WEI XIONG

Computer Science, University of Illinois Urbana-Champaign
wx13@illinois.edu | Website | GitHub

RESEARCH INTERESTS

My research interests focus on reinforcement learning from human feedback (RLHF) for aligning large language model recently. I also spend time on the mathematical foundation of RL, where I am fortunate to collaborate with many great senior mentors and talented peers. I also spent time on deep RL at Microsoft Research.

EDUCATION

University of Illinois Urbana-Champaign *Urbana, USA*
PhD student, Computer Science, GPA: 4.0/4.0 *2023.8 - present*
Advisor: Prof. Tong Zhang and Prof. Nan Jiang

The Hong Kong University of Science and Technology *Hong Kong*
Master of Philosophy, Mathematics *2023.8*
Advisor: Prof. Tong Zhang

University of Science and Technology of China *Hefei, China*
Bachelor of Science, Mathematics & Electronic Engineering *2021.6*
Ranking: 1/72 in Statistics; 2/352 in EE.

EXPERIENCE

Student Researcher of Google Deepmind: *2024.5-present*
Host Dr. Tianqi Liu and Dr. Bilal Piot
Worked on building math agent by multi-turn iterative preference learning algorithm. Developed the multi-turn version of direct preference learning algorithms when the agent is allowed to call tools and make decision based on both the self-decoded tokens and external messages.

University of Illinois Urbana-Champaign: *2024.2-present*
Advisor Prof. Tong Zhang and Prof. Nan Jiang
Worked as the lead of the RLHFlow project, which presents a full recipe for the worklow of online iterative RLHF, including SFT, reward/preference modeling, and iterative RLHF/DPO. The resulting reward/preference functions are the state-of-the-art open-source models, and the final LLM achieves comparable or even better performance compared to LLaMA3-8B-instruct. The code, reward function, and model have contributed to > 30 research projects since then.

The Hong Kong University of Science and Technology: *2023.1-present*
Advisor Prof. Tong Zhang
Worked as a core founding member of the LMFlow project, which allows developing LLMs (fine-tuning, inference, RLHF...) with minimal cost and effort. The project received 8K+ star in github and ranked 2nd in the github trend. I am responsible for developing the RLHF part of the project.

Microsoft Research: *Spring 2021*
Advisor: Dr. Wenxue Cheng and Dr. Li Zhao
Intern: Networking Research and Machine Learning Group
Worked on bandwidth estimation for real-time communications with reinforcement learning.

SELECTED AWARDS AND FELLOWSHIPS

Best Paper Award in Demo Track, NAACL 2024	*2024*
Hong Kong PhD Fellowship	*2021-2023*
Best Teaching Assistant Award at HKUST	*June 2022,2023*
Outstanding graduate (USTC and Anhui province)	*June 2021*
Final list of Guo Moruo scholarship (highest honor of USTC)	*November 2020*
Yuanqing Yang Scholarship	*November 2020*
Chinese Academy of Sciences Institute of Electronics Scholarship	*October 2018*
National Scholarship	*October 2017*
Honor Program in EE/AI at USTC	*2017 - 2019*
Zhuang Caifang Scholarship	*July 2016*

SELECTED PUBLICATIONS AND MANUSCRIPTS

Also see full list in Google Scholar.

(α, β) denotes random/alphabetical order and * denotes equal contribution

[1] Wei Xiong, Hanning Zhang, Nan Jiang, Tong Zhang,"An Implementation of Generative PRM", [Code].

[2] Wei Xiong, Chengshuai Shi, Jiaming Shen, Aviv Rosenberg, Zhen Qin, Daniele Calandriello, Misha Khalman, Rishabh Joshi, Bilal Piot, Mohammad Saleh, Chi Jin, Tong Zhang, Tianqi Liu, "Building Math Agent by Iterative Preference Learning", [Preprint] [Code].

[3] Haoxiang Wang*, Wei Xiong*, Tengyang Xie, Han Zhao, Tong Zhang, "Interpretable Preferences via Multi-Objective Reward Modeling and Mixture-of-Experts", [EMNLP 2024] [Code].

[4] (α, β) Hanze Dong*, Wei Xiong*, Bo Pang*, Haoxiang Wang*, Han Zhao, Yingbo Zhou, Nan Jiang, Doyen Sahoo, Caiming Xiong, Tong Zhang, "RLHF Workflow: From Reward Modeling to Online RLHF", [Transactions on Machine Learning Research (TMLR)] [Code].

[5] Wei Xiong*, Hanze Dong*, Chenlu Ye*, Ziqi Wang, Han Zhong, Heng Ji, Nan Jiang, Tong Zhang, "Iterative Preference Learning from Human Feedback: Bridging Theory and Practice for RLHF under KL-Constraint", [ICML 2024] [Code].

[6] Haoxiang Wang*, Yong Lin*, Wei Xiong*, Rui Yang, Shizhe Diao, Shuang Qiu, Han Zhao, Tong Zhang, "Arithmetic Control of LLMs for Diverse User Preferences: Directional Preference Alignment with Multi-Objective Rewards", [ACL 2024] [Code].

[7] (α, β) Chenlu Ye*, Wei Xiong*, Yuheng Zhang*, Hanze Dong*, Nan Jiang, Tong Zhang, "A Theoretical Analysis of Nash Learning from Human Feedback under General KL-Regularized Preference", [NeurIPS 2024].

