Yimin Wang

Ann Arbor, MI wyimin@umich.edu (+1) 734-358-6981 Homepage LinkedIn

Education

University of Michigan, Ann Arbor — *Ann Arbor, MI*
B.S.E. in Data Science — *Aug 2024 – May 2026*

- **Coursework:** Intro to Machine Learning (A), Intro to NLP (A), Statistics and Artificial Intelligence (A^+), Applied Regression Analysis (A^+), Probability and Statistics (A^+), Linear Algebra (A^+)

Shanghai Jiao Tong University — *Shanghai, China*
B.S. in Mechanical Engineering — *Aug 2022 – Aug 2026*

- **Coursework:** Design and Manufacturing I (A)/II (A), Laboratory I (A), Solid Mechanics (A), Dynamics and Vibrations (A^+)
- **Scholarships & Awards:** 2023 The John Wu & Jane Sun Sunshine Scholarship; Thirteenth Freshman Robotics Competition Best Design Award

Research Experience

LLM Agent Development — *Remote*
Princeton University (Advisor: Prof. Mengdi Wang; Supervisor: Jiahao Qiu) — *Oct 2024 – Present*

- Researched scalable multi-agent architectures for LLMs, focusing on reasoning, safety alignment, and cross-domain adaptability, using AutoGen or SmolAgents.
- Developed modular agent frameworks integrating simulation, distillation, and tool-based multimodal reasoning.
- Contributed to domain-specific and generalizable agent systems including HistAgent, EmoAgent, and Agent-Distill.
- Collaborated on the development of benchmarks and evaluation pipelines for multimodal and human-centered AI agents, emphasizing interpretability, reliability, and cross-domain generalization.

Trustworthy AI and LLM Safety — *Remote*
Northwestern University (Advisor: Prof. Yan Chen; Supervisor: Robin Luo) — *April 2025 – Present*

- Conducted research on adversarial robustness and safety of large language and genomic foundation models, contributing to GenoArmory, a unified benchmark for adversarial attack and defense evaluation.
- Performed literature review and experiments on LLM thinking safety and participated in building and fine-tuning an LLM specialized for cloud-configuration Q&A.

LLM Reasoning with Code — *Ann Arbor, MI, USA*
University of Michigan (Advisor: Prof. Lu Wang; Supervisor: Frederick Zhang, Kaijian Zou) — *June 2025 – Present*

- Conducting research on the reasoning process of large language models in code generation and problem solving, aiming to understand how model reasoning behavior affects overall performance.
- Analyzing reasoning traces across programming and mathematical benchmarks through entropy-based and structural evaluations to reveal consistency patterns, error sources, and potential directions for improving code reasoning reliability.

Low-power Intelligent Gas Sensor — *Shanghai, China*
Shanghai Jiao Tong University (Advisor: Prof. Jianhua Yang; Supervisor: Tao Wang) — *Sep 2023 – May 2024*

- Collaborated on sensor experiments and data collection.
- Built and improved analytical models for electronic nose signals in Python (Keras).
- Integrated four algorithms to predict gas type and concentration for unknown mixtures.

Selected Projects

On Path to Multimodal Historical Reasoning: HistBench and HistAgent

- Developed HistAgent, a tool-augmented multimodal reasoning agent designed to analyze historical sources through text, image, and document inputs.
- Implemented manager orchestration and integrated tools for OCR, translation, image/audio understanding, and literature retrieval using SmolAgent and browser-based APIs.
- Ran large-scale evaluations on HistBench, GAIA, and HLE-History, comparing performance against GPT-4o and Deep Research baselines.
- Coled this project, contributed to paper writing, systematic design, and experiments, focusing on reasoning reliability and multimodal grounding.

AgentDistill: Training-Free Agent Distillation with Generalizable MCP Boxes

- Proposed a **training-free distillation framework** that transfers reasoning and tool-use capabilities between agents via Model Context Protocols (MCPs).
- Led method design, experiment implementation, and paper writing; demonstrated effective skill transfer from large to small LLM agents across biomedical and mathematical tasks.

