

WEI XIONG

Computer Science, University of Illinois Urbana-Champaign

wx13@illinois.edu | [Website](#) | [GitHub](#)

RESEARCH INTERESTS

Machine learning theory and alignment for large language model.

EDUCATION

University of Illinois Urbana-Champaign

PhD student, *Department of Computer Science*

Advisor: Prof. Tong Zhang and Prof. Nan Jiang

Urbana, USA

2023.8 - present

The Hong Kong University of Science and Technology

Master of Philosophy, *Department of Mathematics*

Advisor: Prof. Tong Zhang

Hong Kong, China

2023.8

University of Science and Technology of China

Bachelor of Science, *Department of Mathematics*

Department of Electronic Engineering

ranking: 1/72 in Statistics

Hefei, China

2021.6

Shanghai Jiao Tong University

Exchange student at School of Electronic Information and Electrical Engineering

Shanghai, China

2018

EXPERIENCE

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 is a framework that allows developing LLMs (fine-tuning, inference, RLHF...) with minimal cost and effort. The project received 7K+ star in github and ranked 2nd in the github trend. I am responsible for developing the RLHF part of the project.

Yale University:

2022 Spring

Virtual visit with Prof. Zhuoran Yang

Worked on the mathematical foundation of reinforcement learning.

Microsoft Research Asia (MSRA):

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.

University of Virginia:

2019.8 - 2019.11

Advisor: Prof. Cong Shen

Research Assistant: worked on multi-player multi-armed bandit (MPMAB).

SELECTED AWARDS AND FELLOWSHIPS

Hong Kong PhD Fellowship

2021-2023

Best Teaching Assistant Award at HKUST

June 2022, 2023

Outstanding graduate (USTC and Anhui province)

June 2021

Nomination for Guo Moruo scholarship (1/72 in statistics, highest honor of USTC)	<i>November 2020</i>
Yuanqing Yang Scholarship	<i>November 2020</i>
Chinese Academy of Sciences Institute of Electronics Scholarship	<i>October 2018</i>
National Scholarship	<i>October 2017</i>
Honor Program in EE/AI at USTC	<i>2017 - 2019</i>
Zhuang Caifang Scholarship	<i>July 2016</i>

SELECTED PUBLICATIONS AND MANUSCRIPTS

See full list in [Google Scholar](#).

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

- [1] Wei Xiong*, Hanze Dong*, Chenlu Ye*, Han Zhong, Nan Jiang, Tong Zhang, “Gibbs Sampling from Human Feedback: A Provable KL-constrained Framework for RLHF”, we formulate the real-world RLHF process as a reverse-KL regularized contextual bandits and study its theoretical property by proposing statistically efficient algorithms with finite-sample theoretical guarantee. We also demonstrate that the proposed theoretical algorithms can be practically implemented and empirically outperform existing strong baselines in real-world LLM alignment experiments [\[Preprint\]](#).
- [2] (α, β) Yong Lin*, Hangyu Lin*, Wei Xiong*, Shizhe Diao*, Jianmeng Liu, Jipeng Zhang, Hanning Zhang, Hanze Dong, Haoxiang Wang, Han Zhao, Nan Jiang, Yuan Yao, Heng Ji, and Tong Zhang, “Reward vs Tax: Investigating and Alleviating Alignment Tax of RLHF”, we quantitatively study the alignment tax of popular RLHF algorithms including PPO, RAFT, and DPO and investigate various methods to alleviate the forgetting, including regularization, low-rank finetuning, data replay, reward regularization, and model averaging. We also propose adaptive model averaging, which is most competitive across different tasks of the benchmark.
- [3] (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”, we develop a simple but effective alignment framework with minimal hyper-parameter configuration; the proposed framework is friendly in implementation and memory resources, and also interpretable with clear learning objectives. [\[TMLR\]](#) [\[Code\]](#).
- [4] 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”, we introduce an extensible and lightweight toolkit, LMFlow, which aims to simplify the development of general LLMs; the project received 7K+ GitHub stars and I was responsible for developing the RLHF part of the whole project. [\[Submitted to JMLR\]](#) [\[Code\]](#).
- [5] (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”, we measure the hardness of the sequential decision making problem as the coefficient to generalize in the online manner and show that the online problems in this framework can be reduced to an offline supervised learning in terms of in-sample error estimation; the proposed framework captures most of known trackable MAB, Contextual Banit, MDP and POMDP problems; a generalized posterior sampling framework is also provided. [\[Submitted to Mathematical Operation Research \(MOR\)\]](#) [\[Slide\]](#).
- [6] Wei Xiong, “A Sufficient Condition of Sample-Efficient Reinforcement Learning with General Function Approximation”, *Master Thesis*, we develop GEC in this thesis with a thorough description of the motivation and application; we also develop a new optimization-based framework, as a counterpart of the sampling framework in original GEC paper.
- [7] 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”, An application of weighted regression by using the variance information to achieve a sharper bound. , [\[ICLR 2023\]](#).
- [8] Wei Xiong, Han Zhong, Chengshuai Shi, Cong Shen, and Tong Zhang, “A Self-Play Posterior Sampling Algorithm for Zero-Sum Markov Game”, this is a straightforward extension of the single-agent conditional

posterior sampling, which also provides an extension of the eluder coefficient to the multi-agent case, [ICML 2022].

- [9] Han Zhong*, Wei Xiong*, Jiyuan Tan*, Liwei Wang, Tong Zhang, Zhaoran Wang, and Zhuoran Yang, “Pessimistic Minimax Value Iteration: Provably Efficient Equilibrium Learning from Offline Datasets”, [ICML 2022] [Slide].
- [10] Chengshuai Shi, Wei Xiong, Cong Shen, and Jing Yang, “Heterogeneous Multi-player Multi-armed Bandits: Closing the Gap and Generalization”, we developed a carefully-crafted exploration strategy in the heterogeneous MPMAB setting, as well as a delicate differential communication scheme; the proposed BEACON achieves the minimax optimal regret bound and also demonstrates an impressive empirical performance. [NeurIPS 2021] [Code].

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**); ICML 2022, 2023; AISTATS 2023, 2024.

Journal Reviewer:

Machine Learning, JMLR.

Invited Talks:

Microsoft Research Asia, host: Chuheng Zhang *2024.1*

Hong Kong University, host: Qi Xiaojuan *2023.8*

University of Toronto, host: Qiang Sun *2023.6*

Stanford University, host: Mert Pilanci *2023.5*