[8] (α, β) Yong Lin*, Hangyu Lin*, Wei Xiong*, Shizhe Diao*, Jianmeng Liu, Jipeng Zhang, Rui Pan, Haoxiang Wang, Wenbin Hu, Hanning Zhang, Hanze Dong, Renjie Pi, Han Zhao, Nan Jiang, Yuan Yao, Heng Ji, and Tong Zhang, "Mitigating the Alignment Tax of RLHF", [EMNLP 2024].

[9] (α, β) Hanze Dong*, Wei Xiong*, Deepanshu Goyal, Rui Pan, Shizhe Diao, Jipeng Zhang, Kashun Shum and Tong Zhang, "RAFT: Reward rAnked FineTuning for Generative Foundation Model Alignment" [Transactions on Machine Learning Research (TMLR)] [Code].

[10] Shizhe Diao*, Rui Pan*, Hanze Dong*, KaShun Shen, Jipeng Zhang, Wei Xiong, and Tong Zhang, "LMFlow: An Extensible Toolkit for Finetuning and Inference of Large Foundation Models", [NAACL 2024, Best Paper Award in Demo Track] [Code].

[11] (α, β) Han Zhong*, Wei Xiong*, Sirui Zheng, Liwei Wang, Zhaoran Wang, Zhuoran Yang, and Tong Zhang, "GEC: A Unified Framework for Interactive Decision Making in MDP, POMDP, and Beyond", [Under Major Revision at Mathematical Operation Research (MOR)] [Slide].

[12] Wei Xiong*, Han Zhong*, Chengshuai Shi, Cong Shen, Liwei Wang, and Tong Zhang,"Nearly Minimax Optimal Offline Reinforcement Learning with Linear Function Approximation: Single-Agent MDP and Markov Game", [ICLR 2023].

[13] Wei Xiong, Han Zhong, Chengshuai Shi, Cong Shen, and Tong Zhang,"A Self-Play Posterior Sampling Algorithm for Zero-Sum Markov Game", [ICML 2022].

TEACHING

The Hong Kong University of Science and Technology: *2021.9-2023.6*
Teaching Assistant: MATH 2421 Probability, MATH 2121 Linear Algebra, MATH 6913W Reading Course: Statistical Learning Theory, MATH 2023 Multivariable Calculus (**Best TA Award for all courses**).

University of Science and Technology of China: *2018-2021*
Teaching Assistant: Mathematical Statistics, Data Structures and Databases, Algorithms and Data Structures.

PROFESSIONAL ACTIVITY

Conference Reviewer:
ICLR 2024; NeurIPS 2022, 2023 (**Top Reviewer Award**), 2024; ICML 2022, 2023; AISTATS 2023, 2024, 2025, ARR 2024.

Journal Reviewer:
Machine Learning, JMLR, TMLR.

Talks:
TMLR Young Scientist Seminar Sharing: *2024.11*
Building Math Agent by Iterative Preference Learning Host: Yongqiang Chen

University of Virginia: *2024.11*
Guest Lecture: Iterative Preference Learning for LLM Post Training Host: Yu Meng

Amazon: *2024.11*
Building Math Agent by Iterative Preference Learning Host: Changlong Yu

Informs Annual Meeting: *2024.10*
Building Math Agent by Iterative Preference Learning Host: Ming Yin and Yingru Li

University of Illinois Urbana-Champaign, NLP Seminar: *2024.10*
Building Math Agent by Iterative Preference Learning Host: Chi Han

Simons Institute for the Theory of Computing: *2024.9*
Iterative Preference Learning for Large Language Model Post Training Host: Andrej Risteski

University of Illinois Urbana-Champaign, Machine Learning Seminar: *2024.9*
Iterative Preference Learning for Large Language Model Post Training Host: Rohan Deb

University of Waterloo: *2024.8*
Iterative Preference Learning for Large Language Model Post Training Host: Wenhu Chen

Google Deepmind Sky Team: *2024.8*
Building Math Agent by Iterative Preference Learning, Host: Tianqi Liu

University of Illinois Urbana-Champaign Blender Lab: *2024.7*
Reinforcement Learning from Human Feedback: From Theory to Algorithm, Host: Heng Ji

Mila Alignment Seminar: *2024.7*
Reinforcement Learning from Human Feedback: From Theory to Algorithm, Host: Emiliano Penaloza

Google Multi-turn RLHF Workshop, MTV: *2024.6*
Reinforcement Learning from Human Feedback: From Theory to Algorithm, Host: Lior Shani

Google Learning Theory Seminar, NYC: *2024.5*
Reinforcement Learning from Human Feedback: From Theory to Algorithm, Host: Jacob Abernethy

Center for Machine Learning Research, Peking University: *2024.4*
Reinforcement Learning from Human Feedback: From Theory to Algorithm.

University of Virginia: *2024.4*
Reinforcement Learning from Human Feedback: From Theory to Algorithm, Host: Cong Shen

Yale University, Department of Statistics and Data Science: *2024.3*
Alignment for Foundation Language Models: Mathematical Principle and Algorithmic Designs, Host: Zhuoran Yang

UCLA AGI Lab: *2024.2*
Alignment for Foundation Language Models: Mathematical Principle and Algorithmic Designs, Host: Quanquan Gu

Microsoft Research Asia Machine Learning Group: *2024.1*
Alignment for Foundation Language Models: Mathematical Principle and Algorithmic Designs, Host: Chuheng Zhang

Hong Kong University Deep Vision Lab: *2023.8*
RLHF with Rejection Sampling: A RL-free Approach, Host: Qi Xiaojuan

University of Toronto: *2023.6*
RLHF with Rejection Sampling: A RL-free Approach, Host: Qiang Sun

Stanford University: *2023.5*
RLHF with Rejection Sampling: A RL-free Approach, Host: Mert Pilanci