Yue Wu

Department of Computer Science,
University of California, Los Angeles

✉ ywu@cs.ucla.edu   ☎ (424)440-9841   🌐 http://yuewu.us

Education

2019 – Present   ▉ **University of California, Los Angeles**, Westwood, California.
Doctor of Philosophy in Computer Science
Thesis Advisor: Quanquan Gu

2015 – 2019   ▉ **Peking University**, Beijing, China.
Bachelor of Science in Machine Intelligence
Thesis Advisor: Liwei Wang

Research Interest

▉ My research agenda revolves around AI alignment with human feedback and aims to develop efficient and trustworthy alignment approaches, that are motivated by real-world applications, yield new theoretical insights, and demonstrate tangible practical impacts. I work on designing principled and efficient methods for **preference learning** and **reinforcement learning**. I also work on **trustworthy machine learning** including federated learning and privacy protection.

Honors and Awards

2023   ▉ **Dissertation Year Fellowship**, University of Calofornia, Los Angeles.

2017   ▉ **China National Scholarship**, Peking University.

2016   ▉ **Founder Scholarship**, Peking University.

Publications and Preprints

Di, Q., Jin, T., **Wu**, **Y.**, Zhao, H., Farnoud, F., & Gu, Q. (2023). Variance-aware regret bounds for stochastic contextual dueling bandits. *arXiv preprint arXiv:2310.00968*.

Wu, **Y.**, He, J., & Gu, Q. (2023, 31 Jul–04 Aug). Uniform-PAC guarantees for model-based RL with bounded eluder dimension. In R. J. Evans & I. Shpitser (Eds.), *Proceedings of the thirty-ninth conference on uncertainty in artificial intelligence* (pp. 2304–2313, Vol. 216). PMLR.

Wu, **Y.**, Jin, T., Di, Q., Lou, H., Farnoud, F., & Gu, Q. (2023). Borda regret minimization for generalized linear dueling bandits. *ICML 2023 Workshop The Many Facets of Preference-Based Learning*.

Wu, **Y.**, Zhang, S., Yu, W., Liu, Y., Gu, Q., Zhou, D., Chen, H., & Cheng, W. (2023, 23–29 Jul). Personalized federated learning under mixture of distributions. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.), *Proceedings of the 40th international conference on machine learning* (pp. 37860–37879, Vol. 202). PMLR.

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.

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*, *35*, 23049–23062.

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*, *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*, 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*, 3883–3913.

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

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*, *33*, 17617–17628.

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*, *31*.

Academic Services

Reviewing

2020 – present	▌ ICML, reviewer
	▌ NeurIPS, reviewer
	▌ ICLR, reviewer
	▌ AISTATS, reviewer
2022	▌ AAAI, Senior PC member

Teaching Experience

Winter 2021,22,23 ▌ **UCLA CS 161: Fundamental of Artificial Intelligence**
Teaching Assistant
Re-formulated the course homework and projects, as well as prepared and graded mid-term and final exams.

Spring 2023 ▌ **UCLA CS 31: Introduction to Computer Science**
Teaching Assistant

Fall 2020 ▌ **UCLA CS M51A: Logic Design of Digital Systems**
Teaching Assistant

Professional Experience

2023   ▉   **Bytedance AI Lab**, Los Angeles, California.
Research Scientist Intern, Drug Discovery
Working on multi-conformation generation of large protein molecules. Utilizing diffusion models and molecular dynamics.

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.