

JIAYU (MILA) WANG

[✉ milawang@cs.wisc.edu](mailto:milawang@cs.wisc.edu) [LinkedIn](https://linkedin.com/in/jiayu-mila-wang) [GitHub](https://github.com/jiayuwu) [Website](https://jiayuwu.github.io)
1205 University Avenue, Office 6548, Madison, WI 53706

EDUCATION

University of Wisconsin-Madison

Ph.D. in Computer Sciences | GPA: 4.00 | Advisors: [Aws Albarghouthi](#), [Fred Sala](#)

Jun 2024 – Aug 2026 / **Dec 2026**

Madison, WI, USA

University of Wisconsin-Madison

Master of Science in Economics | GPA: 3.95

Aug 2022 – May 2024

Madison, WI, USA

Peking University

Bachelor of Economics

Sept 2018 – July 2021

Beijing, China

Beijing Foreign Studies University

Bachelor of Arts

Sept 2017 – Jun 2021

Beijing, China

RESEARCH INTERESTS

Large Language Model, Agents, Reasoning, Mid/Post-training, Efficiency, Multimodality

PUBLICATIONS

LiveResearchBench: A Live Benchmark for User-Centric Deep Research in the Wild [\[Paper\]](#) [\[Code\]](#)

Jiayu Wang, Yifei Ming, Riya Dulepet, Qinglin Chen, Austin Xu, Zixuan Ke, Frederic Sala, Aws Albarghouthi, Caiming Xiong, Shafiq Joty

The International Conference on Learning Representations (ICLR), 2026

Beyond Accuracy: Dissecting Mathematical Reasoning for LLMs Under Reinforcement Learning [\[Paper\]](#) [\[Code\]](#)

Jiayu Wang, Yifei Ming, Zixuan Ke, Caiming Xiong, Shafiq Joty, Aws Albarghouthi, Fred Sala

Neural Information Processing Systems (NeurIPS), 2025

Is A Picture Worth A Thousand Words? Delving Into Spatial Reasoning for Vision Language Models [\[Paper\]](#) [\[Code\]](#)

Jiayu Wang, Yifei Ming, Zhenmei Shi, Vibhav Vineet, Xin Wang, Neel Joshi

Neural Information Processing Systems (NeurIPS), 2024

Grammar-Aligned Decoding [\[Paper\]](#) [\[Code\]](#)

Jiayu Wang*, Kanghee Park*, Taylor Berg-Kirkpatrick, Nadia Polikarpova, Loris D'Antoni

Neural Information Processing Systems (NeurIPS), 2024

SkillOrchestra: Learning to Route Agents via Hierarchical Skill Transfer

Jiayu Wang, Aws Albarghouthi, Frederic Sala

preprint, 2026

COSMOS: Are Performance–Cost Tradeoffs Predictable in Model–Strategy Selection? [\[Paper\]](#) [\[Code\]](#)

Jiayu Wang, Aws Albarghouthi, Fred Sala

preprint, 2025

Synthesizing Agentic Data for Web Agent Training with Progressive Difficulty Enhancement

Shrey Pandit, Xuan-Phi Nguyen, Yifei Ming, Austin Xu, Jiayu Wang, Caiming Xiong, Shafiq Joty

preprint, 2025

Helpful Agent Meets Deceptive Judge: Understanding Vulnerabilities in Agentic Workflows [\[Paper\]](#) [\[Code\]](#)

Yifei Ming, Zixuan Ke, Xuan-Phi Nguyen, Jiayu Wang, Shafiq Joty

preprint, 2025

PROJECTS

Multi-agent System for Deep Research

Jun 2025 – present

Advisor: [Shafiq Joty](#)

- Developed a multi-agent research system with memory, planning, and tool-integration modules, achieving SoTA performance on DeepResearch Bench.

- Designed a new evaluation framework and benchmark for diverse deep research tasks in enterprise, featuring checklist-based evaluation and a citation-traced statement quality metric.

Dissecting Mathematical Reasoning of LLMs under Reinforcement Learning

Feb 2025 – May 2025

Advisor: [Aws Albarghouthi](#), [Fred Sala](#)

- Conducted systematic analysis of Large Language Models' mathematical reasoning capabilities, examining the distinct effects of reinforcement learning on knowledge acquisition versus reasoning processes through top-down and sequential evaluation frameworks.

