Xiangyu Liu

✉ xyliu999@umd.edu

Education

University of Maryland, College Park — **College Park, MD**
PhD student in Computer Science — *Sep. 2021 – Jun. 2025 (expected)*
- Advisor: Kaiqing Zhang

Shanghai Jiao Tong University (SJTU) — **Shanghai, China**
B.E. in Computer Science — *Sep.2017 – Jun.2021*
- Zhiyuan Honors Program of Engineering (an elite program for top 5% talented students)

University of California, Berkeley — **Berkeley, CA**
Exchange Student — *Jan. 2020 – May 2020*
GPA: 4.0/4.0

Experience

Bloomberg AI — **NY, USA**
Research intern — *June.2022 – Sep.2022*
- Research on bond pricing with recurrent neural networks.

Research Interests

My research interests are centered around sequential decision-making, with a particular emphasis on the fundamental aspects of ***reinforcement learning (RL)*** in environments characterized by ***multi-agent*** interactions, ***partial observability***, and ***adversarial*** conditions.

Recently, my research is also extended to include two key application domains:
- ***Large Language Model (LLM) Agent“s”***: My interest lies in harnessing the in-context learning capabilities of LLMs for decision-making in dynamic and strategic environments.
- ***Reinforcement Learning from Human Feedback (RLHF)***: We are investigating strategies for identifying and mitigating the impact of noise or adversarial interventions within human feedback, aiming to enhance the reliability and performance of RLHF methodologies.

Working in Progress

- Chanwoo Park*, **Xiangyu Liu***, Asuman E. Ozdaglar, Kaiqing Zhang (* denotes equal contribution)
  Do LLM Agents Have Regret? A Case Study in Online Learning and Games.
  Under Review
- Pankayaraj Pathmanathan, Souradip Chakraborty, **Xiangyu Liu**, Yongyuan Liang, Furong Huang
  Is Poisoning a Real Threat to LLM Alignment? Maybe More So Than You Think
  Under Review

Publications

- **Xiangyu Liu**, Hangtian Jia, Ying Wen, Yujing Hu, Yingfeng Chen, Changjie Fan, Zhipeng Hu, Yaodong Yang
  Towards Unifying Behavioral and Response Diversity for Open-ended Learning in Zero-sum Games

NeurIPS 2021

- **Xiangyu Liu**, Kaiqing Zhang
Partially Observable Multi-agent RL with (Quasi-)Efficiency: The Blessing of Information Sharing
ICML 2023

- **Xiangyu Liu**, Souradip Chakraborty, Yanchao Sun, Furong Huang
Rethinking Adversarial Policies: A Generalized Attack Formulation and Provable Defense in RL.
ICLR 2024 and **Outstanding Paper Award** at NeurIPS 2022 Workshop on Trustworthy and Socially Responsible Machine Learning.

- **Xiangyu Liu**, Chenghao Deng, Yanchao Sun, Yongyuan Liang, Furong Huang
Beyond Worst-case Attacks: Robust RL with Adaptive Defense via Non-dominated Policies.
ICLR 2024 Spotlight (Top 5%).

- Yongyuan Liang, Yanchao Sun, Ruijie Zheng, **Xiangyu Liu**, Tuomas Sandholm, Furong Huang, Stephen McAleer
Game-theoretic Robust Reinforcement Learning Handles Temporally-coupled Perturbations.
ICLR 2024.

- Yang Cai†, **Xiangyu Liu**†, Argyris Oikonomou†, Kaiqing Zhang† († denotes alphabetical order)
Provable Partially Observable Reinforcement Learning with Privileged Information.
NeurIPS 2024.

Awards and Scholarships

- **Outstanding Paper Award**, NeurIPS 2022 Workshop on Trustworthy and Socially Responsible Machine Learning. 2022
- **Dean's Fellowship**, University of Maryland, College Park. 2021
- **National Scholarship** (Top 0.2% in China), Ministry of Education of P.R.China. 2018&2019
- **A-class Scholarship for Excellent Academic Performance** (Top 1% at SJTU), SJTU. 2018
- **1st Prize in Chinese College Mathematics Competitions** (Top 1 at SJTU, selected for final). 2018

Outreach

- TAs: Common-sense reasoning in NLP (Fall 2021); Cryptography (Spring 2022)
- One of student organizers of summer AI camps at UMD 2023 for K-12 students

Skills

- Programming Languages: Python, C/C++, Java, MATLAB, LaTeX
- Deep Learning Packages: PyTorch, TensorFlow