

Haotian Zhong

haotian7.zhong@gmail.com

EDUCATION

Tsinghua University

Dual-Degree Program

B.S. in **Mathematics and Physics**; B.Eng. in **Software Engineering**

Beijing, China

Aug. 2024 – Present

GPA: **3.73/4.00**

University of California, Berkeley, Berkeley, CA

January 2027 – May 2027

Incoming UCEAP exchange student; selected for Tsinghua University's sponsored exchange program (one seat per department)

TECHNICAL SKILLS

Programming Python, C++, SQL

ML and Research PyTorch, Hugging Face Transformers, CUDA; LLM post-training (SFT/RL), agent evaluation, contrastive learning

Systems and Data Linux, Git, NumPy, Pandas, WanDB, Docker, SQLite, Slurm, L^AT_EX

Core Coursework Object-Oriented Programming, Computer Languages and Programming, Data Structure, Algorithms, Computer Organization Principle, Database Systems, Deep Learning, Discrete Math, Probability and Statistics

Languages Mandarin Chinese (native); English (TOEFL iBT 100, Speaking 26)

RESEARCH EXPERIENCE

Research Intern – AI Watermarking and Signature Robustness

June 2026 – Present

Advisor: Prof. Yuheng Bu

Department of Computer Science, University of California, Santa Barbara

- Investigating the robustness of watermarking and signature mechanisms for AI-generated content under common post-processing and adversarial transformations.
- Building an experimental pipeline spanning watermark/signature generation, attack simulation, verification, and robustness measurement.
- Analyzing attack-defense tradeoffs and failure cases to improve the reliability of content provenance mechanisms in practical deployment settings.

Undergraduate Researcher – Defense Against Prompt Injection in LLM Agents

February 2026 – Present

First Author; Advisor: Prof. Yinpeng Dong

College of AI, Tsinghua University, Beijing

- Proposed a non-intrusive inference-time defense for indirect prompt injection by detecting intent drift between hidden states before tool calls and after tool responses.
- Built a Qwen3-8B tool-calling agent pipeline with hooks over the final 4 transformer layers, paired clean/injected trajectories, cosine and contrastive probes, and automatic attack-success labeling.
- Constructed a 1,220-task base benchmark and a 4,000-case cross-domain suite spanning 10 application domains; evaluated task-, user-tool-, attacker-, and cross-dataset generalization.
- Automated injection generation and genetic search over a 3,072-configuration strategy space; first-author manuscript currently under submission.

Research Intern – LLM Agent Security

February 2026 – May 2026

Xuanwu Lab, Tencent

Remote

- Explored early defense ideas with the team and used Tencent-provided compute resources to support large-scale experiments for the Tsinghua-led project.

AI4Science Research Assistant – Crystal-Structure LLM Post-Training

August 2025 – February 2026

Advisor: Prof. Yang Liu

Department of Computer Science and Technology, Tsinghua University

- Contributed to a 3-stage LLM post-training workflow for crystal-structure prediction: scientific-data preparation, supervised fine-tuning, and reinforcement-learning-based optimization.

- Collected, preprocessed, and cleaned crystal-structure data; aligned prompts, model inputs, and target outputs for training and downstream evaluation.
- Designed reasoning-oriented reward functions and explored reward shaping and prompt-alignment strategies to improve LLM reasoning and prediction performance.

PUBLICATION

Understanding Is Done Early: A Depth Division of Labor in Large Language Models and Its Use for Unbounded-Context Memory

July 2026

Joint Second Author

arXiv:2607.28263

- Investigated the depth division of labor in LLMs through layer-wise probes, causal truncation, and split-depth sweeps, showing that semantic information becomes accessible before upper layers specialize representations for prediction.
- Co-developed CoMem, a Qwen3-8B memory architecture that caches intermediate residual states, retrieves a fixed top-12 chunk pack, and recomputes query-conditioned upper layers using a self-distilled rank-32 LoRA.
- Evaluated five long-context and dialogue-memory benchmarks; achieved 97.05 on RULER and 38.27 on LoCoMo (vs. 34.59 for KV-Direct), while an adapter-free 128K H20 test reduced memory from 89.36 GB to 18.26 GB and delivered a $7.83\times$ prefill speedup.

PROJECTS

FinAgent OS: Multi-Agent Wealth Management System

May 2026 – June 2026

Group Project, Major Project of Software Engineering Course

- Built a LangGraph- and FastAPI-based multi-agent system orchestrating user profiling, asset screening, market timing, critic review, and compliance auditing across 54 instruments in 9 asset classes.
- Separated LLM reasoning from deterministic quantitative computation; implemented Markowitz optimization, risk parity, 3,000-path Monte Carlo simulation, VaR/CVaR, stress testing, backtesting, and performance attribution.
- Integrated AkShare and yfinance data, persistent SQLite sessions, Docker deployment, and GitHub Actions CI, maintaining a 269-test passing baseline.

TASIVE: Transformer-Based Siamese Signature Verification

November 2025 – January 2026

Lead Developer, Major Project of Deep Learning Course

- Developed a Swin-T Siamese network with a lightweight MLP projection head, L_2 -normalized 128-D embeddings, and a scaled cosine binary-cross-entropy objective.
- Designed a strict writer-disjoint train/validation split to prevent identity leakage and evaluated on a fixed all-pairs validation set of 5,748 signature pairs.
- Achieved 99.98% accuracy, 1.0000 ROC-AUC, and 0.0000 EER, outperforming a reproduced SigNet baseline under the same evaluation protocol.

AWARDS AND COMPETITIONS

Tsinghua High-Performance Computing Competition 2025

Team ranked **10th of 40 teams**