Linrui Ma

linrui@mit.edu | (+1) (617)-256-8650

Education

- **Massachusetts Institute of Technology (MIT), Cambridge, MA** *Aug 2025 – Present*
  B.E. in Artificial Intelligence & Decision Making & B.S. in Chemistry (Double Major).
  - GPA: 5.0/5.0
  - Core Courses: Deep Learning; Multimodal AI; Symmetry and its Application to Machine Learning; Representation, Reasoning, and Inference in AI; Algorithms; AI & Decision Making in Medicine.
- **Tsinghua University, Beijing, China** *Aug 2024 – Jun 2025*
  Undergraduate student in Yao class, majoring in Computer Science and Artificial Intelligence.
  - GPA: 3.87/4.00
  - Core Courses: Linear Algebra; Programming in C/C++; Artificial Intelligence: Principles & Techniques; Mathematics for CS & AI; Computer Systems; Introduction to LLM Applications.
- Skills: C/C++, Python, PyTorch, NumPy, Matplotlib, LaTeX, Git, CMake, JAX, Linux.

Research and Experiences

- **Undergraduate Researcher, MIT CSAIL (Cambridge, MA)** *Feb 2026 – Present*
  Advisor: Prof. Kaiming He
  Adapting the "Just Image Transformers" (JiT) framework for text-guided image editing in raw pixel space by integrating a pre-trained T5 encoder with cross-attention on pixel-space tokens. Designing experiments, fine-tuning the model, and evaluating outputs against baselines.
- **Undergraduate Researcher, MIT CSAIL (Cambridge, MA)** *Sep 2025 – Present*
  Advisor: Prof. Regina Barzilay
  Collaborate with Ph.D. candidates under Prof. Regina Barzilay's supervision to design deep learning models that predict cell state changes in response to various cellular perturbations. Developing probabilistic models for selective protein binder design.
- **Instructor, Beijing Training Team for Chinese Chemistry Olympiad (CChO)** *May 2024 – Present*
  Delivered over 100 hours of lectures on advanced chemistry topics for the Beijing Training Team for CChO and other training camps. Covered subjects including advanced organic chemistry, thermodynamics, and biochemistry.

Projects

- **SAPER: Structure-Aware Prompt-Enhanced RAPM Framework** *Nov 2025 – Dec 2025*
  Link to project: `https://github.com/winstonqian/SAPER`.
  Developed a novel protein function annotation framework that combines 3D structural information with sequence-based methods, achieving up to 114% performance improvement on Meta-BLEU-2. Designed multi-modal fusion algorithms (Weighted Similarity and Reciprocal Rank Fusion) to integrate structure-aware ProstT5 and evolutionary ESM-2 embeddings for robust protein retrieval. Engineered hierarchical prompt templates with task-specific biological terminology guidance, demonstrating that retrieval quality and prompt design are equally crucial for LLM performance. Built scalable Python infrastructure leveraging FAISS indexing, transformer models, and automated benchmarking across multiple tasks. Validated approach on the rigorous Prot-Inst-OOD dataset, showing particular strength in structure-dependent tasks like catalytic activity and domain motif. Led the project and collaborated with 3 peers.
- **Representation Efficiency in Neural Networks: Evidence in Multilingual Mathematical Benchmarks** *Oct 2025 – Dec 2025*
  Link to project: `https://github.com/bowenyu066/language-shapes-reasoning`.
  Investigated how representation choice affects neural reasoning efficiency in Transformers by multilingual mathematical benchmarks. Built a comprehensive evaluation framework testing SOTA models and 8B-parameter open-source models on GSM8K and MMATH across multiple languages. Discovered that well-trained models achieve representation-invariant reasoning, and when tokenizers align with denser representations, 5-10% token efficiency gains can be achieved. Demonstrated that representation bottlenecks stem from training distribution rather than architecture by reducing English-Chinese performance gap through LoRA fine-tuning. Developed modular Python infrastructure featuring multi-tokenizer analysis pipeline and automated evaluation across API and local models. Collaborated with 2 peers.
- **Labotex: An Intelligent Agent for Lab Reports** *Mar 2025 – Jun 2025*
  Link to project: `https://github.com/Kehan-Liu/LABOTEX`.
  Developed an AI-powered lab assistant to streamline the generation of LaTeX-formatted lab reports from raw data and textbook descriptions. Integrated visual language models (VLMs), chat models, and a NumPy-based data processing framework to autoconvert experiment instruction PDFs and CSV datasets into structured reports (including abstracts, experimental principles, data analysis, and discussion sections). Built a user-friendly Web UI for seamless file uploads, model configuration, and one-click generation of PDF reports and LaTeX source code. Collaborated with 2 peers.
- **High-Performance Matrix Multiplication with Multithreading & Distributed RPC** *Mar 2025 – Jun 2025*
  Designed and implemented a high-speed matrix multiplication system in C to explore how computation and memory layout affect performance. Improved data reuse and reduced memory bottlenecks by restructuring loops and organizing data into small cache-friendly blocks. Leveraged built-in CPU vector operations and multi-core threading to execute many calculations in parallel. Added support for running tasks across machines and verified accuracy and efficiency through automated tests and benchmarks. Collaborated with 2 peers.

Key Awards

- **1st Place, MIT Informatics Tournament (MITIT) 2025 Winter Contest** (Beginner's Round) *Dec 2025*
- **Tsinghua University Leading Talents Scholarship** (Top 1%, First Class Scholarship for Freshmen) *Sep 2024*
- **56th International Chemistry Olympiad (IChO) Gold Medalist** (*2nd in theory. Captain of Chinese Team*) *Jul 2024*
- **37th Chinese Chemistry Olympiad (CChO) Gold Medalist** (*2nd place nationally*) *Nov 2023*