Zhongheng Liu

Email: zhonghen23@mails.tsinghua.edu.cn
Phone: (+86) 138-2703-2150 / (+65) 8620-9423
Github page: zhliu23-mo

Education

Tsinghua University Beijing, China
B.S. in Electronic Information and Technology Aug. 2023 – June 2027 (Expected)

- **GPA: 3.8/4.0** (**Top 30%** among 255 students).
- **Honors Program:** Selected member of the **"AI + System" Talent Program** (20 students per year).
- **Core Courses:** Fundamentals of Programming (A+), Data Structures & Algorithms, Probability & Stochastic Processes, Signals & Systems, Digital Logic & Processors, Electronic Circuit Systems.

Nanyang Technological University (NTU) Singapore
Exchange Student, School of Electrical and Electronic Engineering (EEE) Aug. 2025 – May 2026

- **Overview:** Fully English-taught curriculum; engaged in international academic research environments.
- **Core Courses:** Image Processing & Computer Vision (A+), Computer Communications (A+), Artificial Intelligence, Operating Systems, Pattern Recognition & Deep Learning.

Research Interests

Reinforcement Learning, LLM agents, Multimodal AI, Graph Learning, Combinatorial Optimization.

Research Experience

Research on Memory Compression for Large Video Models Feb. 2026 – Present
Core Member / Prospective First Author Advisor: Prof. Ziwei Liu MMLab@NTU

- **Research Objective**: Addressing the excessive VRAM consumption of KV Cache in Video LLMs during long-video processing by exploring efficient memory representation and compression mechanisms.
- **Methodology**: Investigating dynamic memory update algorithms based on feature streams or compression domains, aiming to achieve multi-fold context compression without sacrificing understanding precision.
- **Current Progress**: Conducting performance evaluation of baseline models and designing novel compression strategies.

LLM-based Reflective Evolution for Dynamic Network Routing 2025.07 – 2025.09
Second Author (LERNA-Flow), Fourth Author (NeRM-Net) Advisor: Prof. Quanming Yao

- **Background:** Proposed a reflective co-evolution framework using large language models (NeRM-Net) for routing in dynamic networks with topological changes and multi-dimensional constraints (bandwidth, latency, compute), with systematic extensions in LERNA-Flow.
- **Personal Contribution:** Conducted performance evaluation, ablation studies, and results analysis; contributed to writing and revising the extended manuscript. Results show that NeRM-Net improves flow acceptance rate by up to **57.49%** in 100-flow scenarios versus Optimal, Greedy, and ACO baselines; LERNA-Flow reduces violation rate to **13.68%** under dynamic migration.
- **Research Outcome:** Published in **ACM CoNEXT 2025 Workshop** (4th author); completed systematic design and experimental validation of Chinese extension LERNA-Flow (2nd author).

Dynamic Graph Neural Networks for Combinatorial Optimization 2024.12 – 2025.06
Independent Researcher Advisor: Prof. Quanming Yao Undergraduate Research Training Program (SRT)

- Investigated scalability challenges in combinatorial optimization, systematically studying graph neural networks (GNNs) and replicating GCN models to analyze their capabilities and limitations in dynamic graph scenarios.
- Surveyed GNN applications in combinatorial optimization (e.g., routing, resource allocation), reviewing representative works and research paradigms in AI for CO.

- Extended research to LLM-driven combinatorial optimization and algorithm evolution, focusing on ReEvo, FunSearch, and NeRM frameworks.
- Gained proficiency in PyTorch and NetworkX, establishing a methodological foundation for subsequent work on NeRM-Net and LERNA-Flow.

Selected Projects

NanoGPT Preference Alignment (DPO) for Arithmetic Tasks Oct. 2025 – Nov. 2025
Core Developer *Course Project*

- Implemented **Direct Preference Optimization (DPO)** on a NanoGPT pretrained model to empower ability in arithmetic and algebraic equation solving.
- **Pipeline:** Designed a two-stage fine-tuning pipeline (SFT + DPO) based on the provided training framework; systematically explored weight coefficients and hyperparameter configurations in a CUDA environment.
- **Data Engineering:** Constructed **100,000+ preference pairs**; generated positive samples via scripts/LLMs (with explicit reasoning steps) and negative samples from the base model.
- **Outcome:** Achieved **>90% accuracy** in arithmetic and one-variable algebra tasks, significantly outperforming the base model ($\sim$0% accuracy).

RAG-based Intelligent Chatbots Development Aug. 2024 & Oct. 2025
Tech Lead (Prompt Engineering & Retrieval Optimization) *Award-winning Projects*

- Developed "Fazhitong" (Legal Assistant) and "Shetuantong" (College Club Assistant) using RAG architecture on Dify/Coze platforms.
- **Optimization:** Proposed a sentence-structure-aware chunking strategy based on special linguistic patterns when scraping hundreds of WeChat articles, improving retrieval precision in RAG pipelines.
- **Prompt Engineering:** Iterated system prompts over 10+ versions, boosting response accuracy from 60% to **>85%** for complex multi-turn queries.
- **Awards:** Won **1st Prize** at Tsinghua "Zhili Cup" and **3rd Prize** at the 1st AI Agent Innovation Competition.

5-Stage Pipelined MIPS Processor Design June 2025 – July 2025
Independent Designer *Course Project*

- Designed a 5-stage pipeline MIPS processor using **Verilog** (Vivado); implemented hazard detection and forwarding units.
- **Feature:** Designed custom MIPS instructions and implemented software-level instruction decoding to drive 7-segment display logic, enabling sparse matrix multiplication result visualization on FPGA.
- **Performance:** Optimized area and timing, achieving a max clock frequency of **66.9MHz** with a CPI of $\sim$1.13.

Skills & Languages

- **Programming:** Python (Proficient), C/C++, MATLAB, Verilog.
- **Frameworks & Tools:** PyTorch, TensorFlow, Dify, Coze, Git, LaTeX.
- **Languages:** English (Working proficiency, Full-English Exchange Experience, IELTS Band 7.0), Mandarin (Native), Cantonese(Native).

Honors & Awards

- **3rd Prize**, 1st AI Agent Innovation Competition, Tsinghua University Oct. 2025
- **Honorable Mention**, MCM/ICM (Mathematical Contest in Modeling) Feb. 2025
- **Winner Prize (Top 30%)**, 27th Hardware Design Competition, Tsinghua University Sept. 2024
- **1st Prize & Best Sharing Award**, "Zhili Cup" Agent Competition, Tsinghua University Aug. 2024
- **1st Prize**, 7th Software Design Competition ("Bambu Lab Cup"), Tsinghua University Apr. 2024
- **Scholarships:** Ministry of Education Scholarship for Students from Hong Kong, Macao and Overseas Chinese(**3rd Prize, twice**); "Tsinghua Friend—Zheng Geru" Scholarship for Tech Innovation.