

Pingzhi Li

Email: pingzhi@cs.unc.edu

Website: pingzhili.github.io

Google Scholar: [🔗](#)

Education

The University of North Carolina at Chapel Hill
Ph.D. in Computer Science, GPA: 4.0

Chapel Hill, NC
Sep. 2024 – Jun. 2028

University of Science and Technology of China
B.Eng. in Computer Science, GPA: 3.6

Hefei, China
Sep. 2019 – Jul. 2023

Experience

Apple AI/ML
Research Intern
Mentor: Dr. Xianzhi Du

Cupertino, CA
May – Aug. 2025

The University of North Carolina at Chapel Hill
Research Intern
Advisor: Prof. Tianlong Chen

Remote
June 2023 – June 2024

University of Science and Technology of China
Teaching Assistant
CS1001A Computer Programming A (C/C++)

Hefei, China
Sep. 2022 – Jan. 2023

University of Science and Technology of China
Undergrad Intern
Advisors: Prof. Qi Liu, Prof. Enhong Chen

Hefei, China
July 2021

Publications (Selected)

(* Equal Contribution) (^ Equal Supervision)

Occult: Optimizing Collaborative Communications across Experts for Accelerated Parallel MoE Training and Inference

Shuqing Luo, **Pingzhi Li**, Jie Peng, Yang Zhao, Yu Cao, Yu Cheng, Tianlong Chen
International Conference on Machine Learning (ICML), 2025

[\[Code\]](#) [\[PDF\]](#)

Advancing MoE Efficiency: A Collaboration-Constrained Routing (C2R) Strategy for Better Expert Parallelism Design

Mohan Zhang*, **Pingzhi Li***, Jie Peng, Mufan Qiu, Tianlong Chen

Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2025 (**SAC Award**)

[\[Code\]](#) [\[PDF\]](#)

PortLLM: Personalizing Evolving Large Language Models with Training-Free and Portable Model Patches

Rana Shahroz, **Pingzhi Li***, Sukwon Yun*, Zhenyu Wang, Shahriar Nirjon, Chau-Wai Wong, Tianlong Chen

International Conference on Learning Representations (ICLR), 2025

[\[Code\]](#) [\[PDF\]](#)

Model-GLUE: Democratized LLM Scaling for A Large Model Zoo in the Wild

Xinyu Zhao*, Guoheng Sun*, Ruisi Cai*, Yukun Zhou*, **Pingzhi Li***, Peihao Wang*, Bowen Tan, Yexiao He, Li Chen, Yi Liang, Beidi Chen, Binhang Yuan, Hongyi Wang^, Ang Li^, Zhangyang Wang^, Tianlong Chen^

Conference on Neural Information Processing Systems (NeurIPS), 2024

[\[Code\]](#) [\[PDF\]](#)

Revisiting Zeroth-Order Optimization for Memory-Efficient LLM Fine-Tuning: A Benchmark

Yihua Zhang*, **Pingzhi Li***, Junyuan Hong*, Jiayang Li*, Yimeng Zhang, Wenqing Zheng, Pin-Yu Chen, Jason D. Lee, Wotao Yin, Mingyi Hong, Zhangyang Wang, Sijia Liu, Tianlong Chen

International Conference on Machine Learning (ICML), 2024

[\[Code\]](#) [\[PDF\]](#)

Merge, Then Compress: Demystify Efficient SMoE with Hints from Its Routing Policy

Pingzhi Li, Zhenyu Zhang, Prateek Yadav, Yi-Lin Sung, Yu Cheng, Mohit Bansal, Tianlong Chen

International Conference on Learning Representations (ICLR), 2024 (Spotlight)

[\[Code\]](#) [\[PDF\]](#)

Awards

SAC Award (Low-resource Methods for NLP track), NAACL 2025	May, 2025
1st Place of ACM/IEEE Quantum Computing for Drug Discovery Challenge	Nov., 2023
Outstanding Graduates Scholarship, USTC	June, 2023
Silver Medal in Kaggle Feedback Prize - Evaluating Student Writing	March, 2022
Outstanding Student Scholarship, USTC	Nov., 2020/21/22

Services

Reviewer: NeurIPS (2024-), ICLR (2025-), ICML (2025-), CVPR (2025-), CPAL (2025-)
Tutorial Organizer: [Zeroth-Order ML \(AAAI 2024\)](#), [MoE \(ICML 2024\)](#)
Lecture: Attention & Transformers (UNC COMP 560)

Skills

Languages: Mandarin (native), English (professional), German (junior)
Programming Languages: Python, C/C++, Bash
Deep Learning: PyTorch, Jax/Flax