Efficient Language Model–Strategy Routing System

May 2024 – Jan 2025

Advisor: [Aws Albarghouthi](#), [Fred Sala](#)

- Proposed a novel dynamic LLM routing system for underrepresented domains, capable of incorporating fine-tuning and prompting to the routing system.
- First to demonstrate that dynamic updates are predictable and further designed a framework that accurately estimates the performance gains and costs of fine-tuning and prompting before incurring actual expenses.
- Constructed a holistic benchmark to assess the domain transfer capabilities of competitive language model routing systems.

Delving into Spatial Reasoning for Vision-Language Models

Jan 2024 – May 2024

Advisor: [Neel Joshi](#), [Vibhav Vineet](#)

- Developed a multi-task benchmark to evaluate the spatial reasoning abilities of multi-modal models.
- Conducted a systematic and comprehensive evaluation of competitive vision-language models and pure language models and revealed new insights on the impact of input modalities, model scale and model modalities.

Grammar-Aligned Decoding with Large Language Models

Jan 2024 – May 2024

Advisor: [Loris D'Antoni](#)

- Established a new problem setting where pre-trained language models are constrained to follow formal specifications while preserving its ground-truth distribution.
- Designed and implemented a Bayesian Trie-based algorithm that automatically adjusts token probabilities during decoding according to derived closed-form solutions.

Text2LaTeX: Potential of Language Models in LaTeX Synthesis [\[Paper\]](#)

Sept 2023 – Dec 2023

- Designed a novel tree-based LaTeX expression generation algorithm for diverse and scalable dataset construction.
- Conducted extensive finetuning on 350M and 2B CodeGen and LLaVA models to better understand the impact of modality and the efficacy of finetuning pre-trained models on LaTeX synthesis based on multi-modal inputs.
- Performed a series of ablation studies and provided analysis to better understand why visual input may not necessarily enhance code generation compared to text-only inputs.

Stochastic Algorithm for Group Distributionally Robust Optimization [\[Paper\]](#)

Sept 2023 – Dec 2023

- Derived a new formulation for solving GDRO with prior shifts via Benders Decomposition under convex conditions.
- Validated the algorithm's effectiveness through extensive testing, demonstrating superior performance over competitive baselines such as ERM and GDRO with Online Optimization.

INDUSTRY EXPERIENCE

Salesforce

Jun 2025 – present

Research Intern, AI Research

Palo Alto, CA, USA

- Developed a multi-agent system to perform complex deep research tasks and designed corresponding evaluation methods and benchmarks.

Genesis MedTech Group

Apr 2022 – Aug 2022

Mergers and acquisitions (M&A) Intern, Business Development

Shanghai, China

- M&A Strategy: Assisted in 7 M&A initiatives in the medical devices sector by performing comprehensive desk research, conducting expert interviews, and due diligence while aiding negotiation including Term Sheet review and risk & opportunities assessment.
- Financial Benchmarking: Developed financial analysis templates for 100+ comparable listed companies, facilitating automatic performance tracking against industry benchmarks.
- Streamlined financial valuation processes for the Group's B+ round financing, accelerating valuations by 2 weeks through the implementation of WACC and LoMD methodologies.

Alpha Startup Fund

Jun 2021 – Sept 2021

Intern, Post-investment Management

Beijing, China

- Supported fundraising of \$50M for Phase II fund and the launch of USD and RMB private equity funds.
- Negotiated terms and conditions of Limited Partnership Agreement & Side Letter with Limited Partners, finalizing LPA & saving time by 50%.
- Risk Management: Engaged with 15 portfolio companies across AI, FinTech, SaaS, and IoT sectors to evaluate and advise on risk management and strategic opportunities.

INVITED TALKS

Salesforce AI Research

Multi-agent Systems for Complex Enterprise Research Tasks

Aug 2025

MadPL Seminar, University of Wisconsin–Madison

Automated Post-training Development of Large Language Models

Apr 2025

Microsoft Research

Spatial Reasoning in Vision–Language Models

Sep 2024

AWARDS & HONORS

NeurIPS 2024 Travel Award

Oct 2024

National Scholarship (Top 1%)

Dec 2020

China “Internet+” Innovation Competition, Regional Prizewinner

Sep 2020

SKILLS

Programming & Frameworks: Python (PyTorch, NumPy, Pandas, TensorFlow), UNIX shell, Docker, Kubernetes
Languages: English (Fluent), Laotian (Fluent), Thai (Basic), Mandarin (Native)

Data Visualization & Database: Tableau, Power BI, MySQL, MongoDB

Statistics & Optimization: Gurobi, Stata, R, SPSS

Finance: ACCA (passed 11/13), SAC (Securities Qualification of China), Bloomberg, Wind, Pitchbook, Capital IQ

Interests: Tennis (National Second Level Athlete), Pipa, Chinese Dance, Classical Guitar