

CHENLU YE

Computer Science, University of Illinois Urbana-Champaign

chenluy3@illinois.edu | Website | Scholar

RESEARCH INTERESTS

- AI Alignment and Safety, Scalable Oversight (Process Reward Models)
- Reinforcement learning for reasoning and agent tool-using in LLM post-training;
- Decision-making problems;

EDUCATION

University of Illinois Urbana-Champaign

Ph.D. student, *Computer Science*

Advisor: Prof. Tong Zhang

Urbana, USA

2024.08 - present

The Hong Kong University of Science and Technology

MPhil, *Artificial Intelligence*

Advisor: Prof. Tong Zhang

Hong Kong, China

2021.09 - 2024.08

University of Science and Technology of China

Bachelor of Science, *Statistics*

Hefei, China

2017.09 - 2021.06

RESEARCH AND EXPERIENCE

Applied Scientist Intern, Amazon

2025.05 - Present

Hosts: [Dr. Yu Zhou](#), [Dr. Ziji Zhang](#)

Leader and main contributor of **Adaptive Layerwise Perturbation (ALP)** to make off-policy RL **robust to system biases and noises**, like train-inference mismatch [\[ARXIV\]](#):

- Added a small perturbation to all layers of policy models during training to improve the model's smoothness and robustness.
- Large scale experiments on Math and **multi-turn agent tasks**, such as TIR showed that ALP improves the performance by enhancing training stability and improving exploration.

Leader and main contributor of an RL reasoning framework that improve the **reliability and safety of reasoning process** [\[ARXIV\]](#):

- 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:

- Developed an **adaptive-sampling** framework that dynamically allocates inference budget across prompts for online RL post-training to avoid signal elimination and increase signal diversity. [\[ARXIV\]](#) [\[GITHUB\]](#)

- Proposed a **self-rewarding correction** framework to enhance the policy model’s ability to perform self-verification and correction for mathematical reasoning. [ARXIV]
- Proposed **online iterative RLHF** and implemented it with online DPO, and then, we extended the framework to general preference settings. [ARXIV] [GITHUB]

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.

SKILLS

Programming Languages and Tools: Python PyTorch(Expert), C++

Developer Tools: Git, Docker

HONORS AND AWARDS

Gold Prize for Outstanding Student Scholarship (1/40) 2020.9

Bronze Prize for Outstanding Student Scholarship 2019.9

Bronze Prize for Outstanding Student Scholarship 2018.9

SELECTED PUBLICATIONS AND PREPRINTS

(* denotes alphabetical order or equal contribution)

- [1] Wei Xiong*, Chenlu Ye*, Baohao Liao*, Hanze Dong*, Xinxing Xu, Christof Monz, Jiang Bian, Nan Jiang, Tong Zhang, “Reinforce-Ada: An Adaptive Sampling Framework for Reinforce-Style LLM Training”, [Preprint].
- [2] 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].
- [3] Wei Xiong*, Hanning Zhang*, Chenlu Ye*, Lichang Chen, Nan Jiang, Tong Zhang, “Self-rewarding correction for mathematical reasoning”, [Preprint].
- [4] 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].
- [5] 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]
- [6] Heyang Zhao* Chenlu Ye*, Wei Xiong, Quanquan Gu, Tong Zhang, “Logarithmic Regret for Online KL-Regularized Reinforcement Learning”, [ICML 2025].
- [7] Chenlu Ye, Yujia Jin, Alekh Agarwal, Tong Zhang, “Catoni Contextual Bandits are Robust to Heavy-tailed Rewards”, [Spotlight of ICML 2025].

- [8] 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](#)].
- [9] Heyang Zhao Chenlu Ye, Quanquan Gu, Tong Zhang, “Sharp Analysis for KL-Regularized Contextual Bandits and RLHF”, [[NeurIPS 2025](#)].
- [10] Chenlu Ye*, Jiafan He*, Quanquan Gu, Tong Zhang, “Towards robust model-based reinforcement learning against adversarial corruption”, [[ICML 2024](#)].
- [11] Chenlu Ye*, Rui Yang*, Quanquan Gu and Tong Zhang, “Corruption-Robust Offline Reinforcement Learning with General Function Approximation”, [[NeurIPS 2023](#)].
- [12] 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](#)].
- [13] 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](#)].
- [14] Jianqing Fan*, Zhaoran Wang*, Zhuoran Yang*, Chenlu Ye*, “Provably Efficient High-Dimensional Bandit Learning with Batched Feedbacks”, [[Preprint](#)].
- [15] Xingyuan Pan*, Chenlu Ye*, Joseph Melkonian, Jiaqi W. Ma, Tong Zhang, “Daunce: Data Attribution through Uncertainty Estimation”, [[Preprint](#)].
- [16] 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.