

Xu Wan

wanxu@zju.edu.cn

I am an AI Researcher at ByteDance Seed Robotics and received my Ph.D. from Zhejiang University in June 2026, advised by Prof. Mingyang Sun. My research focuses on **robot use agents, LLM post-training, and scalable and safe reinforcement learning**. I have published **20+** papers, including 12 first- or co-first-author papers (including NeurIPS, ICLR, ICML, etc.).

Education

Jun. 2026	Zhejiang University
Sep. 2021	Ph.D. in Control Science and Engineering • National Scholarship • Tencent Scholar
Jun. 2021	China University of Geosciences (Wuhan)
Sep. 2017	B.Eng. in Artificial Intelligence • GPA: 2/182 • National Scholarship

Work Experience

Present	ByteDance • Seed Robotics • AI Researcher
Apr. 2026	Agentic Robotics • Embodied VLM
Feb. 2026	Tencent • HY Team • Research Intern
Oct. 2025	Code Agent • Efficient CoT • Multi-turn RL
Sep. 2025	ByteDance • Seed Robotics • Research Intern
May 2025	LLM Reasoning • RLVR • Spatial Reasoning
Mar. 2025	Alibaba • DAMO Academy • Research Intern
Dec. 2023	LLM-assisted RL • Multi-agent RL

Selected Publications

* denotes equal contribution; # denotes corresponding author.

LLM Post-training

- **Xu Wan**, Wenyue Xu, Shengjie Zhao, Mingyang Sun, Mitigating the Length-Scaling Tax with Online Distillation. **arXiv 2026**
- Jiashun Liu*, Runze Liu*, **Xu Wan***, Jing Liang*, Hongyao Tang, Ling Pan†, Reformulate LLM Reinforcement Learning for Efficient Training under Black-box Discrepancy. **NeurIPS 2026**
- **Xu Wan***, Speed Zhu*, Jiawei Cai, Guang Chen, Ximing Huang, Wiggin Zhou, Mingyang Sun, The Shadow Price of Reasoning: Economic Perspective on Optimal Budget Allocation for LLMs. **ICML 2026**
- Chao Shen*, Zihan Guo*, **Xu Wan***, Zhenghao Yang, Yifan Zhang, Wengi Huang, Jie Song, Zongyan Zhang, Mingyang Sun, ProOPF: Benchmarking and Improving LLMs for Professional-Grade Power Systems Optimization Modeling. **ICML 2026**
- Yifu Luo, Haoyuan Sun, Xinhao Hu, **Xu Wan**, et al., Principled RL for Flow Matching Emerges From the Chunk-level Policy Optimization. **ICML 2026**
- **Xu Wan**, Wenyue Xu, Wei Wang, Wotao Yin, Wenqi Huang, Shengjie Zhao, Mingyang Sun, AdapThink: Adaptive Thinking Preferences for Reasoning Language Model. **ACL 2026 (Findings)**
- **Xu Wan**, Yansheng Wang, Wenqi Huang, Mingyang Sun, Buffer Matters, Unleashing the Power of Off-Policy Reinforcement Learning in Large Language Model Reasoning. **ICLR 2026**
- **Xu Wan**, Wenyue Xu, Chao Yang, Mingyang Sun, Think Twice, Act Once: A Co-Evolution Framework of LLM and RL for Large-Scale Decision Making. **ICML 2025**

Scalable and Safe Reinforcement Learning

- › **Xu Wan**, Chao Yang, Cheng Yang, Jie Song, Mingyang Sun, Fuz-RL: A Fuzzy-Guided Robust Framework for Safe Reinforcement Learning under Uncertainty. **NeurIPS 2025**
- › **Xu Wan**, Wenyue Xu, Chao Yang, Mingyang Sun, SrSv: Integrating Sequential Rollouts with Sequential Value Estimation for Multi-agent Reinforcement Learning. **AAAI 2025 (Oral)**
- › Liyu Zhang, **Xu Wan**, Haochi Wu, Quan Kong, Ruilong Deng, Mingyang Sun, SAMG: Offline-to-Online Reinforcement Learning via State-Action-Conditional Offline Model Guidance. **arXiv 2024**
- › **Xu Wan**, Mingyang Sun, Boli Chen, Zhongda Chu, Fei Teng, AdapSafe: Adaptive and Safe-Certified Deep Reinforcement Learning-Based Frequency Control for Carbon-neutral Power Systems. **AAAI 2023**
- › **Xu Wan**, Lanting Zen, Mingyang Sun, Exploring the Vulnerability of Deep Reinforcement Learning-based Emergency Control for Low Carbon Power Systems. **IJCAI 2022**

AI Applications in Large-scale Systems

- › Yupeng Zhang*, **Xu Wan***, Xiangyun Kong*, Chao Yang, Binda Ma, Wotao Yin, Jian Zhou, IVMR suite: An Industrial-scale Virtual Machine Rescheduling Dataset and Benchmark for Elastic Cloud Service. **SIGKDD 2025**
- › **Xu Wan**, Mingyang Sun, AdapSafe2: Prior-Free Safe-Certified Reinforcement Learning for Multi-Area Frequency Control. **IEEE Transactions on Power Systems**
- › Zhongwei Li, Yang Liu, Peng Qiu, Hongyan Yin, **Xu Wan**[#], Mingyang Sun, Highly Transferable Adversarial Attack Against Deep-Reinforcement-Learning-Based Frequency Control. **Energy Conversion and Economics**
- › Lanting Zen, Mingyang Sun, **Xu Wan**, Zhenyong Zhang, Ruilong Deng, Yan Xu, Physics-Constrained Vulnerability Assessment of Deep Reinforcement Learning-Based SCOPF. **IEEE Transactions on Power Systems**