Zhongmou He

412-589-0987 | hzm20021210@gmail.com | linkedin.com/in/zhongmou-he-0739b6280

Education

Carnegie Mellon University — Aug. 2025 – Dec. 2026 (Expected)
Master of Science in Machine Learning (MSML); GPA: N/A — Pittsburgh, PA

University of Michigan — Aug. 2023 – May 2025
B.S. in Data Science (Computer Science and Engineering); Minor in Math; GPA: 3.96/4.0 — Ann Arbor, MI

Shanghai Jiao Tong University — Sept. 2021 – July 2023
B.S.E. in Electrical and Computer Engineering; Minor in Computer Science; GPA: 3.83/4.0 (top 5%) — Shanghai, China

Research Experiences

Research Assistant, CMU — Aug. 2024 – Present
Topic: LLM for Code Generation; Advisor: Prof. Lei Li — Pittsburgh, PA

- Addressed the lack of reliable test data for post-training code LLMs by developing a framework that synthesizes high-quality test cases, **improving verifier precision and recall by +11.2% and +11.0%.** Built a multi-process sandboxed execution engine with bwrap, generating 390 GB of test cases from 26.6K problems collected via large-scale web scraping.
- Validated the effectiveness of the generated tests through downstream rejection-sampling instruction fine-tuning and reinforcement learning (GRPO) experiments on Qwen3-4B with 8 A100 GPUs, **showing +2.5% pass@1 and +7.6% pass@10 gains** on LiveCodeBench.
- Currently working with an Amazon team on implementing a pipeline that generates issues, test suites, and patches from a codebase to enhance the performance of memory construction and test-time training on software engineering tasks.

Research Assistant, University of Michigan — Aug. 2023 – Aug. 2024
Topic: LLM for Link Prediction on Graphs; Advisor: Prof. Danai Koutra — Ann Arbor, MI

- Tackled the challenge of incorporating graph structure into LLMs by designing two specialized node encoders based on graph transformers and implementing a LoRA-based multimodal instruction-tuning pipeline using PyTorch.
- Achieved best performance on 6 benchmarks (*e.g.*, Amazon recommendation networks) with **up to 11% higher MRR and 13% higher Hit@1** than strong baselines, and demonstrated outstanding zero- and few-shot generalization.
- Reduced the LLM inference cost through designing a novel retrieval-reranking scheme, **achieving 10× speedup with minimal accuracy loss**, and validated scalability on graphs with 170K nodes and 1.2M edges.

Research Assistant, Shanghai Jiao Tong University — Mar. 2022 – Sept. 2023
Topic: Scientific Fundamental LLM in Geoscience; Advisor: Prof. Luoyi Fu — Shanghai, China

- Contributed to the large-scale further pre-training and fine-tuning of a LLaMA-7B model using a 5.5B-token geoscience corpus for domain adaptation and 40K curated SFT records for better alignment, requiring 214 A100 GPU hours.
- Led the evaluation of the trained model on real-world college-level geoscience exams. Designed a logit-based metric for multiple-choice questions and a perplexity-based metric for subjective questions. The model outperformed strong baselines by **+8.8% on multiple-choice and +7.8% on subjective tasks,** and demonstrated strong tool-using capability.
- This work has since received **139 citations and 204 GitHub stars**.

Publication

[ICLR 2026 (under review). 1st Author] HardTests: Synthesizing High-Quality Test Cases for LLM Coding. [paper] [code]
[CIKM 2025. 1st Author] LinkGPT: Leveraging LLMs for Enhanced Link Prediction in Text-Attributed Graphs. [paper] [code]
[CVPR 2025. 4th Author] Mosaic of Modalities: A Comprehensive Benchmark for Multimodal Graph Learning. [paper] [code]
[WSDM 2024. 3rd Author] K2: A Foundation LM for Geoscience Knowledge Understanding and Utilization. [paper] [code]

Projects

Vision Language Model (VLM)-Driven Video Replanning for Robotic Manipulation (Course Project) — Jan. 2024 – April 2024

- Addressed subgoal failures in video generation-based robotic manipulation by building a VLM reasoning and replanning module with visual feedback loops. The system detected and corrected trajectory errors, improving the success rate.

Skills

Programming Languages: Python, C/C++, MATLAB
Frameworks & Libraries: PyTorch, HuggingFace Transformers, NumPy, Pandas, Scikit-learn, LangChain, DeepSpeed, vLLM, AWS, OpenAI API, wandb, veRL, XGBoost, Slime
Tools: Git, Docker, SQL, LaTeX
Specialized Areas: LLMs, AI Agent, NLP, CV, RL, ML, Graph Learning, Deep Learning, CNN, RNN