Xiaoxuan Wang

xw27@cs.ucla.edu | Google Scholar | mandyyyyii.github.io

OBJECTIVE

My research centers on **Agentic Reinforcement Learning** and **LLM Reasoning**, with the goal of building stable, scalable agents for real-world interaction and improving scientific reasoning through planning and tool use. I also organize UCLA DM reading group.

EDUCATION

University of California, Los Angeles Sept. 2022 – Present
Ph.D. in Computer Science Advisor: Wei Wang Los Angeles, CA

University of Illinois Urbana-Champaign Aug. 2018 – May 2022
B.S. in Computer Science, Grainger College of Engineering Urbana, IL

- Minor in Mathematics James Scholar Honor Dean's List

SELECTED PROJECTS

ARLArena: Demystifying Policy Gradient Stability in Agentic Reinforcement Learning 2025 – Present
Project Lead UCLA

- Leading a large-scale project on agentic LLMs, developing principled training and inference recipes across 8+ tasks.
- Building a unified and extensible pipeline supporting 16+ policy optimization algorithms, with extensions to asynchronous agent training and MoE models.
- Developing a multi-agent system for solving interactive agentic tasks and extending it to support multi-agent training.
- Conducting extensive empirical analyses of training stability, scalability, and system efficiency, including cross-hardware comparisons across NVIDIA and AMD GPUs.

SciBench: Evaluating Scientific Problem-Solving in Large Language Models 2023 – 2024

- Developed SciBench to evaluate college-level scientific reasoning and problem solving in LLMs.
- Designed evaluation protocols and analyses for model capabilities and failure modes across scientific domains.
- Published at *ICML 2024*; received media coverage in a **Nature News Feature**.

PUBLICATIONS

ARLArena: Demystifying Policy Gradient Stability in Agentic Reinforcement Learning
Xiaoxuan Wang*, Han Zhang*, Haixin Wang*, Yidan Shi, Ruoyan Li, Kaiqiao Han, Chengyi Tong, Haoran Deng, Alex Taylor, Yanqiao Zhu, Renliang Sun, Jingsheng Cong, Yizhou Sun, Wei Wang
Submitted to ICML, 2026

From Solving to Verifying: A Unified Objective for Robust Reasoning in LLMs
Xiaoxuan Wang, Bo Liu, Song Jiang, Jingzhou Liu, Jingyuan Qi, Xia Chen, Baosheng He
Preprint, 2025

MatSciBench: Benchmarking the Reasoning Ability of Large Language Models in Materials Science
Junkai Zhang*, Jingru Gan*, **Xiaoxuan Wang**, Zian Jia, Changquan Gu, Jianpeng Chen, Yanqiao Zhu, Mingyu Derek Ma, Dawei Zhou, Ling Li, Wei Wang
Preprint, 2025

EAST: Entropy-Based Adaptive Weighting for Self-Training
Xiaoxuan Wang, Yihe Deng, Mingyu Derek Ma, Wei Wang
Submitted to ARR, 2025

SciBench: Evaluating College-Level Scientific Problem-Solving Abilities of Large Language Models
Xiaoxuan Wang*, Ziniu Hu*, Pan Lu*, Yanqiao Zhu*, Jieyu Zhang, Satyen Subramaniam, Arjun R. Loomba, Shichang Zhang, Yizhou Sun, Wei Wang
ICML 2024 **Media Coverage: Nature News Feature**

CliBench: Multifaceted Evaluation of LLMs in Clinical Decision Making
Mingyu Derek Ma, Chenchen Ye, Yu Yan, **Xiaoxuan Wang**, Peipei Ping, Timothy S. Chang, Wei Wang
Preprint, 2024

Memorize and Rank: Evaluating LLMs for Clinical Diagnosis Prediction
Mingyu Derek Ma, **Xiaoxuan Wang**, Yijia Xiao, Anthony Cuturrufo, Wei Wang
AAAI 2025; NeurIPS GenAI4Health 2024

STAR: Boosting Low-Resource Event Extraction via Structure-to-Text Generation
Mingyu Derek Ma, **Xiaoxuan Wang**, Po-Nien Kung, P. Jeffrey Brantingham, Nanyun Peng, Wei Wang
AAAI 2024

Learning under Label Proportions for Text Classification
Jatin Chauhan, **Xiaoxuan Wang**, Wei Wang

EMNLP Findings 2023
Global Responses to the COVID-19 Pandemic: Evidence Finding and Verification
Rotem Dror, **Xiaoxuan Wang**, Dan Roth
Preprint, 2022
Seamless Equal Accuracy Ratio for Inclusive CTC Speech Recognition
Heting Gao, **Xiaoxuan Wang**, Sunghun Kang, et al.
Speech Communication, 2022

INTERNSHIP EXPERIENCE

Meta Platforms, Inc. Summer 2025
Research Intern (AI)
- Developed **GRPO-Verif**, a novel reinforcement learning algorithm that jointly optimizes solution generation and self-verification in LLMs via a unified loss formulation.
- Demonstrated improved self-verification while preserving reasoning performance, with tunable hyperparameters controlling verification signal strength.
- Conducted large-scale empirical evaluations validating robustness and generalization.

Amazon Development Center U.S., Inc. Summer 2024
Applied Scientist Intern
- Implemented and evaluated RLHF methods including REST-EM, Iterative DPO, and variants for tool-integrated problem solving in LLMs.
- Identified sensitivity of REST-EM to data quality and quantity relative to online preference optimization methods.
- Showed that regularization mitigates overfitting from self-generated data and improves tool usage through controlled log-probability optimization.

SKILLS

Research Areas: Agentic RL, Post-Training, LLM Reasoning, Tool-Augmented Agents, RLHF
Languages: Python, C++, Java, SQL
Frameworks: PyTorch, Ray, vLLM, HuggingFace, FSDP