Xingjian Zhang jimmyzxj@umich.edu | xingjianz.com | 734-604-4231

Research Focus

Benchmark & Eval · Post-Training · Synthetic Data for Agents · Personalization · AI for Science · Graph ML

Education

University of Michigan, Ph.D. in Information Science, *Advised by Prof. Qiaozhu Mei* Sep 2022 – Present
University of Michigan, B.S. in Computer Science; Minor in Mathematics Sep 2020 – Apr 2022
Shanghai Jiao Tong University, B.S. in Electrical and Computer Engineering Sep 2018 – Aug 2020

Work Experience

Meta Superintelligence Labs, Muse Product Post-Training, *Research Scientist Intern* May 2026 – Present

- **Guided Synthetic Data Generation for Personal Agents**
  - Co-developed steering mechanisms to guide agent rollout towards desired behavior including reflection, plan-and-execute, and adaptive replanning.
  - Contributed personalized-research SFT data to Muse Spark post-training; the trained model improved by 3pp across three internal agent benchmarks.
  - Migrated rollout env from a fast-evolving production black box to a controlled white-box sandbox, enabling reproducible experiments and faster team iteration.

Google DeepMind, Gemini Post-Training, *Research Intern* May 2025 – Aug 2025

- **Agentic RL for Multimodal Tool-Using Agents**
  - Curated a benchmark linking YouTube Shorts to real-world events to evaluate video grounding through web search.
  - Contributed agentic RL and multimodal-input support to `simply`, a minimal JAX codebase used broadly within DeepMind with 500+ GitHub stars. [simply]

Google DeepMind, Gemini Post-Training, *Research Intern* May 2023 – Aug 2023

- **Adaptive Computation for Efficient Transformer Fine-Tuning and Serving** [ICLR Workshop'24]
  - Proposed GPU/TPU-friendly adaptive layer skipping based on input-sequence complexity.
  - Matched full-compute performance on language and sequential user modeling at 60% compute; integrated the method into TensorFlow Recommenders.

University of Michigan, *PhD Research Assistant* Sep 2022 – Present

- **Foundation Models Reasoning and Evaluation**
  - Inferred human thinking traces from label-only ratings via rejection sampling, improving agreement between LLM raters and human judgments. [TMLR'26]
  - Separated perception from reasoning across three ARC-style datasets; found that about 80% of inspected failures stemmed from perception errors. [ACL'26]
  - Co-developed an open foundation model and benchmarks for human behavior modeling. [In Submission]
- **Agents for Scientific and Data Workflows**
  - Built an end-to-end benchmark for curating structured knowledge from genetics literature; identified distinct cost-performance trade-offs across agent architectures. [In Submission]
  - Extracted scientific workflows from research papers to support downstream generation and recommendation tasks. [NAACL'25]
  - Benchmarked agents' ability to detect real-world dataset issues. [KDD'25]
- **Graph Machine Learning** [TMLR'24, NeurIPS Workshop'24, WSDM'22, LoG'22]

Leadership & Honors

Workshop Organizer: Graph Learning Benchmark @ KDD'23; LLMs for Individuals, Groups, and Society @ WSDM'24.
Selected Honors: Rackham Doctoral Fellowship; James B. Angell Scholar; Tang Junyuan Scholarship; J. Wu & J. Sun Sunshine Scholarship.

Publications/Preprints

[In Submission] FlyAOC: Evaluating Agentic Ontology Curation of Drosophila Scientific Knowledge Bases
Xingjian Zhang, Sophia Moylan, Ziyang Xiong, Qiaozhu Mei, Yichen Luo, Jiaqi W. Ma

[In Submission] Be.FM: Open Foundation Models for Human Behavior
Yutong Xie, Zhuoheng Li, Xiyuan Wang, Yijun Pan, Qijia Liu, Xingzhi Cui, Kuang-Yu Lo, Ruoyi Gao, **Xingjian Zhang**, Jin Huang, Walter Yuan, Matthew O. Jackson, Qiaozhu Mei

[ACL'26] Your Reasoning Benchmark May Not Test Reasoning: Revealing Perception Bottleneck in Abstract Reasoning Benchmarks
Xiyuan Wang, Jin Huang, **Xingjian Zhang**, Tianhao Wang, Jiaqi W. Ma

[TMLR'26] Through the Judge's Eyes: Inferred Thinking Traces Improve Reliability of LLM Raters
Xingjian Zhang, Tianhong Gao, Suliang Jin, Tianhao Wang, Teng Ye, Eytan Adar, Qiaozhu Mei

[KDD'25] DCA-Bench: A Benchmark for Dataset Curation Agents
Benhao Huang, Yingzhuo Yu, Jin Huang, **Xingjian Zhang**, Jiaqi W. Ma

[NAACL'25] MASSW: A New Dataset and Benchmark Tasks for AI-Assisted Scientific Workflows
Xingjian Zhang*, Yutong Xie*, Jin Huang, Jinge Ma, Zhaoying Pan, Qijia Liu, Ziyang Xiong, Tolga Ergen, Dongsub Shim, Honglak Lee, et al.

[KDD'25] MapExplorer: New Content Generation from Low-Dimensional Visualizations
Xingjian Zhang, Ziyang Xiong, Shixuan Liu, Yutong Xie, Tolga Ergen, Dongsub Shim, Hua Xu, Honglak Lee, Qiaozhu Mei

[NeurIPS'24] **Spotlight** · dattri: A Library for Efficient Data Attribution
Junwei Deng, Ting-Wei Li, Shiyuan Zhang, Shixuan Liu, Yijun Pan, Hao Huang, Xinhe Wang, Pingbang Hu, **Xingjian Zhang**, Jiaqi W. Ma

[TMLR'24] Can LLMs Effectively Leverage Graph Structural Information: When and Why
Jin Huang, **Xingjian Zhang**, Qiaozhu Mei, Jiaqi Ma

[NeurIPS Workshop'24] Leveraging LLM-Generated Structural Prior for Causal Inference with Concurrent Causes
Xingjian Zhang, Shixuan Liu, Yixin Wang, Qiaozhu Mei

[ICLR Workshop'24] Conditional Transformer Fine-Tuning by Adaptive Layer Skipping
Xingjian Zhang, Jiaxi Tang, Yang Liu, Xinyang Yi, Li Wei, Lichan Hong, Qiaozhu Mei, Ed H. Chi

[WSDM'22] Fast Learning of MNL Model from General Partial Rankings with Application to Network Formation Modeling
Jiaqi Ma*, **Xingjian Zhang***, Qiaozhu Mei

[LoG'22] **Oral** · Graph Learning Indexer: A Contributor-Friendly and Metadata-Rich Platform for Graph Learning Benchmarks
Jiaqi Ma*, **Xingjian Zhang***, Hezheng Fan, Jin Huang, Tianyue Li, Ting Wei Li, Yiwen Tu, Chenshu Zhu, Qiaozhu Mei

* Equal contribution.

Open Source Projects

[simply] **google-deepmind/simply**
Minimal JAX codebase for rapid iteration on LLM research; 500+ GitHub stars.

[dattri] **TRAIS-Lab/dattri**
PyTorch library for efficient data attribution.

[gli] **Graph-Learning-Benchmarks/gli**
Metadata-rich graph learning benchmark platform.