

SHENAO ZHANG

shenao@u.northwestern.edu | shenao-zhang.github.io

EDUCATION

Northwestern University

Ph.D. student in IEMS (Industrial Engineering & Management Sciences)
Advisor: Prof. Zhaoran Wang

Sep. 2023 - Present

Evanston, IL

Georgia Institute of Technology

M.S. in ECE (Electrical and Computer Engineering), GPA: 3.81/4.00
Advisors: Prof. Tuo Zhao and Prof. Bo Dai

May 2020 - May 2022

Atlanta, GA

South China University of Technology

B.Eng. in EE (Electronic and Information Engineering, Innovation Class)

Aug. 2016 - May 2020

Guangzhou, China

University of California, Berkeley

Visiting student at the Department of EECS, GPA: 3.90/4.00

Jan. 2019 - May 2019

Berkeley, CA

RESEARCH INTERESTS

My research centers on **LLM** and **RL**, with a focus on **reasoning**, **alignment**, and **agents**. I develop models that learn strong foundations from offline data and self-improve via online interactions:

- **Building capable models from offline data:** RL algorithms that extract easier-to-learn hidden rationales [12] and achieve better credit assignment for reasoning and agents [11]; techniques that enhance RLHF by augmenting the offline data [10] and mitigating reward hacking [8].
- **Self-improvement via online exploration and adaptation:** A formal study of how to efficiently (or even correctly) scale test-time compute with exploration [13]; self-exploring language models [9]; LLM agents that quickly adapt by orchestrating reasoning and acting [7]; and RL algorithms for data-efficient exploration [2, 4, 5], long-horizon tasks [3, 6], and multi-agent systems [1].

PREPRINTS

- [13] **Beyond Markovian: Reflective Exploration via Bayes-Adaptive RL for LLM Reasoning.**
Shenao Zhang, Yaqing Wang, Yinxiao Liu, Tianqi Liu, Peter Grabowski, Eugene Ie, Zhaoran Wang[†], Yunxuan Li[†].
[We formally derive why, how, and when LLMs should self-reflect and explore at test time.](#)
Preprint, 2025.

PUBLICATIONS

- [12] **BRiTE: Bootstrapping Reinforced Thinking Process to Enhance LLM Reasoning.**
Han Zhong*, Yutong Yin*, **Shenao Zhang***, Xiaojun Xu*, Yuanxin Liu*, Yifei Zuo*, Zhihan Liu*, Boyi Liu, Sirui Zheng, Hongyi Guo, Liwei Wang, Mingyi Hong, Zhaoran Wang.
[A probabilistic framework that unifies previous LLM reasoning methods and unlocks new ones.](#)
International Conference on Machine Learning (ICML), 2025.
- [11] **Offline Reinforcement Learning for LLM Multi-Step Reasoning.**
Huajie Wang*, Shibo Hao*, Hanze Dong, **Shenao Zhang**, Yilin Bao, Ziran Yang, Yi Wu.
[An offline RL algorithm for LLM reasoning and language agents, adopted by Kimi k1.5.](#)
Findings of the Association for Computational Linguistics (ACL), 2025.
ICLR Workshop on Reasoning and Planning for LLMs (**Oral**), 2025.
- [10] **Reward-Augmented Data Enhances Direct Preference Alignment of LLMs.**
Shenao Zhang*, Zhihan Liu*, Boyi Liu, Yufeng Zhang, Yingxiang Yang, Liyu Chen, Tao Sun, Zhaoran Wang.
[A simple data augmentation method to enhance direct preference alignment algorithms.](#)
International Conference on Machine Learning (ICML), 2025.
- [9] **Self-Exploring Language Models: Active Preference Elicitation for Online Alignment.**
Shenao Zhang, Donghan Yu, Hiteshi Sharma, Ziyi Yang, Shuohang Wang, Hany Hassan, Zhaoran Wang.
[The first algorithm for LLMs to self-explore and self-improve during online RLHF.](#)
Transactions on Machine Learning Research (TMLR).
ICML Workshop on AutoRL (**Best Paper Award**), 2024.

