

Education

University of Illinois Urbana-Champaign (UIUC) <i>Ph.D. in Information Sciences</i> Advisor: Prof. Jingrui He	Aug 2024 – May 2029
University of Illinois Urbana-Champaign (UIUC) <i>Master of Computer Science</i>	Aug 2022 – Dec 2023
Xi'an Jiaotong University (XJTU) <i>B.E. in Computer Science</i>	Sep 2018 – Jul 2022

Publications

* denotes equal contribution.

• Peer-Reviewed Conference Publications

C1	Does Accuracy Equal Evidence? Reasoning Faithfulness Under KV Cache Compression Mengting Ai, Jingrui He, Yue Guo EMNLP 2026 Findings KV Cache Compression Reasoning Faithfulness 29% acc. rate
C2	Learning What Matters: End-to-End Adaptive Calibration Data Selection for LLM Pruning Mengting Ai, Sirui Chen, Xinrui He, Maggie Zarekarizi, Kommy Weldemariam, Jingrui He EMNLP 2026 Findings LLM Pruning Calibration Data Selection 29% acc. rate
C3	NIRVANA: Structured Pruning Reimagined for Large Language Model Compression Mengting Ai, Tianxin Wei, Sirui Chen, Jingrui He COLM 2026 LLM Efficiency Structured Pruning 29% acc. rate
C4	MedConceal: A Benchmark for Clinical Hidden-Concern Reasoning Under Partial Observability Yikun Han, Joey Chan, Jingyuan Chen, Mengting Ai, Simo Du, Yue Guo COLM 2026 Clinical AI Benchmark 29% acc. rate
C5	Event Graph-Based Reliable Clinical Data Imputation Xinrui He, Mengting Ai, Junting Wang, Curtiss Cook, Jingrui He KDD 2026 Clinical Data Graph ML 18.5% acc. rate
C6	AutoTool: Dynamic Tool Selection and Integration for Agentic Reasoning Jiaru Zou, Ling Yang, Yunzhe Qi, Sirui Chen, Mengting Ai, Ke Shen, Jingrui He, Mengdi Wang ICML 2026 Agentic AI Tool Use 26.6% acc. rate
C7	Competing Visions of Ethical AI: A Case Study of OpenAI Melissa Wilfley*, Mengting Ai*, Madelyn Rose Sanfilippo iConference 2026 AI Ethics Case Study 24.5% acc. rate
C8	MC-Search: Evaluating and Enhancing Multimodal Agentic Search with Structured Long Reasoning Chains Xuying Ning, Dongqi Fu, Tianxin Wei, Mengting Ai, Jiaru Zou, Ting-Wei Li, Jingrui He ICLR 2026 Multimodal Agents Benchmark 1.2% acc. rate Oral
C9	Influence-Preserving Proxies for Gradient-Based Data Selection in LLM Fine-Tuning Sirui Chen, Yunzhe Qi, Mengting Ai, Yifan Sun, Ruizhong Qiu, Jiaru Zou, Jingrui He ICLR 2026 Data Selection LLM Fine-Tuning Efficiency 28.0% acc. rate

C10

ResMoE: Space-Efficient MoE Module Approximation via Wasserstein Barycenter and Residual RestorationMengting Ai*, Tianxin Wei*, Yifan Chen*, Zhichen Zeng, Ritchie Zhao, Girish Varatkar, Bitu Darvish Rouhani, Xianfeng Tang, Hanghang Tong, Jingrui He

KDD 2025 LLM Efficiency Mixture-of-Experts 18.7% acc. rate

C11

APEX²: Adaptive and Extreme Summarization for Personalized Knowledge GraphsZihao Li, Dongqi Fu, Mengting Ai, Jingrui He

KDD 2025 Knowledge Graphs Summarization 18.7% acc. rate

C12

PyG-SSL: A Graph Self-Supervised Learning ToolkitLecheng Zheng, Baoyu Jing, Zihao Li, Zhichen Zeng, Tianxin Wei, Mengting Ai, Xinrui He, Lihui Liu, Dongqi Fu, Jiaxuan You, Hanghang Tong, Jingrui He

