

Yuxuan Li

Second-year Ph.D. Student

Human-Computer Interaction Institute, School of Computer Science
Carnegie Mellon University

yuxuanll@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

[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

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

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

Lihang Pan*, Yuxuan Li*, Chun Yu, Yuanchun Shi (*denotes equal contribution)

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	<u>Yuxuan Li</u> , Wesley Hanwen Deng, Xuhui Zhou, Kevin Klyman, Chun Yu, Yuanchun Shi, Nicholas Vincent, Amy X. Zhang, Maarten Sap, Sauvik Das, Hirokazu Shirado <i>Under review at ACM Conference on Human Factors in Computing Systems, 2026</i>

SELECTED MEDIA COVERAGE

2025	CMU researchers find selfish traits in cutting-edge AI By <i>TribLive</i>
2025	As AI grows smarter, it may also become increasingly selfish By <i>Tech Xplore</i>
2025	Smarter AI models show more selfish behavior By <i>PsyPost</i>
2025	Is Artificial Intelligence Developing Self-Interested Behavior? By <i>BioEngineer.org</i>
2025	The smarter AI gets, the more selfish it becomes, study warns By <i>Knowridge Science Report</i>
2025	Smarter AI, Selfish AI: Carnegie Mellon Study Reveals Downside Of Machine Reasoning By <i>AZoAI</i>
2025	Is AI Becoming Selfish? CMU Researchers Discover Certain AI Models Adopt Self-Seeking Behavior By <i>CMU SCS News</i>
2025	As AI Develops Further, it Could Become More Self-Serving By <i>Scitake</i>

MENTORSHIP

2025–present	Leyang Li , undergrad at University of Notre Dame Ongoing project for IEEE S&P 2026
2025–present	Jason Alexander , undergrad at University of Massachusetts, Amherst Ongoing project for FAccT 2026

INVITED TALKS AND GUEST LECTURES

11/2025	Spontaneous Giving and Calculated Greed in Language Models Carnegie Mellon University, Foundations of Cooperative AI Lab
---------	--

10/2025	Examine Machine Behavior and Simulate Society with Social Agents Georgia Institute of Technology, Graph Computation and Machine Learning Lab
10/2025	Examine Machine Behavior and Simulate Society with Social Agents Carnegie Mellon University, Social Web Course
10/2025	Examine Machine Behavior and Simulate Society with Social Agents 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