MIAO LU

Tel: 650-250-9790 | E-mail: miaolu@stanford.edu | Web: miaolu3.github.io

Google Scholar | LinkedIn | Last update: September, 2025

Research Interests

My research interest centers around *reinforcement learning for LLM agents and training* [C9][P4][P5]. With backgrounds in probability and statistics, my past research includes mathematical theory and algorithm design of provably sample-efficient reinforcement learning [C4][C5], reinforcement learning from human feedback for LLM [C9], robust reinforcement learning [C1][C6][C8], training dynamics and generalization of optimization in deep learning landscapes [C2][C7][C10][C11][P3], and reinforcement learning for operations and economics [C3][P1][P2].

Education

Stanford University — *Stanford, USA*
Ph.D. Candidate in Operations Research — *Sep.2023 – present*
Advisor: Prof. Jose Blanchet

University of Science and Technology of China — *Hefei, China*
B.S. in Mathematics & Applied Mathematics — *Sep.2018 – Jun.2022*
Ranking: 2/140, with summa cum laude (Guo Moruo Scholarship)

Industrial Experiences

ByteDance Seed, Research Scientist Intern — *San Jose, USA*
Host: Dr. Jiecao Chen — *Jun.2025 – present*
Developed end-to-end agentic RL algorithms to scale LLM agent training for long horizon tasks beyond a fixed context limit, with techniques of summarization and context-folding driven context management. Achieved significant improvements over baseline agentic RL in deep research and coding agent with Seed-OSS-36B-Instruct. Progress results in papers [P4][P5].

Ubiquant Investment, Quantitative Research Intern — *Shanghai, China*
Interned at AI department of Ubiquant, research on DL and RL for quantitative trading. — *Jun.2022 – Sep.2022*

Research Experiences

Stanford University, Graduate Research Assistant — *Stanford, USA*
Advisor: Prof. Jose Blanchet — *Sep.2023 – present*
Projects: (i) LLM RLHF from limited data and distributional shifts, the first to propose SFT as regularization in preference learning with theoretical grounding [C9]; (ii) theory and algorithm of distributionally robust decision-making [C6][C8][P2].

Toyota Technological Institute at Chicago, Student Visitor — *Chicago, USA*
Host: Prof. Tianhao Wang and Prof. Zhiyuan Li — *July.2024 – Sep.2024*
Projects: (i) theoretical understanding of heavy-ball momentum acceleration with large learning rates in river-valley loss landscapes [P3]; (ii) optimal computational-statistical trade-off of learning single-index models via neural networks [C10].

Northwestern University, Remote Research Assistant — *Remote*
Host: Prof. Zhaoran Wang — *Sep.2021 – Aug.2023*
Selected projects: Principled exploration for online RL under function approximations via *value-incentivized regularization* (Maximize to Explore [C5]). The method inspires a long line of following work in online/offline RL, multi-agent RL, RLHF.

Awards and Honors

Xinhe Scholarship (outstanding undergraduate researchers, School of the Gifted Young, USTC) — *Mar.2023*
Yuanqing Yang Scholarship (top scholarship, School of Mathematical Sciences, USTC) — *Jan.2022*
The 41st Guo Moruo Scholarship (highest honor, USTC) — *Dec.2021*
Chinese National Scholarship (top scholarship, Ministry of Education of China) — *Nov.2019, 2020*

Invited Talks

Theoretical Foundations of Distributionally Robust Decision Making [C6][C8][P2]

- 2025 INFORMS annual meeting, Atlanta GA, USA [P2] *Oct.2025*
- 2024 INFORMS annual meeting, Seattle, WA, USA [C8] *Oct.2024*
- 58th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA [C6] *Mar.2024*
- 2023 INFORMS annual meeting, Phoenix, AZ, USA [C6] *Oct.2023*

Computational-statistical Trade-off of Learning Single-index Models via Neural Networks [C10]

- 2nd Mathematics of Modern Machine Learning Workshop (M3L), Vancouver, BC, Canada *Dec.2024*

Skills

Quantitative Skills: Statistics, Optimization, Reinforcement learning, Large language models

Programming Languages and Tools: Fluent: Python, Pytorch, VeRL, Git, LaTeX; Familiar: C/C++

Academic Services

Journal Reviewer: Annals of Applied Probability (AOAP), Operations Research (OR), Mathematics of Operations Research (MOR), Transactions on Machine Learning Research (TMLR)

Conference Reviewer: Neural Information Processing Systems (NeurIPS; 2023-2025), International Conference on Machine Learning (ICML; 2024-2025), International Conference on Learning Representations (ICLR; 2024-2026), International Conference on Artificial Intelligence and Statistics (AISTATS; 2025), ICML Workshop on Aligning Reinforcement Learning Experimentalists and Theorists (ARLET; 2024), ICML Workshop on Exploration in AI Today (EXAIT; 2025), ICML Workshop on Methods and Opportunities at Small Scale (MOSS; 2025), NeurIPS Workshop on Mathematics of Modern Machine Learning (M3L; 2024), Association for the Advancement of Artificial Intelligence (AAAI; 2025)

Publications & Preprints

Authors with * contributed equally to the work, and † represents alphabetical order.

