

Andy Liu

andyliu.github.io
andyliu@cs.cmu.edu
Last updated October 14, 2026

Language Technologies Institute
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213

◇ EDUCATION

- 2023 – Present **Carnegie Mellon University**, Pittsburgh, PA
PH.D. IN LANGUAGE AND INFORMATION TECHNOLOGIES
Advisors: Mona Diab, Daniel Fried
Cumulative GPA: 4.05
Relevant Coursework: LLM Ethics, Safety, & Social Impact, Large Language Models, Advanced Natural Language Processing, Quantitative Evaluation of Language Technologies, Deep Reinforcement Learning, Multimodal Machine Learning, Machine Learning & Game Theory, Cooperative AI
- 2019 – 2023 **Harvey Mudd College (HMC)**, Claremont, CA
B.S. IN COMPUTER SCIENCE & MATHEMATICS
CONCENTRATION IN SCIENCE, TECHNOLOGY, AND SOCIETY

◇ HONORS & AWARDS

- 2025 NSF GRADUATE RESEARCH FELLOWSHIP
2023 DEPARTMENTAL HONORS, HMC
2019 - 2023 HARVEY S. MUDD MERIT AWARD, HMC
2019 - 2023 LOCKHEED MARTIN STEM SCHOLARSHIP
2019 - 2023 NATIONAL MERIT SCHOLARSHIP

◇ PUBLICATIONS

— Conference Proceedings —

- 2026 Saket Reddy and **Andy Liu**. “A Geometric Perspective on Stabilizing Value Conflict Resolution.” *Asia-Pacific Chapter of the Association for Computational Linguistics (ACL)*, 2026.
- 2026 Kshitish Ghate, **Andy Liu**, Devansh Jain, Taylor Sorensen, Atoosa Kasirzadeh, Aylin Caliskan, Mona T. Diab, and Maarten Sap. “EVALUESTEER: Measuring Reward Model Steerability Towards Values and Preferences.” *In Findings of the Association for Computational Linguistics: EMNLP 2026*.
- 2026 **Andy Liu**, Kshitish Ghate, Mona Diab, Daniel Fried, Atoosa Kasirzadeh, and Max Kleiman-Weiner. “Generative Value Conflicts Reveal LLM Priorities.” *The Fourteenth Annual International Conference on Learning Representations*.
- 2025 Wenkai Li*, Jiarui Liu*, **Andy Liu**, Xuhui Zhou, Mona Diab, and Maarten Sap. “BIG5-CHAT: Shaping LLM Personalities Through Training on Human-Grounded Data.” *Annual Meeting of the Association for Computational Linguistics*, 2025.
- 2025 **Andy Liu***, Abhishek Kulkarni*, Jean-Raphaël Gaglione, Daniel Fried, and Ufuk Topcu. “Dynamic Coalition Structure Detection in Natural-Language-based Interactions.” *The Twenty-Fourth International Conference on Autonomous Agents and Multi-Agent Systems*.
- 2024 **Andy Liu**, Mona Diab, and Daniel Fried. “Evaluating Large Language Model Biases in Persona-Steered Generation.” *Findings of the Association for Computational Linguistics*, 2024.

2023 **Andy Liu**, Hao Zhu, Emmy Liu, Yonatan Bisk, and Graham Neubig. “Computational Language Acquisition with Theory of Mind.” *The Eleventh Annual International Conference on Learning Representations*.

— **Workshop Papers** —

2025 Kushal Agrawal*, Verona Teo*, Juan J Vazquez*, Sudarsh Kunnavakkam, Vishak Srikanth, and **Andy Liu**. “Evaluating LLM Agent Collusion in Double Auctions.” *Multi-Agent Systems in the Era of Foundation Models @ ICML 2025*.

— **Preprints & Manuscripts** —

2026 **Andy Liu**, Mehar Bhatia, Karolina Stanczak, Mona Diab, Vered Shwartz, and Daniel Fried. “Predicting Alignment Generalization with Value Representations.”

2025 Wenkai Li, Lynnette Hui Xian Ng, **Andy Liu**, and Daniel Fried. “Measuring Fine-Grained Negotiation Tactics of Humans and LLMs in Diplomacy.”

◇ **TEACHING**

Fall 2026 AI AGENTS (11-768), Teaching Assistant, CMU
Spring 2026 ADVANCED NATURAL LANGUAGE PROCESSING (11-711), Teaching Assistant, CMU
Fall 2022 NATURAL LANGUAGE PROCESSING (CSCI 159), Teaching Assistant, HMC
Fall 2021 PRINCIPLES OF COMPUTER SCIENCE (CSCI 060), Teaching Assistant, HMC

◇ **INTERNSHIPS & ASSISTANTSHIPS**

— **Industry** —

8/25 – 3/26 **Meta**, Seattle, WA
RESEARCH SCIENTIST INTERN, FAIR Communication & Language
Supervisors: Ruta Desai, Ansong Ni
Implemented environments to evaluate how well LLM self-play generalizes to novel partners in collaborative tasks, then developed persona-based population play methods to diversify training and improve zero-shot coordination abilities.

— **Academia** —

1/22 – 5/23 **HMC Department of Computer Science**, Claremont, CA
UNDERGRADUATE RESEARCHER, Lab for CATS (Cognition and Attention over Time and Space)
Advisor: Calden Wloka
Fine-tuned models for action recognition on 3D video input, then used GradCAM-based network visualization techniques to analyze neural network attention and interpret model failure cases.

5/22 – 8/22 **CMU Language Technologies Institute**, Pittsburgh, PA
RESEARCH INTERN, NeuLab
Advisors: Graham Neubig, Yonatan Bisk
Studied computational models of reinforcement learning-based language learning agents to study the influences of theory of mind modeling and environmental pressures on language acquisition.