necessary?
- How can these data be used to improve domain practices?
- What ethical concerns and considerations arise from using LMMs in behavioral research (e.g., consent, privacy, and bias)?
- What measures can be implemented to prevent the potential misuse of these technologies?
- Who should be involved to ensure transparency, accountability, and ethical use of LMMs in behavioral studies?
- What frameworks or guidelines should be established?

5 Expected Outcome

During the SIG itself and in related discussions, we will foster the formation of new research collaborations and larger initiatives. Participants will be particularly encouraged to interact with colleagues from different disciplines with whom they have not previously had the chance to engage. The SIG will act as a pivotal gathering for researchers, educators, and practitioners from diverse fields, and the environment will promote rich, cross-disciplinary exchanges that are essential for advancing the use of LMMs in social computing research.

Post-SIG, a LinkedIn group will be established to facilitate continued dialogue and coordination on issues related to the using LMMs. This group will enable attendees to sustain interactions and further the discussions initiated during the SIG. Key guidelines, insights, and findings from the SIG will be documented and targeted for publication in prominent journals such as Interactions or Communications of the ACM.

6 Acknowledgement

findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Institute of Education Sciences, or the U.S. Department of Education.

References

[1] Riku Arakawa, Kiyosu Maeda, and Hiromu Yakura. 2024. Supporting Experts with a Multimodal Machine-Learning-Based Tool for Human Behavior Analysis of Conversational Videos. *arXiv preprint arXiv:2402.11145* (2024).

[2] Chenhao Chen, Yutaka Arakawa, Ko Watanabe, and Shoya Ishimaru. 2022. Quantitative Evaluation System for Online Meetings Based on Multimodal Microbehavior Analysis. *Sensors & Materials* 34 (2022).

[3] Xinyue Chen, Shuo Li, Shipeng Liu, Robin Fowler, and Xu Wang. 2023. MeetScript: Designing Transcript-based Interactions to Support Active Participation in Group Video Meetings. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–32.

[4] Bruno Carlos Dos Santos Melicio, Linyun Xiang, Emily Dillon, Latha Soorya, Mohamed Chetouani, Andras Sarkany, Peter Kun, Kristian Fenech, and Andras Lorincz. 2023. Composite AI for Behavior Analysis in Social Interactions. In *Companion Publication of the 25th International Conference on Multimodal Interaction*. 389–397.

[5] Ankita Gandhi, Kinjal Adhvaryu, Soujanya Poria, Erik Cambria, and Amir Hussain. 2023. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion* 91 (2023), 424–444.

[6] Lin Guan, Yifan Zhou, Denis Liu, Yantian Zha, Heni Ben Amor, and Subbarao Kambhampati. 2024. " Task Success" is not Enough: Investigating the Use of Video-Language Models as Behavior Critics for Catching Undesirable Agent Behaviors. *arXiv preprint arXiv:2402.04210* (2024).

[7] Matthew Jackson, Anne H Anderson, Rachel McEwan, and Jim Mullin. 2000. Impact of video frame rate on communicative behaviour in two and four party groups. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. 11–20.

[8] Jitesh Jain, Jianwei Yang, and Humphrey Shi. 2024. VCoder: Versatile Vision Encoders for Multimodal Large Language Models. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[9] Qiao Jin, Ishan Sodhi, Angie Chen, and Svetlana Yarosh. 2024. Interaction Forms of Collaborative VR Video Learning: An Exploratory Study. (2024).

[10] Chunyuan Li, Zhe Gan, Zhengyuan Yang, Jianwei Yang, Linjie Li, Lijuan Wang, Jianfeng Gao, et al. 2024. Multimodal foundation models: From specialists to general-purpose assistants. *Foundations and Trends® in Computer Graphics and Vision* 16, 1-2 (2024), 1–214.

[11] Tianyi Li and Francisco Iacobelli. 2023. Purposeful AI. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 563–565. https://doi.org/10.1145/3584931.3606954

[12] Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024. Improved Baselines with Visual Instruction Tuning. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[13] Xi Lu, Yunan Chen, and Daniel A. Epstein. 2021. A Model of Socially Sustained Self-Tracking for Food and Diet. *Proceedings of the ACM on Human-Computer Interaction* 5 (10 2021), 451. Issue CSCW2. https://doi.org/10.1145/3479595

[14] Hedda Meadan, James D Lee, Michelle M Sands, Moon Y Chung, and Pau García-Grau. 2023. The Coaching Fidelity Scale (CFS): Development and Evaluation of an Observational Measure of Coaching Fidelity. *Infants & Young Children* 36, 1 (2023), 37–52.

[15] Arshi Naim. 2023. Consumer behavior in marketing patterns, types, segmentation. *European Journal of Economics, Finance and Business Development* 1, 1 (2023), 1–18.

[16] Hong Shen, Tianshi Li, Toby Jia-Jun Li, Joon Sung Park, and Diyi Yang. 2023. Shaping the Emerging Norms of Using Large Language Models in Social Computing Research. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 569–571. https://doi.org/10.1145/3584931.3606955

[17] Ignacio Villagrán. 2021. Exploring the Effects of Applying Learning Analytics for Teaching Procedural Skills in Health Sciences Education. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 299–302.

[18] Ridwan Whitehead, Andy Nguyen, and Sanna Järvelä. 2024. The Generative Multimodal Analysis (GMA) Methodology for Studying Socially Shared Regulation in Collaborative Learning. In *The International Conference on Learning Analytics & Knowledge (LAK24)*.

[19] Christy D Yoon, Adriana Kaori Terol, Hedda Meadan, and James D Lee. 2024. Gaze Behaviors and Social Communication Skills of Young Autistic Children: A Scoping Review. *Review Journal of Autism and Developmental Disorders* (2024), 1–15.

[20] Qingxiao Zheng, Parisa Rabbani, Yu-Rou Lin, Daan Mansour, and Yun Huang. 2024. SOAP.AI: A Collaborative Tool for Documenting Human Behavior in Videos through Multimodal Generative AI. In *Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24)* (San Jose, Costa Rica, 2024-11-09/2024-11-13). ACM, New York, NY, USA, 4. https://doi.org/10.1145/3678884.3681819

[21] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 570, 24 pages. https://doi.org/10.1145/3491102.3501855

[22] Qingxiao Zheng, Shengyang Xu, Lingqing Wang, Yiliu Tang, Rohan C Salvi, Guo Freeman, and Yun Huang. 2023. Understanding safety risks and safety design in social VR environments. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–37.