- [8] **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.
[We show that adding SFT loss mitigates RLHF reward hacking, adopted by Llama 3 and Nemotron 4.](#)
Neural Information Processing Systems (NeurIPS), 2024.
- [7] **Reason for Future, Act for Now: A Principled Framework for Autonomous LLM Agents with Provable Sample Efficiency.**
 Zhihan Liu*, Hao Hu*, **Shenao Zhang***, Hongyi Guo, Shuqi Ke, Boyi Liu, Zhaoran Wang.
[The first provably efficient framework to orchestrate reasoning and acting for LLM agents.](#)
International Conference on Machine Learning (ICML), 2024.
- [6] **Adaptive-Gradient Policy Optimization: Enhancing Policy Learning in Non-Smooth Differentiable Simulations.**
 Feng Gao*, Liangzhi Shi*, **Shenao Zhang**, Zhaoran Wang, Yi Wu.
[An adaptive policy gradient method for variance reduction in long-horizon tasks.](#)
International Conference on Machine Learning (ICML), 2024.
- [5] **Model-Based Reparameterization Policy Gradient: Theory and Practical Algorithms.**
Shenao Zhang, Boyi Liu, Zhaoran Wang[†], Tuo Zhao[†].
[We analyze first-order policy gradients, obtained by differentiating through policy, dynamics, and reward.](#)
Neural Information Processing Systems (NeurIPS), 2023.
- [4] **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.
[A simple RL objective that integrates estimation and planning for sample-efficient exploration.](#)
Neural Information Processing Systems (NeurIPS) (Spotlight), 2023.
- [3] **Adaptive Barrier Smoothing for First-Order Policy Gradient with Contact Dynamics.**
Shenao Zhang, Wanxin Jin, Zhaoran Wang.
[A smoothing technique for RL policy gradients that balances the bias-variance tradeoff.](#)
International Conference on Machine Learning (ICML), 2023.
- [2] **Conservative Dual Policy Optimization for Efficient Model-Based Reinforcement Learning.**
Shenao Zhang.
[A theoretically and practically sample-efficient exploration algorithm for model-based RL.](#)
Neural Information Processing Systems (NeurIPS), 2022.
- [1] **Learning Meta Representation for Agents in Multi-Agent Reinforcement Learning.**
Shenao Zhang, Li Shen, Lei Han, Li Shen.
[A meta-RL algorithm that enables agents to quickly adapt to new multi-agent environments.](#)
Conference on Lifelong Learning Agents (CoLLAs) (Oral), 2023.

INTERNSHIP EXPERIENCE

Google	<i>Dec. 2024 - May 2025</i>
<i>Student Researcher</i>	Advisors: Eric Li, Yaqing Wang, Canoe Liu, and Tianqi Liu
• Worked on test-time exploration and Bayes-adaptive RL for reflective reasoning [13].	
Microsoft GenAI	<i>Jan. 2024 - June 2024</i>
<i>Student Researcher</i>	Advisor: Donghan Yu
• Worked on active preference elicitation for online alignment [9].	
ByteDance Seed	<i>June 2024 - Sep. 2024</i>
<i>Research Intern</i>	<i>June 2023 - Aug. 2023</i>
• Worked on RL with LLM policy prior [*] and reward-augmented alignment [10].	
Microsoft Research	<i>Feb. 2023 - May 2023</i>
<i>Research Intern</i>	Advisor: Li Zhao
• Worked on autonomous LLM agents that actively gather information [*].	
Tencent AI Lab	<i>Aug. 2019 - Sep. 2020</i>
<i>Research Intern</i>	Advisors: Li Shen, Lei Han and Li Shen
• Worked on visual attention representation [*] and multi-agent RL [1].	

TEACHING EXPERIENCE

Head TA of the graduate course [CS 7648: Interactive Robot Learning](#) (Fall 2021) at Georgia Tech.

PROFESSIONAL SERVICE

Conference Review: NeurIPS 20-25, ICLR 22-25, ICML 22-25, AISTATS 22-25, COLM 24-25.

Journal Review: Neurocomputing, TPAMI, TMLR.

HONORS AND AWARDS

Meshy Fellowship Finalist	<i>2025</i>
NeurIPS Top Reviewer	<i>2024</i>
NeurIPS Scholar Award	<i>2022-2023</i>
ICML Travel Award	<i>2023</i>
Georgia Tech Level A Premier Merit-Based Scholarship	<i>2020-2021</i>
Outstanding Freshman Scholarship (Awarded to 30 among 6,500 students)	<i>2016</i>