

Zhang Jiahao

Email: jiahao.zhang.public@gmail.com

<https://jiahaozhang-public.github.io/>
or jiahao.zhang@mbzuai.ac.ae

RESEARCH INTERESTS

- **Explainable AI (XAI):** Developing interpretable ML methods to enhance transparency and trust, with a focus on biomedical applications.
- **Controllable Generative Models for Biology:** Designing diffusion/LM-based controllable generators for biological sequences/structures to enable targeted discovery in protein engineering and drug design.

EDUCATION

- **Ph.D. in Machine Learning** Aug. 2025 – Present
Mohamed bin Zayed University of Artificial Intelligence (MBZUAI) Abu Dhabi, UAE
 - **Supervisor:** Lijie Hu
 - **Research Focus:** Explainable AI (XAI), Controllable Generation Models for Biological Research
- **B.Eng in Electronic Information** Sept. 2021 – June 2025 (Expected)
Huazhong University of Science and Technology (HUST) (985, QS Top 100) Wuhan, China
 - **GPA:** 86/100
 - **Specialization:** Electronic Packaging Technology
 - **Relevant Coursework:** Probability Theory and Mathematical Statistics, Stochastic Processes, Computational Methods, Digital Circuits, Python Programming, Signals and Systems, Information Retrieval, Principles of Microcomputer, Data Structures
- **UC Berkeley** Jan. 2024 – May 2024
Visiting Student | GPA: 3.57/4 Berkeley, CA
 - **Key Courses:** Data Structures, Artificial Intelligence

PUBLICATIONS

- **Jiahao Zhang, Jerry Wang. *TIC: A Unified Framework for Temporal and Causal Inference in Tumor Microenvironments*. Manuscript in preparation.**: Developed a computational framework integrating graph-based pseudotime trajectory analysis and causal inference methods for understanding EMT progression in tumor microenvironments. Python package: <https://github.com/cellethology/tic>.

RESEARCH EXPERIENCE

- **Laboratory of Cell Ethology, CIS, Westlake University** Jan. 2025 – July 2025
Research Assistant (Advisor: Dr. Jerry Wang) Hangzhou, China
 - **TIC: Temporal and Causal Inference of Cellular States:** Led the development of TIC, a modular Python framework for analyzing spatial transcriptomics in tumor microenvironments. Integrated LLMs for cell type annotation, GNNs for feature extraction, pseudotime analysis for trajectory inference, and causal inference for biomarker interactions during EMT.
 - **Key Contributions:**
 - * Designed the computational pipeline: preprocessing, embedding strategies, and model integration.
 - * Implemented graph-based algorithms for pseudotime inference and validated through downstream analyses.
 - * Open-sourced and maintained the Python package (v2.0.0) to support reproducibility and community adoption.
- **Representation Learning Lab, Westlake University** June 2024 – Dec. 2024
Research Assistant Hangzhou, China

- **AI-Driven Protein Affinity Design Using Diffusion Models:** Developed a generative model (RF-diffusion-inspired) to design proteins with controllable binding strength for molecular circuits in synthetic biology and biocomputing.
- **Personal Responsibilities::**
 - * Designed and implemented a custom generative model optimized for high/low protein affinity tasks.
 - * Conducted experiments and fine-tuning to improve model performance.

AI Lab (Chaowei Xiao), University of Wisconsin–Madison

Oct. 2023 – Dec. 2023

Research Assistant

Remote

- **Benchmark for AI in Detecting Software Vulnerabilities:** Developed a benchmark to evaluate AI models' effectiveness in detecting and repairing software vulnerabilities, improving reliability of AI-assisted software security.
- **Personal Responsibilities::**
 - * Conducted literature review to define key evaluation metrics for vulnerability detection.
 - * Helped design and run experiments on real-world vulnerability datasets.
 - * Led data analysis to refine benchmark protocols and identify improvement areas.

PROJECT EXPERIENCE

GPT Insight Generator for Survey Data

Jan. 2024 – May 2024

Data Science Research Intern, Grapedata

UC Berkeley

- **AI-Based Survey Data QA System for Business Insights:** Co-developed a QA system that analyzed B2B survey data to generate actionable insights via RAG and zero-shot inference, learning user preferences beyond explicit responses. Collaborated with industry partners and academic mentors.
- **Personal Responsibilities::** Led a two-stage workflow to fine-tune GPT for robust QA; resolved model tuning and data integration challenges to deliver a production-ready prototype.
- **Results/Impact::** Commercialized product priced at **£10,000+** per unit; received the Cloud Computing Application Award at UC Berkeley Data Science Discovery Program.

COMPETITIONS & HONORS

- **Cloud Computing Application Award — *UCB Data Science Discovery Program* (May 2024):** Recognized for developing an LLM-based insight generator for business survey analytics. More details.
- **Undergraduate Academic Excellence Scholarship — *2023–2024*:** Awarded for outstanding academic performance and integrity.
- **Second Prize — *College Student Mathematics Competition, Hubei Division* (Mar. 2023):** Recognized for strong problem-solving in advanced mathematics.
- **Second Prize — *Chinese Mathematics Competition* (Jan. 2023):** Awarded for outstanding performance in mathematical problem-solving.
- **Third Prize — *China Undergraduate Mathematical Contest in Modeling* (Sept. 2022):** Led a team analyzing and classifying ancient glassware, demonstrating rigorous modeling and data analysis skills.

TECHNICAL SKILLS

- **Programming Languages:** Python (Proficient), C++ (Intermediate), Java (Basic)
- **AI Frameworks & Tools:** PyTorch (Proficient), Git (Intermediate), Diffusion Models (Intermediate), LLMs (Intermediate; focus on generative tasks)

Last Edited: August 14, 2025