Fanjiang Ye

WORK fy27@rice.edu DAILY yvanphys@gmail.com
PHONE +1 (812) 322-7150 HOME home.fanjiang.net
PROFILES Google Scholar • LinkedIn

Education

Ph.D. Student in Computer Science

Rice Univeristy

2025–Present

Houston, TX, USA

Advisor: Dr. Yuke Wang

Ph.D. Student in Computer Engineering 2023–2025

Indiana University

Bloomington, IN, USA

Advisor: Dr. Dingwen Tao

B.S. in Physics 2019–2023

University of Science and Technology of China Hefei, Anhui, China

Advisor: Dr. Changling Zou

Research Experience

Rice University, Yuke's Laboratory

Graduate Research Assistant

Houston, TX, USA

Indiana University, HiPDAC Laboratory 2023–2025

Graduate Research Assistant

Bloomington, IN, USA

Hong Kong University of Science and Technology, JÄCK LaboratoryUndergraduate Research Intern

06/2022–10/2022

Kowloon, Hong Kong

University of Science and Technology of China, Zou Laboratory

2020–2023

Undergraduate Research Assistant

2020–2023

Hefei, Anhui, China

Research Interests

- Efficient systems for generative AI
- System optimization for multimodal LLMs
- Data compression and communication in HPC/ML training

Publications

- ICML'25 Spotlight Xiyuan Wei, Ming Lin, <u>Fanjiang Ye</u>, Fengguang Song, Liangliang Cao, My T. Thai, Tianbao Yang. Model Steering: Learning with a Reference Model Improves Generalization Bounds and Scaling Laws. <u>arXiv</u>
- ICS'25 Best Paper Runner-up Boyuan Zhang, Bo Fang, Fanjiang Ye, Luanzheng Guo, Fengguang Song, Tallent Nathan, Dingwen Tao. BMQSim: Overcoming Memory Constraints in Quantum Circuit Simulation with a High-Fidelity Compression Framework.
- SC'24 Hao Feng, Boyuan Zhang, <u>Fanjiang Ye</u>, Min Si, Ching-Hsiang Chu, Jiannan Tian, Chunxing Yin, Zhaoxia (Summer) Deng, Yuchen Hao, Pavan Balaji, Tong Geng, Dingwen Tao. Accelerating Communication in Deep Learning Recommendation Model Training with Dual-Level Adaptive Lossy Compression. <u>doi</u>
- PPoPP'25 Poster Boyuan Zhang, Luanzheng Guo, Jiannan Tian, Jinyang Liu, Daoce Wang, Fanjiang Ye, Chengming Zhang, Jan Strube, Nathan R. Tallent, Dingwen Tao. High-performance Visual Semantics Compression for Al-Driven Science. doi
- Preprint Fanjiang Ye, Zepeng Zhao, Yi Mu, Jucheng Shen, Renjie Li, Kaijian Wang, Desen Sun, Saurabh Agarwal, Myungjin Lee, Triston Cao, Aditya Akella, Arvind Krishnamurthy, T. S. Eugene Ng, Zhengzhong Tu, Yuke Wang. SUPERGEN: An Efficient Ultra-high-resolution Video Generation System with Sketching and Tiling. arXiv
- Preprint Xinrui Zhong, Xinze Feng, Jingwei Zuo, Fanjiang Ye, Yi Mu, Junfeng Guo, Heng Huang, Myungjin Lee, Yuke Wang. An Efficient and Adaptive Watermark Detection System with Tile-based Error Correction. arXiv

• Preprint Xiyuan Wei, Fanjiang Ye, Ori Yonay, Xingyu Chen, Dingwen Tao, Tianbao Yang. FastCLIP: A Suite of Optimization Techniques to Accelerate CLIP Training with Limited Resources.

Honors and Awards

Indiana University Travel Awards (\$1500), Indiana University Bloomington
 USTC Fellowship (\$2000), University of Science and Technology of China

Outstanding Student Scholarship (Top 25%), University of Science and Technology of China
 2020, 2021, 2022

Professional Service

- ASPLOS'26 Artifact Evaluation Committee
- SOSP'25 Artifact Evaluation Committee
- QCE'24 Sub-Reviewer

Teaching Experience

• Teaching Assistant of ENGR-E 516: Cloud Computing, Spring 2025, Indiana University