Yue Wu

✉ frankwupku@gmail.com ☎ (424)440-9841 🌐 http://yuewu.us

Industrial Experience

2025.4-2026.1 **xAI**, Palo Alto, California *Member of Technical Staff, Reasoning*
◇ Co-led RL-Science team, spanning RL scaling, RL recipe design (multi-agent RL, CoT compression, continual learning), and RL fundamentals (stability & exploration).

2024 **Meta**, Bellevue, Washington *Research Scientist Intern*, Gen AI
◇ Worked on token-level reward modeling and new architecture design for general human preference and general preference optimization.

2023 **Bytedance AI Lab**, Los Angeles, California. *Research Scientist Intern*, Drug Discovery
◇ Worked on multi-conformation generation of large protein molecules. Incorporated physical priors of molecular dynamics into diffusion-based generative models. The paper is accepted in ICML 2024.

2022 **NEC Laboratories America**, Princeton, New Jersey
Research Intern, Data Science and System Security
◇ Worked on personalized federated learning and developed a method based on mixture models. The paper is accepted in ICML 2023.

Education & Academic Employment

2024 – Present **Princeton University AI Lab**, Princeton, New Jersey.
Postdoctoral Research Fellow

2019 – 2024 **University of California, Los Angeles**, Westwood, California.
Doctor of Philosophy in Computer Science
Advisor: Quanquan Gu
Dissertation Committee: Quanquan Gu, Guy Van den Broeck, Lieven Vandenberghe, Aditya Grover, Mengdi Wang

2015 – 2019 **Peking University**, Beijing, China.
Bachelor of Science in Machine Intelligence
GPA: 3.83/4.00, Rank: 1/53, Summa Cum Laude.
Thesis Advisor: Liwei Wang

Highlighted Projects

2024.4 **SPPO: Self-Play Preference Optimization for LLM Alignment**
Propose to directly align LLM with the preference instead of using an approximate reward model such as Bradley-Terry, and a new RL-based learning objective to maximize the probability of being preferred. Design principled self-play training framework and approximate solution based on iterative fine-tuning on synthetic data generated by the reference model.

2024.9 **General Preference Model with Preference Representations**
Propose a more principled, efficient way of modeling general preferences instead of the ad-hoc pairwise prompting and prediction. The new method can efficiently predict preference among a group of text completions and be further utilized to align LLMs.

Publications and Preprints

Qiu, J., Lu, Y., Zeng, Y., Guo, J., Geng, J., Wang, H., Huang, K., **Wu, Y.**, & Wang, M. (2024). Treebon: Enhancing inference-time alignment with speculative tree-search and best-of-n sampling. *arXiv preprint arXiv:2410.16033*.

Wang, Y., Wang, L., Shen, Y., Wang, Y., Yuan, H., **Wu, Y.**, & Gu, Q. (2024). Protein conformation generation via force-guided se (3) diffusion models. *Proceedings of the 40th International Conference on Machine Learning* (**ICML 2024**).

Wu, Y., Jin, T., Di, Q., Lou, H., Farnoud, F., & Gu, Q. (2024). Borda regret minimization for generalized linear dueling bandits. *Proceedings of the 40th International Conference on Machine Learning* (**ICML 2024**).

Wu, Y., Sun, Z., Yuan, H., Ji, K., Yang, Y., & Gu, Q. (2024). Self-play preference optimization for language model alignment. *arXiv preprint arXiv:2405.00675*.

Yuan*, H., Zeng*, Y., **Wu*, Y.**, Wang, H., Wang, M., & Leqi, L. (2024). A common pitfall of margin-based language model alignment: Gradient entanglement. *arXiv preprint arXiv:2410.13828*.

Zhang*, Y., Zhang*, G., **Wu*, Y.**, Xu, K., & Gu, Q. (2024). General preference modeling with preference representations for aligning language models. *https://arxiv.org/abs/2410.02197*.

Di, Q., Jin, T., **Wu, Y.**, Zhao, H., Farnoud, F., & Gu, Q. (2023). Variance-aware regret bounds for stochastic contextual dueling bandits. *International Conference on Learning Representations* (**ICLR 2024**).

Wu, Y., He, J., & Gu, Q. (2023). Uniform-PAC guarantees for model-based RL with bounded eluder dimension. *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence* (**UAI 2023**), 2304–2313.

Wu, Y., Zhang, S., Yu, W., Liu, Y., Gu, Q., Zhou, D., Chen, H., & Cheng, W. (2023). Personalized federated learning under mixture of distributions. *Proceedings of the 40th International Conference on Machine Learning* (**ICML 2023**).

Xiao, Y., Jin, Y., Bai, Y., **Wu, Y.**, Yang, X., Luo, X., Yu, W., Zhao, X., Liu, Y., Chen, H., et al. (2023). Large language models can be good privacy protection learners. *arXiv preprint arXiv:2310.02469*.

Yang, X., Cheng, W., **Wu, Y.**, Petzold, L., Wang, W. Y., & Chen, H. (2023). Dna-gpt: Divergent n-gram analysis for training-free detection of gpt-generated text. *International Conference on Learning Representations Proceedings of the 40th International Conference on Machine Learning* (**ICLR 2024**).

Chen, Z., Deng, Y., **Wu, Y.**, Gu, Q., & Li, Y. (2022). Towards understanding the mixture-of-experts layer in deep learning. *Advances in neural information processing systems* (**NeurIPS 2022**).

Lou, H., Jin, T., **Wu, Y.**, Xu, P., Gu, Q., & Farnoud, F. (2022). Active ranking without strong stochastic transitivity. *Advances in neural information processing systems* (**NeurIPS 2022**), *35*, 297–309.

Wu, Y., Jin, T., Lou, H., Xu, P., Farnoud, F., & Gu, Q. (2022). Adaptive sampling for heterogeneous rank aggregation from noisy pairwise comparisons. *International Conference on Artificial Intelligence and Statistics* (**AISTATS 2022**), 11014–11036.

Wu, Y., Zhou, D., & Gu, Q. (2022). Nearly minimax optimal regret for learning infinite-horizon average-reward mdps with linear function approximation. *International Conference on Artificial Intelligence and Statistics* (**AISTATS 2022**).

Cao, Y., Fang, Z., **Wu, Y.**, Zhou, D.-X., & Gu, Q. (2021). Towards understanding the spectral bias of deep learning. *International Joint Conference on Artificial Intelligence* (**IJCAI 2021**).

Wu, Y., Zhang, W., Xu, P., & Gu, Q. (2020). A finite-time analysis of two time-scale actor-critic methods. *Advances in Neural Information Processing Systems* (**NeurIPS 2020**).

Wang, L., Hu, L., Gu, J., **Wu, Y.**, Hu, Z., He, K., & Hopcroft, J. (2018). Towards understanding learning representations: To what extent do different neural networks learn the same representation. *Advances in neural information processing systems* (**NeurIPS 2018**).

Honors and Awards

2023	**Dissertation Year Fellowship**, University of Calofornia, Los Angeles.
2017	**China National Scholarship**, Peking University.
2016	**Founder Scholarship**, Peking University.