CHENLU YE

Computer Science, University of Illinois Urbana-Champaign
chenluy3@illinois.edu | Website | Scholar

RESEARCH INTERESTS

- Reinforcement learning for reasoning and agent tool-using in LLM post-training;
- Decision-making problems;

EDUCATION

University of Illinois Urbana-Champaign *Urbana, USA*
Ph.D. student, *Computer Science* *2024.08 - present*
Advisor: Prof. Tong Zhang

The Hong Kong University of Science and Technology *Hong Kong, China*
MPhil, *Artificial Intelligence* *2021.09 - 2024.08*
Advisor: Prof. Tong Zhang

University of Science and Technology of China *Hefei, China*
Bachelor of Science, *Statistics* *2017.09 - 2021.06*

RESEARCH AND EXPERIENCE

Applied Scientist Intern, Amazon *2025.05 - Present*
Hosts: Dr. Yu Zhou, Dr. Ziji Zhang
Leader of an RL reasoning framework that robustly applies PRMs via consistency-oriented filtration:

- Proposed **PRocess cOnsistency Filtering (PROF)** to robustly integrate noisy Process Reward Models (PRMs) with Outcome Reward Models (ORMs) in RL, incorporating data consistency filtration and balancing the correct-incorrect ratio.
- Conducted extensive studies to demonstrate that PROF-GRPO not only increases the final outcome accuracy but also shapes the intermediate reasoning steps and improves the process reasoning quality.
- Conducted a series of ablation studies to illustrate the importance of separating the correct and incorrect responses during the filtration.

Ph.D. Student, University of Illinois Urbana-Champaign *2024.09 - Present*
Advisor: Prof. Tong Zhang
Core developer of several works across RLHF and reasoning tasks:

- Developing a general adaptive-sampling framework where the rollout number is tailored to the model's ability to solve it.
- Proposed a self-rewarding correction framework to enhance the policy model's ability to perform self-verification and correction for mathematical reasoning.
- Proposed online iterative RLHF and implemented it with online DPO, and then, we extended the framework to general preference settings.

Master, The Hong Kong University of Science and Technology *2021.9 - 2024.8*
Advisor:Prof. Tong Zhang

- Proposed algorithm designs in RLHF and formulated the real-world RLHF process as a reverse-KL regularized contextual bandits for preference satisfying BT model, respectively. We studied its

theoretical property by proposing statistically efficient algorithms with finite-sample theoretical guarantee.
- Developed a series of corruption-robust algorithms based on uncertainty weighting for online and offline, value-based and model-based settings.

Visiting Research Scholar, University of California, Los Angeles:
Host: Prof. Quanquan Gu *2023.8 - 2023.12*

- Proposed several RL algorithms robust to adversarial corruption for both online and offline decision-making processes.

SELECTED PUBLICATIONS AND PREPRINTS

(* denotes alphabetical order or equal contribution)

[1] Chenlu Ye, Zhou Yu, Ziji Zhang, Hao Chen, Narayanan Sadagopan, Jing Huang, Tong Zhang, Anurag Beniwalg, "Beyond Correctness: Harmonizing Process and Outcome Rewards through RL Training", [Preprint].

[2] Wei Xiong*, Hanning Zhang*, Chenlu Ye*, Lichang Chen, Nan Jiang, Tong Zhang, "Self-rewarding correction for mathematical reasoning", [Preprint].

[3] Chenlu Ye*, Wei Xiong*, Yuheng Zhang*, Hanze Dong*, Nan Jiang, Tong Zhang, "Online iterative reinforcement learning from human feedback with general preference model", [NeurIPS 2024].

[4] Wei Xiong*, Hanze Dong*, Chenlu Ye*, Han Zhong, Nan Jiang, Tong Zhang, "Iterative preference learning from human feedback: Bridging theory and practice for rlhf under kl-constraint", [ICML 2024]

[5] Heyang Zhao* Chenlu Ye*, Wei Xiong, Quanquan Gu, Tong Zhang, "Logarithmic Regret for Online KL-Regularized Reinforcement Learning", [ICML 2025].

[6] Chenlu Ye, Yujia Jin, Alekh Agarwal, Tong Zhang, "Catoni Contextual Bandits are Robust to Heavy-tailed Rewards", [Spotlight of ICML 2025].

[7] Yifan Hao*, Xingyuan Pan*, Hanning Zhang*, Chenlu Ye, Rui Pan, Tong Zhang, " Understanding Over-adaptation in Supervised Fine-Tuning: The Role of Ensemble Methods", [ICML 2025].

[8] Heyang Zhao Chenlu Ye, Quanquan Gu, Tong Zhang, "Sharp Analysis for KL-Regularized Contextual Bandits and RLHF", [NeurIPS 2025].

[9] Chenlu Ye*, Jiafan He*, Quanquan Gu, Tong Zhang, "Towards robust model-based reinforcement learning against adversarial corruption", [ICML 2024].

[10] Chenlu Ye*, Rui Yang*, Quanquan Gu and Tong Zhang, "Corruption-Robust Offline Reinforcement Learning with General Function Approximation", [NeurIPS 2023].

[11] Yong Lin*, Chen Liu*, Chenlu Ye*, Qing Lian, Yuan Yao and Tong Zhang,"Optimal Sample Selection Through Uncertainty Estimation and Its Application in Deep Learning", [JMLR].

[12] Chenlu Ye, Wei Xiong, Quanquan Gu, and Tong Zhang, "Corruption-Robust Algorithms with Uncertainty Weighting for Nonlinear Contextual Bandits and Markov Decision Processes", [ICML 2023].

[13] Jianqing Fan*, Zhaoran Wang*, Zhuoran Yang*, Chenlu Ye*, "Provably Efficient High-Dimensional Bandit Learning with Batched Feedbacks", [Preprint].

[14] Xingyuan Pan*, Chenlu Ye*, Joseph Melkonian, Jiaqi W. Ma, Tong Zhang, "Daunce: Data Attribution through Uncertainty Estimation", [Preprint].

[15] Yifan Hao*, Chenlu Ye*, Chi Han, Tong Zhang, "Transformers as Multi-task Learners: Decoupling Features in Hidden Markov Models", [Preprint].

PROFESSIONAL ACTIVITY

Conference Reviewer: ICML, NeurIPS, ICLR, AISTAT.

Journal Reviewer: JMLR, Machine Learning, Artificial Intelligence.