

Zeyu (Andy) Liang

Male, born in March 2002.

Email: liang.zey@northeastern.edu

Research Interests: Reinforcement Learning, Multi-Agent Systems, LLM Agents, AI Safety

Github: zedyellion.github.io

Phone: 18701388750



🎓 Education

Master of Artificial Intelligence,

Northeastern University

Sep. 2024 - Dec. 2026

Bachelor of Mathematics,

University of California San Diego

Sep. 2020 - Jan. 2024

📖 Publications

[1] Shuo Liu, **Zeyu Liang**, Xueguang Lyu, Christopher Amato, LLM Collaboration With Multi-Agent Reinforcement Learning. AAAI, 2026. <https://ojs.aaai.org/index.php/AAAI/article/view/40487/44448>

[2] Dongyu Ma*, **Zeyu Liang***, Luming Yin, Hongliang Liang, Symbolic testing of floating-point bugs and exceptions, Journal of Systems and Software. 2024, <https://doi.org/10.1016/j.jss.2024.112226>

[3] **Zeyu Liang**, Lei Wang, Parallel Computing on RTEMS Operating System, 2024 BenchCouncil International Symposium on Intelligent Computers, Algorithms, and Applications (IC 2024). https://link.springer.com/chapter/10.1007/978-981-96-6310-1_15

[4] **Zeyu Liang**, Zihao Li, Xiao Liang, Zhen Han, MADPO: Multi-Agent Direct Preference Optimization for Cooperative LLM Agents, (under review) 2026. <https://zedyellion.github.io/files/madpo.pdf>

🏆 Honors and Awards

2022Winter, 2021Spring, 2020Fall: Provost Honors at UCSD.

2019: Distinguished Graduate Honor at BeiJing National Day School (BNDS).

2018: Duke Math Competition, China Regional Team Gold Award

💼 Internship Experience

MADPO: Direct Preference Optimization for Cooperative LLM Agents Amazon, London, UK
May 2026 — Aug. 2026

Proposed MADPO, a direct preference-learning framework for cooperative LLM-agent teams.

- Extended Direct Preference Optimization (DPO) from a single-agent policy to a factored team of role-specialized agents, so that the whole team is trained directly from preferences over joint outputs, without an engineered reward, an intermediate reward model, or online environment rollouts.
- Established the method's theoretical properties — its equivalence to DPO on the joint policy, a shared team-level learning signal analogous to credit assignment in cooperative multi-agent RL, and the boundary at which joint and independent training coincide — and validated it on both sequentially-coupled and structurally-independent multi-agent domains, comparing against reward-model-based RL and inference-time coordination baselines.
- Responsible for the method design, implementation, experiments, and manuscript writing.

STAR-MARL: Multi-Agent Reinforcement Learning DeepFlow, London, UK Jan. 2026 — Apr. 2026

- Contributed to advancing STAR-MARL by integrating LLM-generated hierarchical sub-task trees, executable training scenarios, and reward functions into a multi-agent reinforcement learning pipeline.
- Improved a trajectory-informed, retrieval-augmented generation (RAG) workflow that uses a historical component-evaluation knowledge base to assess and iteratively refine reward functions for new sub-tasks offline.
- Conducted training and benchmark evaluations in CookingZoo and Google Research Football; validated faster performance gains over MAPPO in CookingZoo and analyzed knowledge-base quality and cross-scenario transfer bottlenecks in the more complex football task.

Parallel Computing on RTEMS Operating System Beijing Huayu Zhuocheng Software Co. Mar. 2024 — Aug. 2024

- Implemented a heterogeneous parallel computing framework on the RTEMS operating system.

- Designed the task scheduling mechanism and a Master-Slave communication protocol to support distributed execution.
- Integrated BLAS and user-defined computational operators (validated 142 BLAS operators).
- Deployed and evaluated the system on an x86-ARM heterogeneous cluster with up to $14.1\times$ speedup.
- Conducted comprehensive experimental evaluation and wrote the paper (published, IC 2024).

Academic Projects

LLM Collaboration with Multi-Agent RL (MAGRPO) (*published at AAAI 2026*) May 2025 — August 2025

We proposed MAGRPO, a multi-agent reinforcement learning method that trains cooperating LLMs with group-relative policy optimization (GRPO) under a shared team reward. My core contributions:

- Designed and built the data pipeline for the multi-agent collaboration tasks used to train and evaluate the method.
- Identified and fixed a critical evaluation flaw that had significantly inflated reported model performance by counting incorrect predictions as correct, correcting the study's central empirical results.
- Built the evaluation pipeline and ran the benchmark experiments reported in the paper.
- Produced the experimental figures and contributed to the manuscript.

GOSSIP Summer School (Shanghai Jiao Tong University)

July, 2024

Completed lectures and hands-on labs on real-world software security, covering binary and code analysis, vulnerability discovery and exploitation, and cryptographic implementation flaws.

Symbolic Testing of Floating-Point Bugs and Exceptions (*published in JSS 2024*) Oct. 2023 — Mar. 2024

Participated in the research and development of the Yuse tool, which leverages mathematical function modeling and symbolic execution to detect floating-point defects and exceptions in software. My core contributions:

- Implemented floating-point constraint construction, interval propagation, and math function modeling.
- Designed two-phase constraint solving combining range solver and SMT solver for performance optimization.
- Participated in manuscript writing

Technical Skills

Machine Learning: Transformer, RLHF, PPO, GRPO

Data Analysis: R, MATLAB

Tools and Programming: Python, Java, C/C++, PyTorch, Git, Linux, Vim, LaTeX

Volunteer Experience and Hobbies

2024: Go (Weiqi) Tutor at the Maine Go Club, USA.

2019: Go (Weiqi) Tutor at the Haidian Go Club, Beijing, China.

2020, 2018: Beijing Marathon (Half Marathon).