Zhuhan Bao

Zhejiang, China | zhuhan.bao@duke.edu | github.com/BaoZhuhan

Education

Hangzhou City University, B.Eng. in Software Engineering — Sep 2023 – Jul 2027
GPA: 3.77/4.0 Average: 86.29/100

Publications

SIMAX: A Scalable and Interpretable Framework for Multi-fidelity and Annotated Clinical Dialogue Simulation
Zhuhan Bao[†], Rui Yang[†], Bohao Yang, et al., Chuan Hong[*]
Working Paper (intended for submission to NEJM AI)
arXiv:2606.30491 AutoEvaluation/SIMAX
As first author, I built a workflow to generate behaviorally validated doctor-patient dialogues using large language models and TTS models. The framework supports targeted simulation based on codebook and parameter settings. I completed the preliminary research, code implementation, experiments, evaluation design, and paper writing.

Toward Realistic Evaluation of Clinical Diagnostic Reasoning through Multimodal Evidence Synthesis
Rui Yang, Weihao Xuan, **Zhuhan Bao**, Huitao Li, et al., Nan Liu[*], Yifan Peng[*]
Working Paper (not publicly indexed)
Developed the ClinMM-Bench to evaluate multi-turn multimodal clinical diagnosis, where I executed the reasoning evaluation experiments through atomic fact decomposition and comparison.

Research Experience

Visiting Student, Advisor: Prof. **Chuan Hong** — July 2025 – Present
Department of Biostatistics & Bioinformatics, Duke University School of Medicine
Duke University, Durham, NC, USA
Conduct research on clinical dialogue simulation, medical AI agents, and synthetic healthcare data generation. Lead the development of LLM-based simulation frameworks for behaviorally grounded doctor–patient interaction modeling, multimodal clinical communication analysis, and large-scale synthetic dataset generation.
In parallel, support the laboratory's computational infrastructure by maintaining unified research environments on the DCC computing cluster and conducting H200 GPU benchmarking and workload evaluation for large scale AI research deployment.

Visiting Student, Advisor: Prof. **Peiyu Liu** — May 2025 – Aug 2025
Zhejiang University NESA Research Lab
Zhejiang University, Hangzhou Zhejiang China
Systematic Security Impacts Investigation of MCP Servers in AI Code Editors
Conducted research on MCP server ecosystems and their security implications in AI-assisted code editors. Explored SDLC-oriented plugin analysis using LLMs and web crawling techniques, and developed MCP-Collector for automated MCP Marketplace data collection and classification.

Undergraduate, Advisor: Prof. **Lin Sun** — Mar 2025 – Mar 2026
The School of Computer and Computing Science
Hangzhou City University, Hangzhou Zhejiang China
Medical Multimodal Dataset Construction and Large Model Research Based on Online Teaching Videos
This project aims to construct a high-quality bilingual (Chinese-English) pathology image-text dataset from online educational videos and fine-tune the CLIP model using the Lora method to enhance the performance and accuracy of intelligent pathology diagnosis and medical knowledge retrieval. As project leader, I am mainly responsible for developing web crawlers, aligning the dataset, and fine-tuning the model.

Teaching Experience

Teaching Assistant, Advisor: Prof. **Kui Su** 26 Spring
High Performance Computing
Designed and graded assignments; led problem-solving sessions; assisted in homework evaluation. Topics included MPI, OpenMP, and parallel algorithms.

Teaching Assistant, Advisor: Dr. **Yuan Gao** 25 Fall
German Basic Writing (B08106), German Translation: Theory and Practice (C08130)
Delivered two independent teaching sessions per course; introduced fundamentals of artificial intelligence; demonstrated the use of AI agents for literary presentation and basic web development.

Project Experience

Embodied World Model Inference Optimization Jan 2026 – Mar 2026
2026 ASC Student Supercomputer Challenge
This project focused on accelerating the inference of UnifoLM-WMA-0, a state-of-the-art embodied world model, while maintaining a PSNR threshold of ≥ 25 dB. I conducted systematic performance profiling using Nsight Systems to identify bottlenecks in GEMM kernels and memory I/O. I implemented a multi-level optimization strategy including mixed precision (FP16/TF32), Flash Attention (SDPA) integration, and CUDA Graph execution to reduce CPU launch overhead. Additionally, I developed fused CUDA kernels for DDIM updates and optimized the data pipeline using pinned memory and asynchronous H2D transfers. The final pipeline achieved a $2.68\times$ speedup over the FP32 baseline, successfully passing all 20 robotic evaluation scenarios.

RNA 5-Methylcytosine Construction Optimization Jan 2025 – Feb 2025
2025 ASC Student Supercomputer Challenge
This project focuses on optimizing the detection workflow for RNA 5-methylcytosine (m5C) modification sites, aiming to enhance both precision and computational efficiency. I was responsible for designing and implementing automated scripts for data cleaning, alignment, deduplication, and statistical filtering, as well as tuning parameters to achieve high accuracy and reduced runtime. The project achieved significant improvements in automation, runtime, and memory usage, and established a robust, reproducible pipeline for m5C site detection.

Learning Experience

Competitions
1st National Service Computing Innovation Competition **Excellent Award**
2025 ASC Student Supercomputer Challenge **Second Prize**
2026 ASC Student Supercomputer Challenge **Second Prize**
16th Lanqiao Cup C/C++ Programming, Zhejiang Province **Second Prize**
15th Lanqiao Cup C/C++ Programming, Zhejiang Province **Third Prize**

Captain of School High-Performance Computing Lab Mar 2025 – Mar 2026
Hangzhou City University Supercomputing Center

Captain of the Voluntary Maintenance Team Sep 2024 – Sep 2025
Voluntary Maintenance Team of the CS School

Skills

Certificates: CET-4, CET-6, PRC Driver's License (C1)
Programming Languages: C/C++, MPI ,Python, Bash, Triton, CUDA
Skills: Infrenence Optimization, Parallel Computing, Machine Learning, Data Analysis, Algorithm Design
Software & Platforms: Linux, Slurm, Vtuner, Perf, SSH, Docker, Git, Conda, PyTorch