maxiao.info 🌐 xm75@cornell.edu ✉

Xiao Ma

Summary

I am passionate about building AI systems that improve the human experience in a meaningful way. I work on Bard/Gemini safety eval and improvement. I have also invented new human-AI interactions for exploratory tasks and video content reviews.

Professional Experience

2019-present **Google**

I am a senior software engineer on the Responsible ML team in Google Research. I land impactful projects making AI systems safer and more aligned with human values, by combining foundational research in the intersection of machine learning and human-centered computing, and collaboration with product teams.

- **Bard/Gemini safety:** Critical safety launch human evals (3+ launches) and mitigations through safe fine-tuning data generation (5+ launches).
- **Training AI to assist raters:** Developed and launched a first-of-its-kind ML-generated hints system to better direct rater attention for video content reviews (3+ launches and a best paper award).
- **Fairness:** Contributed a new loss to the open-source tensorflow library for model remediation (2 launches).

2018, 2017 **Facebook**, *Core Data Science*, PhD Research Intern

Conducted large-scale graph-based modeling on what network features predict social trust, informing recommendation algorithms for groups for engagement.

2016 **Airbnb**, PhD Research Intern

Analyzed how trust affects booking on Airbnb to improve booking-related metrics.

Education

2014-2019 **Cornell Tech, Cornell University**, New York, NY
Ph.D., Information Science, Minor in Computer Science
Committee: Mor Naaman (advisor), Karen Levy, Serge Belongie, Jeff Hancock
Networked Trust: Computational Understanding of Interpersonal Trust Online

2010-2014 **Peking University**, Beijing, China
Bachelor of Science, Microelectronics

Publications

2023 **Gemini: A Family of Highly Capable Multimodal Models**. Team, Gemini.

Beyond ChatBots: ExploreLLM for Structured Thoughts and Personalized Model Responses. Ma, Xiao and Mishra, Swaroop and Liu, Ariel and Su, Sophie and Chen,

Jilin and Kulkarni, Chinmay and Cheng, Heng-Tze and Le, Quoc and Chi, Ed. *In submission.*
Filed Patent

Improving Diversity of Demographic Representation in Large Language Models via Collective-Critiques and Self-Voting. Preethi Lahoti, Nicholas Blumm, Xiao Ma, Raghavendra Kotikalapudi, Sahitya Potluri, Qijun Tan, Hansa Srinivasan, Ben Packer, Ahmad Beirami, Alex Beutel, Jilin Chen. *In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP'23)*

Let's Do a Thought Experiment: Using Counterfactuals to Improve Moral Reasoning. Xiao Ma, Swaroop Mishra, Ahmad Beirami, Alex Beutel, Jilin Chen. *In Proceedings of the 2023 International Conference on Machine Learning Neural Conversational AI Workshop (ICML'23)*

A Mixed-Methods Approach to Understanding User Trust after Voice Assistant Failures. Amanda Baughan, Allison Mercurio, Ariel Liu, Xuezhi Wang, Jilin Chen, Xiao Ma. *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI'23)*
Honorable Mention for Best Paper

2022 **A Human-ML Collaboration Framework for Improving Video Content Reviews**. Meghana Deodhar, Xiao Ma, Yixin Cai, Alex Koes, Alex Beutel, Jilin Chen. *In Proceedings of the 31st ACM International Conference on Information and Knowledge Management Human-in-the-Loop Data Curation Workshop (CIKM'22)*
Best Paper

2020 **Challenges in Supporting Exploratory Search through Voice Assistants**. Xiao Ma, Ariel Liu. *In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Workshop (CHI'20)*

AI-Mediated Exchange Theory. Xiao Ma, Taylor W. Brown. *In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Workshop (CHI'20)*

2019 **Understanding Image Quality and Trust in Peer-to-Peer Marketplaces**. Xiao Ma, Lina Mezghani, Kimberly Wilber, Hui Hong, Robinson Piramuthu, Mor Naaman, Serge Belongie. *In Proceedings of 2019 IEEE Winter Conference on Applications of Computer Vision (WACV'19)*

