

YIBIN WANG

✉ yibinwang@gmail.com · 🎓 [Google Scholar](#) · 💻 yibinwang.netlify.app · 🌐 [Github](#)

Research Interests: My research interests focus on **LLMs and reinforcement learning**, particularly in the areas of generalization and calibration.

🎓 EDUCATION

Huazhong University of Science and Technology (HUST) Sept. 2019 – June. 2024

B.E. in Computer Science (CS) (Excellent Class), GPA: 3.82/4.00 - [Transcripts](#)

I got injured and took a one-year leave of absence from school in 2019.

████ EXPERIENCE

Bayesian LLMs & Efficient LLMs RL Fine-tuning June. 2024 – June. 2025

Research Intern | University of Illinois Urbana-Champaign (UIUC)

Advised by Prof. Huan Zhang

Generalization, Calibration and Robustness of LLM Sept. 2023 – May. 2024

Remote Research Intern | Rutgers Machine Learning Lab, Rutgers University

Advised by Prof. Hao Wang

Certified Adversarial Robustness in NLP Sept. 2021 – Aug. 2023

Research Intern | John Hopcroft Lab for Data Science, HUST

Advised by Prof. Kun He

📄 PUBLICATIONS

* indicates equal contribution

Robustness-Aware Word Embedding Improves Certified Robustness to Adversarial Word Substitutions Findings of ACL 2023

- Yibin Wang*, Yichen Yang*, Di He, Kun He

Continual Learning of Large Language Models: A Comprehensive Survey

ACM Computing Surveys

- Haizhou Shi, Zihao Xu, Hengyi Wang, Weiyi Qin, Wenyuan Wang, Yibin Wang, Zifeng Wang, Sayna Ebrahimi, Hao Wang

BLoB: Bayesian Low-Rank Adaptation by Backpropagation for Large Language Models

NeurIPS 2024

- Yibin Wang*, Haizhou Shi*, Ligong Han, Dimitris Metaxas, Hao Wang

Training-Free Bayesianization for Low-Rank Adapters of Large Language Models

NeurIPS 2025

- Haizhou Shi*, Yibin Wang*, Ligong Han, Huan Zhang, Hao Wang

Improving Data Efficiency for LLM Reinforcement Fine-tuning Through Difficulty-targeted Online Data Selection and Rollout Replay NeurIPS 2025

- Yifan Sun*, Jingyan Shen*, Yibin Wang*, Tianyu Chen, Zhendong Wang, Mingyuan Zhou, Huan Zhang

Token-Level Uncertainty Estimation for Large Language Model Reasoning

ICLR 2026

- Tunyu Zhang*, Haizhou Shi*, Yibin Wang, Hengyi Wang, Xiaoxiao He, Zhuowei Li, Haoxian Chen, Ligong Han, Kai Xu, Huan Zhang, Dimitris Metaxas, Hao Wang

i SERVICE

- Reviewer for NeurIPS, ICLR, ACL, EMNLP