EmoAgent: Assessing and Safeguarding Human-AI Interaction for Mental Health Safety

- Built a multi-agent framework to evaluate and mitigate psychological risks in human–AI conversations, integrating sentiment detection and content moderation.
- Conducted ablation experiments, refined evaluation modules, and analyzed model sensitivity across different emotional scenarios.
- Participated in writing and rebuttal preparation for the submitted paper.

AI for Quantum Materials Automation: Multi-Agent Exfoliation System

- Developed a multi-agent robotic framework for end-to-end automation of 2D material exfoliation and stacking, coordinating robotic, vision, and control agents.
- Integrated computer vision for crystal alignment and implemented closed-loop coordination between hardware subsystems to enable reliable autonomous operation.
- Designed system-level state monitoring and safety mechanisms to enhance consistency, fault recovery, and overall process stability.

Interactive Web Game Development

- Designed a responsive and accessible web game interface using Elm, emphasizing user feedback and visual clarity.
- Implemented in-game communication and event handling with Elm's message-passing architecture.
- Created original hand-drawn art assets and integrated them into a consistent UI design.

Technologies

Languages: C/C++, Python, HTML/CSS, JavaScript, SQLite, R, Elm

Frameworks: PyTorch, scikit-learn, LangChain, AutoGen, HuggingFace/SmolAgent

Tools: Git, Linux, Shell, Jupyter Notebook, LaTeX

Publications

** indicates equal contribution.*

On Path to Multimodal Historical Reasoning: HistBench and HistAgent 2026 ICLR under review

Jiahao Qiu*, Fulian Xiao*, **Yimin Wang***, Yuchen Mao*, Yijia Chen*, Xinzhe Juan, Siran Wang, Xuan Qi, Tongcheng Zhang, Zixin Yao, and others

AgentDistill: Training-Free Agent Distillation with Generalizable MCP Boxes 2026 ICLR under review

Jiahao Qiu*, Xinzhe Juan*, **Yimin Wang***, Ling Yang*, Xuan Qi, Tongcheng Zhang, Jiacheng Guo, Yifu Lu, Zixin Yao, Hongru Wang, Shilong Liu, Xun Jiang, Liu Leqi, Mengdi Wang

EmoAgent: Assessing and Safeguarding Human-AI Interaction for Mental Health Safety 2025 EMNLP oral

Jiahao Qiu*, Yinghui He*, Xinzhe Juan*, **Yimin Wang**, Yuhan Liu, Zixin Yao, Yue Wu, Xun Jiang, Ling Yang, Mengdi Wang

GenoArmory: A Unified Evaluation Framework for Adversarial Attacks on Genomic Foundation Models 2026 ICLR under review

Haozheng Luo*, Chenghao Qiu*, **Yimin Wang**, Shang Wu, Jiahao Yu, Han Liu, Binghui Wang, Yan Chen

Alita: Generalist Agent Enabling Scalable Agentic Reasoning with Minimal Predefinition and Maximal Self-Evolution 2026 ICLR under review

Jiahao Qiu*, Xuan Qi*, Tongcheng Zhang*, Xinzhe Juan, Jiacheng Guo, Yifu Lu, **Yimin Wang**, Qihan Ren, Xun Jiang, Xing Zhou, Dongrui Liu, Ling Yang, Yue Wu, Kaixuan Huang, Shilong Liu, Hongru Wang, Mengdi Wang

High-precision control of an antagonistic soft continuum robot for dexterous objects grasping and assembly Sensors and Actuators A: Physical

Shoulu Gong, Xinchen Ye, **Yimin Wang**, Wenbo Li, Wenming Zhang, Lei Shao

A Novel Approach to Air Quality Monitoring: Towards Miniature, Self-organized, and Low-power Device 2023 IEEE SENSORS

Tao Wang, Yu Wu, Wangze Ni, Jianhua Yang, **Yimin Wang**, Jiaqing Zhu, Ming Zeng, Nantao Hu, Zhi Yang