CIKM 2025 Graph SSL Toolkit 43.2% acc. rate

• Workshop Papers

W1

AutoTool: Dynamic Tool Selection and Integration for Agentic ReasoningJiaru Zou, Ling Yang, Yunzhe Qi, Sirui Chen, Mengting Ai, Ke Shen, Jingrui He, Mengdi Wang

ICCV 2025 Workshop MMRAgl Agentic AI Tool Use Best Paper Award

W2

TASP: Preserving Training Dynamics in Transformers via NTK-Aware Structured PruningMengting Ai, Tianxin Wei, Jingrui He

ICLR 2025 Workshop SLLM Structured Pruning Efficient Fine-Tuning

• Preprints & Manuscripts

P1

MLP Fusion: Towards Efficient Fine-Tuning of Dense and Mixture-of-Experts Language ModelsMengting Ai*, Tianxin Wei*, Yifan Chen*, Jingrui He

Under Review Mixture-of-Experts Efficient Fine-Tuning

P2

Radiohead: Remote Sensing Earth Foundation Model Fusion with Mixture-of-ExpertsMengting Ai, Lecheng Zheng, Zihao Li, Xinrui He, Maggie Zarekarizi, Prasanth Meiyappan, Kommy Weldemariam, Jingrui He

Under Review Earth Observation FMs Model Fusion

P3

Evo-Memory: Benchmarking LLM Agent Test-Time Learning with Self-Evolving MemoryTianxin Wei, Naveen Sachdeva, Benjamin Coleman, Zhankui He, Yuanchen Bei, Xuying Ning, Mengting Ai, Yunzhe Li, Jingrui He, Ed H. Chi, Chi Wang, Shuo Chen, Fernando Pereira, Wang-Cheng Kang, Derek Zhiyuan Cheng

Under Review Self-Evolving Memory Benchmark

P4

Agentic Reasoning for Large Language ModelsTianxin Wei, Ting-Wei Li, Zhining Liu, Xuying Ning, Ze Yang, Jiaru Zou, Zhichen Zeng, Ruizhong Qiu, Xiao Lin, Dongqi Fu, Zihao Li, Mengting Ai, Duo Zhou, Wenxuan Bao, Yunzhe Li, Gaotang Li, Cheng Qian, Yu Wang, Xiangru Tang, Yin Xiao, Liri Fang, Hui Liu, Xianfeng Tang, Yuji Zhang, Chi Wang, Jiaxuan You, Heng Ji, Hanghang Tong, Jingrui He

Under Review Agentic Reasoning Survey

P5

Heterogeneous Scientific Foundation Model CollaborationZihao Li, Jiaru Zou, Feihao Fang, Xuying Ning, Mengting Ai, Tianxin Wei, Sirui Chen, Xiyuan Yang, Jingrui He

Under Review Scientific FMs Model Collaboration

P6

Harnessing Generalist Agents for Contextualized Time SeriesZihao Li, Kaifeng Jin, Yuanchen Bei, Jiaru Zou, Avaneesh Kumar, Xuying Ning, Yanjun Zhao, Mengting Ai, Baoyu Jing, Hanghang Tong, Jingrui He

Under Review Time Series Agents Harness

P7

Code as Agent HarnessXuying Ning, Katherine Tieu, Dongqi Fu, Tianxin Wei, Zihao Li, Yuanchen Bei, Jiaru Zou, Mengting Ai, Zhining Liu, Ting-Wei Li, Lingjie Chen, Yanjun Zhao, Ke Yang, Bingxuan Li, Cheng Qian, Gaotang Li, Xiao Lin, Zhichen Zeng, Ruizhong Qiu, Sirui Chen, Yifan Sun, Xiyuan Yang, Ruida Wang, Rui Pan, Chenyuan Yang, Dylan Zhang, Liri Fang, Zikun Cui, Yang Cao, Pan Chen, Dorothy Sun, Ren Chen, Mahesh Srinivasan, Nipun Mathur, Yinglong Xia, Hong Li, Hong Yan, Pan Lu, Lingming Zhang, Tong Zhang, Hanghang Tong, Jingrui He

