Jiayi Yuan

jiayiy9@cs.washington.edu | yuanjiayiy.github.io | (540) 255-9188

RESEARCH INTERESTS AI4Science; automated scientific discovery; multi-agent systems; reinforcement learning; LLM alignment and safety; open-endedness.

EDUCATION

University of Washington 2025 - Present
Ph.D. in Computer Science and Engineering
Advisor: Natasha Jaques

University of Washington 2022 - 2024
Professional Master of Science in Computational Linguistics
Advisor: Shane Steinert-Threlkeld

Carnegie Mellon University 2018 - 2021
Bachelor of Science in Neuroscience and Computer Science
University Honor, summa cum laude

PUBLICATIONS

Jiayi Yuan*, Hangoo Kang*, James Jihao Liu*, Yejin Choi, Vikram Iyer, Liwei Jiang, and Natasha Jaques. "Forty Shades of Blue: Quality-Diversity Alignment via Mode-Conditioned Reinforcement Learning". NeurIPS 2026 submission.

Jiayi Yuan*, Jonas Noether, Natasha Jaques, and Goran Radanovic. "AgenticRed: Optimizing Agentic Systems for Automated Red-teaming". Preprint. arXiv:2601.13518.

Julian Lehmkuhl*, Abel Ilyes-Kun*, Nicholas Bremes*, C. K. Ozaltan*, Felix Muthers*, and **Jiayi Yuan.** "Generating Piano Music with Transformers: A Comparative Study of Scale, Data, and Metrics". NeurIPS 2025 Workshop: AI4Music. arXiv:2511.07268.

Xiaoxuan Hou*, **Jiayi Yuan*,** Joel Z. Leibo, and Natasha Jaques. "InvestESG: A Multi-Agent Reinforcement Learning Benchmark for Studying Climate Investment as a Social Dilemma". ICLR 2025.

Marcel Torne*, Arhan Jain*, **Jiayi Yuan***, Vidyaaranya Macha*, Lars Ankile, Anthony Simeonov, Pulkit Agrawal, and Abhishek Gupta. "Robot Learning with Super-Linear Scaling". RSS 2025.

INVITED TALKS

"Automated Red-Teaming of LLM and the Future of LLM Safety". Google Red Team Seminar. Seattle, USA. March 2026.

"Addressing Artificial Hivemind Through Post-Training". DARPA In the Moment (ITM) PM Meeting. Seattle, USA. February 2026.

"Hour of Code: Automated Scientific Discovery". Ballard High School Seattle Public School Hour of Code program. Seattle, USA. November 2025.

"DiZCo: Planning Zero-Shot Coordination in World Models". Allen School Annual Industry Affiliates Research Showcase. Seattle, USA. November 2025.

"AI for Climate Change: Multi-Agent Reinforcement Learning Approaches". Cooperation for Climate Change Action Workshop. Lausanne, Switzerland. June 2025.

"Idea and Direction about Emotionally-intelligent AI". Cooperative AI Summer School. Marlow, UK. July 2025.

"InvestESG: A Multi-Agent Reinforcement Learning Benchmark for Studying Climate Investment as a Social Dilemma". ICLR 2025. Singapore. April 2025.

"InvestESG: A Multi-Agent Reinforcement Learning Benchmark for Studying Climate Investment as a Social Dilemma". NeurIPS 2024 Workshop on Tackling Climate Change with Machine Learning. Vancouver, BC. December 2024.

"Adapting Data Preparation Tools to the Era of LLM: Introducing iDPS CLI and Data Quality Report Tool". 1st Amazon AGI Engineering Workshop. Bellevue, WA. April 2024.

SERVICE

Reviewer, AMLC 2024; ICML 2025, 2026; ICLR 2026; NeurIPS 2025, 2026; ICLR RSI Workshop 2026; ICML Human-AI Co-creativity Workshop 2026.

Organizer, SocialRL Reading Group.

GPSS Senator, Allen School for Computer Science & Engineering.

Session Chair, DUB Community Day.

Mentor, Allen School Pre-Application Mentorship Service (PAMS), the University of Washington Math AI Lab, UW Undergraduate Guided Research.

Tutorial Organizer, WEIRDLab, 2024.

RESEARCH EXPERIENCE

Paul G. Allen Center for Computer Science & Engr., University of Washington
2025 - Present
Research Assistant

- Apply modern machine learning methods to scientific discovery and real-world decision-making, with applications in scientific ideation, LLM safety, Earth and environmental sciences, and robot learning.
- Mentor junior researchers on research projects.

Max Planck Institute for Software Systems (MPI-SWS), Saarbruecken, Germany
2025
Research Intern

- Led AgenticRed, an evolutionary framework for automatically optimizing LLM red-teaming agents, *supervised by Goran Radanovic.*

INDUSTRY EXPERIENCE

Amazon AGI, Seattle, WA 2021 - 2024
Software Development Engineer II

- Built iDPS CLI for launching and monitoring large-scale data processing jobs for Amazon Nova models, reducing development effort by 80%.
- Led development of AGI's first Data Quality Report Tool, automating training-data inspection and reducing development time by 95%.
- Developed SentenceBERT-based inconsistency checkers for Alexa NLU data, improving model accuracy by 25% and reducing annotation resolution time by 50%.
- Presented tools and tutorials to 300+ engineers and scientists; received Amazon AGI org-level Peer Recognition Awards in 2022 and 2023.

RECOGNITION

Allen School First-Year Fellowship	2025
Cooperative AI Summer School Scholarship	2025
NYU Abu Dhabi Global PhD Fellowship (offered)	2025
NYU Tandon School of Engineering Graduate Fellowship (offered)	2025
Amazon AGI Organization Annual Peer Recognition Award	2023
Amazon AGI Organization Annual Peer Recognition Award	2022
University Honor, summa cum laude, Carnegie Mellon University	2021
Summer Undergraduate Research Fellowship, Carnegie Mellon University	2019

MENTORING

Abel Ilyes-Kun, Master student, RWTH.
Yuxin Li, Undergraduate student, Columbia University.
Shima Rezai, Master student, Sharif University of Technology.
Nalin Poungpeth, Undergraduate student, Northwestern University.

SKILLS

Programming: Python, Java, C/C++, TypeScript, MATLAB, Bash, SQL.
Machine learning: PyTorch, JAX, Verl, TensorFlow, scikit-learn, Apache Spark, OpenCV.
Robotics and RL: ROS, Gym, MoveIt, Gazebo, IsaacSim, PettingZoo, SB3.
Systems: AWS, CUDA, GCP, Azure, Docker, Git, Conda, Slurm.