

# 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

- [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<sup>◊</sup>, 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	<b>Academic Excellence Scholarship</b> , Tsinghua University
2021	<b>Sport Excellence Scholarship</b> , Tsinghua University
2020	<b>Freshman Scholarship</b> , 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  
*Under review at ACM Conference on Human Factors in Computing Systems, 2026*

## SELECTED MEDIA COVERAGE

---

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

## MENTORSHIP

---

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

## INVITED TALKS AND GUEST LECTURES

---

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

10/2025	<b>Examine Machine Behavior and Simulate Society with Social Agents</b> Georgia Institute of Technology, Graph Computation and Machine Learning Lab
10/2025	<b>Examine Machine Behavior and Simulate Society with Social Agents</b> Carnegie Mellon University, Social Web Course
10/2025	<b>Examine Machine Behavior and Simulate Society with Social Agents</b> Carnegie Mellon University, FEAT (Fairness, Ethics, Accountability, Transparency) AI Seminar
06/2025	<b>Actions Speak Louder than Words: Agent Decisions Reveal Implicit Biases in Language Models</b> Microsoft Research, alt.FAccT
11/2024	<b>A Practical Guide to Building Social Agents</b> 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