

Yuxuan Li

Second-year Ph.D. Student

Human-Computer Interaction Institute, School of Computer Science

Carnegie Mellon University

yuxuanli@andrew.cmu.edu • <https://yuxuanli.com/>

EDUCATION

- 2024–present **Carnegie Mellon University**
Ph.D. in Human-Computer Interaction, School of Computer Science
Advisors: Hirokazu Shirado & Sauvik Das
- 2023 **University of California, Berkeley**
Research Intern, Berkeley AI Research (BAIR) & School of Information
Advisor: Coye Cheshire
- 2020–2024 **Tsinghua University**
B.S. in Computer Science, Computer Science and Technology Department
Advisors: Chun Yu & Yuanchun Shi

PUBLICATIONS

Conference (Peer-reviewed)

- [C2] **Spontaneous Giving and Calculated Greed in Language Models**
EMNLP '25 Yuxuan Li, Hirokazu Shirado
Conference on Empirical Methods in Natural Language Processing, 2025
Oral Presentation | Extensive Media Coverage
- [C1] **Actions Speak Louder than Words: Agent Decisions Reveal Implicit Biases in Language Models**
FAccT '25 Yuxuan Li, Hirokazu Shirado, Sauvik Das
ACM Conference on Fairness, Accountability, and Transparency, 2025

Preprint

- [P5] **What Makes LLM Agent Simulations Useful for Policy? Insights From an Iterative Design Engagement in Emergency Preparedness**
CHI '26 Yuxuan Li, Sauvik Das, Hirokazu Shirado
Under review at ACM Conference on Human Factors in Computing Systems, 2026
- [P4] **HiddenBench: Assessing Collective Reasoning in Multi-Agent LLMs via Hidden Profile Tasks**
ICLR '26 Yuxuan Li, Aoi Naito, Hirokazu Shirado
Under review at International Conference on Learning Representations, 2026
- [P3] **CoSim: LLM-based Simulation System Trains Student Counselors to Communicate with College Students Experiencing Stress and Anxiety**
CHI '26 Jie Cai, Wenjing Deng, Yunfei Chen, Yanshan Lin, Dongzhe Zheng, He Zhang, Yuxuan Li[◊], John M. Carroll, Weite Zhang, Chun Yu (◊denotes corresponding author)
Under review at ACM Conference on Human Factors in Computing Systems, 2026
- [P2] **A Human-Computer Collaborative Tool for Training a Single Large Language Model Agent into a Network through Few Examples**
ARXIV '24 Lihang Pan*, Yuxuan Li*, Chun Yu, Yuanchun Shi (*denotes equal contribution)
arXiv, 2024

[P1] **Say Your Reason: Extract Contextual Rules In Situ for Context-aware Service Recommendation**
 ARXIV '24 **Yuxuan Li**, Jiahui Li, Lihang Pan, Chun Yu, Yuanchun Shi
arXiv, 2024

AWARDS & HONORS

2023, 2022	Academic Excellence Scholarship , Tsinghua University
2021	Sport Excellence Scholarship , Tsinghua University
2020	Freshman Scholarship , Tsinghua University

WORKSHOPS ORGANIZED

[W1] **PoliSim@CHI 2026: LLM Agent Simulation for Policy**
 CHI 2026 **Yuxuan Li**, Wesley Hanwen Deng, Xuhui Zhou, Kevin Klyman, Chun Yu, Yuanchun Shi, Nicholas Vincent, Amy X. Zhang, Maarten Sap, Sauvik Das, Hirokazu Shirado
ACM Conference on Human Factors in Computing Systems, 2026

SELECTED MEDIA COVERAGE

2025	Smarter AI Models Show Selfish Behavior in Study By <i>The Chosun Ilbo</i>
2025	CMU researchers find selfish traits in cutting-edge AI By <i>TribLive</i>
2025	AI showing signs of selfishness, researchers warn of troubling trend By <i>The News International</i>
2025	Is AI Becoming Selfish? CMU Researchers Discover Certain AI Models Adopt Self-Seeking Behavior By <i>CMU SCS News</i>
2025	When AI LLMs think more, groups suffer, CMU study finds By <i>Geneva Internet Platform (Dig.watch)</i>
2025	Yapay zeka akıl yürütme kapasitesi arttıkça bencilleşiyor By <i>NTV</i>

MENTORSHIP

2025–present	Leyang Li , undergrad at University of Notre Dame Ongoing project for USENIX Security 2026
2025–present	Jason Alexander , undergrad at University of Massachusetts, Amherst Ongoing project for FAccT 2026

INVITED TALKS AND GUEST LECTURES

12/2025	Examine Machine Behavior and Simulate Society with Social Agents Microsoft Research Asia, Social Computing Group
12/2025	Tsinghua University, Pervasive HCI Group
10/2025	Georgia Institute of Technology, Graph Computation and Machine Learning Lab
10/2025	Carnegie Mellon University, Social Web Course
10/2025	Carnegie Mellon University, FEAT (Fairness, Ethics, Accountability, Transparency) AI Seminar

- 06/2025 **Actions Speak Louder than Words: Agent Decisions Reveal Implicit Biases in Language Models**
Microsoft Research, alt.FAcCT
- 11/2024 **A Practical Guide to Building Social Agents**
Carnegie Mellon University, Social Agent Course

SKILLS

Programming: Python, C, C++, Java, JavaScript, QT, System Verilog

Machine Learning: PyTorch, TensorFlow, Keras, LLM Post-training, AI Evaluation

Data: Statistical Modeling, Large-scale Data Analysis, Survey Instrumentation