AI-Mediated Communication: How the Perception that Profile Text was Written by AI Affects Trustworthiness. Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T Hancock, Mor Naaman. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*

When Do People Trust Their Social Groups?. Xiao Ma, Justin Cheng, Shankar Iyer, Mor Naaman. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*

2018 **Web-Based VR Experiments Powered by the Crowd**. Xiao Ma, Megan Cackett, Leslie Park, Eric Chien, Mor Naaman. *In Proceedings of the Web Conference 2018 (WWW'18)*

2017 **A Computational Approach to Perceived Trustworthiness of Airbnb Host Profiles**. Xiao Ma, Trishala Neeraj, Mor Naaman. *In Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM'17)*

What Happens in happn? The Warranting Power of Location History. Xiao Ma, Emily Sun, and Mor Naaman. *In Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW'17)*

Self-disclosure and Perceived Trustworthiness of Airbnb Host Profiles. Xiao Ma, Jeffrey T Hancock, Kenneth Lim Mingjie, Mor Naaman. *In Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW'17)*
Honorable Mention for Best Paper

"People are Either too Fake or too Real": Opportunities and Challenges in Tie-based Anonymity. Xiao Ma, Nazanin Andalibi, Louise Barkhuus, Mor Naaman. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI'17)*

2016 **Movement: A Secure Community Awareness Application and Display**. Xiao Ma, Ross McLachlan, Donghun Lee, Mor Naaman, Emily Sun. *In CSCW '16 Companion: Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW'16)*

Anonymity, Intimacy and Self-Disclosure in Social Media. Xiao Ma, Jeffrey T Hancock, Mor Naaman. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16)*

Awards & Honors

2023	Honorable Mention for Best Paper, ACM CHI
2022	Best Paper, CIKM Human-in-the-Loop Workshop
2019	Backslash Art Microgrant
2018	Cornell Outstanding Service Award in Information Science
2017	Women in Technology and Entrepreneurship in New York (WiTNY) Fellowship
2017	Facebook PhD Fellowship Finalist
2017	Honorable Mention for Best Paper, ACM CSCW
2017	Honorable Mention for Snap Research Fellowship

Teaching & Mentoring

Teaching

Spring 2017 **Teaching Assistant**, Connective Media Technologies, Cornell Tech

2014-2016, 2018 **Teaching Assistant**, Psychological and Social Aspects of Connective Media, Cornell Tech

Mentoring

2021	Amanda Baughan, PhD in Computer Science & Engineering, University of Washington
2018	Lina Mezghani, Masters in Computer Science, École Polytechnique
2018	Eric Chien, Masters in Connective Media, Cornell Tech
2017	Trishala Neeraj, Masters in Connective Media, Cornell Tech
2015	Shanshan Zhang, Undergraduate, Beijing University of Posts and Telecommunications

Service

Program Committee

2023-2024	ACM Conference on Fairness, Accountability, and Transparency (FAccT)
2019	Truth and Trust Online (TTO)
2018-2020	International Conference on Computational Social Science (IC2S2)
2018-2019	International Conference on Web and Social Media (ICWSM)
2018	WWW 2018 Satellites: Journalism, Misinformation and Fact Checking
2017,2019	International Conference on Social Informatics (Socinfo)

Conference Reviewer

2023	Empirical Methods in Natural Language Processing (EMNLP)
2023-2024	International Conference on Machine Learning Workshops (ICML)
2023	Conference on Neural Information Processing Systems (NeurIPS)
2021	ACM Conference on Fairness, Accountability, and Transparency (FAccT)
2016-2020	ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)
2016-2024	ACM CHI Conference on Human Factors in Computing Systems (CHI)
2018	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
2018,2020	AAAI International Conference on Web and Social Media (ICWSM)

Journal Reviewer

Social Science Computer Review
ACM Transactions on Social Computing
Media and Society
Journal of Hospitality Management

Recognitions for Outstanding Reviews

CHI 2016, CSCW 2017, CHI 2020, CHI 2021

Languages

Programming: Python, TypeScript (React, Next.js), Prompt Engineering

Natural: English (fluent), Chinese (native), French (intermediate)