[P5] **Scaling Long-Horizon Agent via Context Folding**
Weiwei Sun, **Miao Lu**, Zhan Ling, Xuesong Yao, Kang Liu, Yiming Yang, Jiecao Chen
Preprint under conference review

[P4] **Scaling LLM Multi-Turn RL with End-to-end Summarization-based Context Management**
Miao Lu, Weiwei Sun, Weihua Du, Zhan Ling, Xuesong Yao, Kang Liu, Jiecao Chen
Preprint under conference review

[P3] **Towards Understanding Momentum Acceleration in River-Valley Loss Landscape**
Miao Lu, Kaiyue Wen, Zeyu Bian, Beining Wu, Siyu Chen, Tianhao Wang, Zhiyuan Li
Preprint under conference review

[P2] **Robust Assortment Optimization from Observational Data with Near-Optimal Sample Complexity**
Miao Lu, Yuxuan Han, Han Zhong, Zhengyuan Zhou, Jose Blanchet
Preprint in preparation

[P1] **Learning an Optimal Assortment Policy under Observational Data**
Yuxuan Han, Han Zhong, **Miao Lu**, Jose Blanchet, Zhengyuan Zhou
Preprint under review at Management Science (MS)

[C11] **Towards Theoretical Understanding of Transformer Test-Time Computing: Investigation on In-Context Linear Regression**
Xingwu Chen, **Miao Lu**, Beining Wu, Difan Zou
NeurIPS Workshop on Foundations of Reasoning in Language Models (FoRLM) 2025
Preprint under conference review

[C10] **Can Neural Networks Achieve Optimal Computational-statistical Tradeoff? An Analysis on Single-Index Model**
Siyu Chen*, Beining Wu*, **Miao Lu**, Zhuoran Yang, Tianhao Wang
International Conference on Learning Representations (ICLR) 2025 **Oral presentation**
NeurIPS Workshop on Mathematics of Modern Machine Learning (M3L) 2024 **Oral presentation**

[C9] **Provably Mitigating Overoptimization in RLHF: Your SFT Loss is Implicitly an Adversarial Regularizer**
Zhihan Liu*, **Miao Lu***, Shenao Zhang, Boyi Liu, Hongyi Guo, Yingxiang Yang, Jose Blanchet, Zhaoran Wang
Neural Information Processing Systems (NeurIPS) 2024
ICML Workshop on Aligning Reinforcement Learning Experimentalists and Theorists (ARLET) 2024

[C8] **Distributionally Robust Reinforcement Learning with Interactive Data Collection: Fundamental Hardness and Near-Optimal Algorithm**
Miao Lu*, Han Zhong*, Tong Zhang, Jose Blanchet
Neural Information Processing Systems (NeurIPS) 2024
ICML Workshop on Aligning Reinforcement Learning Experimentalists and Theorists (ARLET) 2024

[C7] **Benign Oscillation of Stochastic Gradient Descent with Large Learning Rates**
Miao Lu*, Beining Wu*, Xiaodong Yang, Difan Zou
International Conference on Learning Representations (ICLR) 2024
NeurIPS Workshop on Mathematics of Modern Machine Learning (M3L) 2023

[C6] **Double Pessimism is Provably Efficient for Distributionally Robust Offline Reinforcement Learning: Generic Algorithm and Robust Partial Coverage**
Jose Blanchet†, **Miao Lu**†, Tong Zhang†, Han Zhong†
Neural Information Processing Systems (NeurIPS) 2023
Extended version under major revision at Mathematics of Operations Research (MOR)

[C5] **Maximize to Explore: One Objective Function Fusing Estimation, Planning, and Exploration**
Zhihan Liu*, **Miao Lu***, Wei Xiong*, Han Zhong, Hao Hu, Shenao Zhang, Sirui Zheng, Zhuoran Yang, Zhaoran Wang
Neural Information Processing Systems (NeurIPS) 2023 **Spotlight**
Extended version under review at Operations Research (OR)

[C4] **Pessimism in the Face of Confounders: Provably Efficient Offline Reinforcement Learning in Partially Observable Markov Decision Processes**
Miao Lu, Yifei Min, Zhaoran Wang, Zhuoran Yang
International Conference on Learning Representations (ICLR) 2023

[C3] **Welfare Maximization in Competitive Equilibrium: Reinforcement Learning for Markov Exchange Economy**
Zhihan Liu*, **Miao Lu***, Zhaoran Wang, Michael I. Jordan, Zhuoran Yang
International Conference on Machine Learning (ICML) 2022

[C2] **Learning Pruning-Friendly Networks via Frank-Wolfe: One-Shot, Any-Sparsity, and No Retraining**
Miao Lu*, Xiaolong Luo*, Tianlong Chen, Wuyang Chen, Dong Liu, Zhangyang Wang
International Conference on Learning Representations (ICLR) 2022 **Spotlight**

[C1] **Learning Robust Policy against Disturbance in Transition Dynamics via State-Conservative Policy Optimization**
Yufei Kuang, **Miao Lu**, Jie Wang, Qi Zhou, Bin Li, Houqiang Li
Association for Advancement of Artificial Intelligence (AAAI) 2022