Under Review Code Agents Survey

Community Service

- 2026 ICML (Silver Reviewer); NeurIPS; COLM; TMLR; AAAI
- 2025 ICLR; TKDD; iConference (Reviewer & Moderator); FM4RoboPlan@RSS
- 2024 AAAI

Research Experience

Capital One

Applied Research Intern

Jun 2026 – Aug 2026

- Conducted research on the pretraining of state-space models.

Institute of Software, Chinese Academy of Sciences

NLP Algorithm Intern

Oct 2021 – Jan 2022

- Built a cybersecurity question-answering system using DAPT-BERT.
- Crawled security-related Stack Overflow questions and constructed a semantic-similarity training dataset.
- Developed a DAPT-BERT binary classifier and achieved 92% accuracy with the DAPT100k model.

Baidu

NLP Algorithm Research Intern

Feb 2021 – Aug 2021

- Built a three-layer XGBoost text classifier with Word2Vec features, achieving an F1 score of approximately 93%.
- Expanded an employee skill-tag pool with 4-grams and TF-IDF, increasing tag coverage from 50% to 80%.
- Improved hard-skill labels through BERT-based encoding and iterative centroid refinement.

Professional Experience

UIUC Prairie Research Institute

Software Engineer Intern

Jul 2023 – Aug 2024

- Modeled time-series rainfall and streamflow data with PyTorch.
- Developed an interactive map that calculates and displays watershed areas from user-selected streamflow points.
- Implemented the system with HTML, CSS, JavaScript, Leaflet, Node.js, and the ArcGIS Pro API.

TE Connectivity

Digital Intern

Jun 2022 – Mar 2023

- Researched emerging AI technologies, including ChatGPT, and presented findings to the digital team.
- Developed RESTful data-processing interfaces with FastAPI for CRUD operations on cloud-hosted MySQL and MongoDB databases.
- Deployed the Mingdao low-code platform on AWS EC2 with Docker and Nginx load balancing and reverse proxying.
- Automated SQL generation for backend data access and manipulation, substantially reducing development costs.

Selected Projects

AI Research Agent

Independent Project

May 2025 – Jul 2025

- Engineered a privacy-preserving local-LLM agent that tracks and recommends academic papers.
- Built a daily pipeline to collect new papers, update a local vector knowledge base, and send personalized email summaries.
- Orchestrated ingestion, retrieval, generation, and notification modules with LangChain.

Farmer's Market Order System

Full-Stack Web Project

Oct 2022 – Nov 2022

- Implemented product browsing, shopping cart, order placement, order history, and farmer-facing product CRUD workflows.
- Built the front end with HTML, CSS, and JavaScript and the backend with MySQL, Spring Boot, JDBC, and MyBatis.

Web Chatroom Application

Apr 2022 – May 2022

Full-Stack Web Project

- Implemented authentication, group chat, and file-upload functionality.
- Built the application with HTML, CSS, JavaScript, iFrame, Node.js, PHP, MySQL, Express, and Socket.IO.

Awards

ICML 2026 Silver Reviewer	May 2026
UIUC Fall 2025 Facilitating Learning Excellence Reward (IS 577 Data Mining TA)	Apr 2026
ICCV 2025 MMRAgI Workshop Outstanding Paper	Oct 2025
KDD Student Travel Award	Jun 2025
Outstanding Student (top 15%), XJTU	Oct 2021
Third-Class Scholarship (top 20%), XJTU	Oct 2020 & 2021
2nd Place, KPMG × Yiyou Youth Center “STEM is HER Future” (2/48)	Aug 2021

Invited Talk

ResearchTrend.AI: On Structured Pruning	Nov 2025
---	----------