Yixiao Li

H. Milton School of Industrial and Systems Engineering
Georgia Institute of Technology, 755 Ferst Drive, Atlanta, GA 30332

E-mail: yixiaoli@gatech.edu

Research Interest

My research focuses on applications of machine learning algorithms. I am interested in large language models (LLMs) and their efficiency. I am also interested in data synthesis for LLM post-training.

Education

2022 - present Georgia Institute of Technology
Ph.D. Candidate in Machine Learning
Expected to graduate by May 2026

2018 - 2022 The University of Science and Technology of China (USTC)
B.Eng. in Electronic Information Engineering

Publications

- **LoftQ: LoRA-Fine-Tuning-Aware Quantization for Large Language Models**
  Yixiao Li*, Yifan Yu*, Chen Liang, Nikos Karampatziakis, Pengcheng He, Weizhu Chen and Tuo Zhao
  International Conference on Learning Representations (ICLR), 2024
- **LoSparse: Structured Compression of Large Language Models based on Low-Rank and Sparse Approximation**
  Yixiao Li*, Yifan Yu*, Qingru Zhang, Chen Liang, Pengcheng He, Weizhu Chen and Tuo Zhao
  International Conference on Machine Learning (ICML), 2023
- **Adaptive Preference Scaling for Reinforcement Learning with Human Feedback**
  Ilgee Hong, Zichong Li, Alexander Bukharin, **Yixiao Li**, Haoming Jiang, Tianbao Yang, Tuo Zhao
  Conference on Neural Information Processing Systems (NeurIPS), 2024
- **Deep Reinforcement Learning from Weak Hierarchical Preference Feedback**
  Alexander Bukharin, **Yixiao Li**, Pengcheng He, Weizhu Chen and Tuo Zhao
  International Conference on Machine Learning (ICML), 2025
- **NoWag: A Unified Framework for Shape Preserving Com- pression of Large Language Models**
  Lawrence Ray Liu, Inesh Chakrabarti, **Yixiao Li**, Mengdi Wang, Tuo Zhao, Lin Yang
  Conference on Language Modeling (COLM), 2025
- **IDEA Prune: An Integrated Enlarge-and-Prune Pipeline in Generative Language Model Pre-training**
  Yixiao Li, Xianzhi Du, Ajay Kumar Jaiswal, Tao Lei, Tuo Zhao, Chong Wang, Jianyu Wang
  Under Review
- **Finding Fantastic Experts in MoEs: A Unified Study for Expert Dropping Strategies and Observations**
  Ajay Kumar Jaiswal, Jianyu Wang, **Yixiao Li**, Pingzhi Li, Tianlong Chen, Zhangyang Wang, Chong Wang, Ruoming Pang, Xianzhi Du
  Under Review

Professional Experiences

2025.8 - Present **Applied Scientist Intern, Amazon SFAI, Part-time**

2025.5 - 2025.8 **Applied Scientist Intern, Amazon SFAI, Full-time**

Weakness-Driven Prompt Synthesis for Small Language Models in Post-Training: identified small model's weaknesses by a language model as a verifier; controlled prompt difficulty by a teacher model; showed better performance than including more generic data; demonstrated synthesis efficiency than data selection.

2024.12 - 2025.5 **Machine Learning Research Intern, Apple AIML, Part-time**

Tanh-based Attention for KV Head Compression: explored differentiable attention mechanism in pretraining; designed tanh-based attention to allow negative attention scores to compress KV heads.

2024.5 - 2024.12 **Machine Learning Research Intern, Apple AIML, Full-time**

Enlarge-and-Prune Pipeline for Small Language Model Pretraining: developed a small language model training algorithm by enlarge the small model and then prune; studied the token efficiency of the pruning; showed our algorithm scalable (up to 2T tokens) in pretraining; demonstrate higher token efficiency compared to naively increasing tokens.

2023.5 - 2024.5 **Research Intern, Microsoft Azure, Part-time**

LoRA-Fine-Tuning-Aware Quantization: quantized the backbone weights and fine-tuned the model with LoRA; designed an iterative algorithm to initialize the quantized weight and LoRA adapter; demonstrated better performance of DeBERTa, BART, and LLAMA.

Teaching Experiences

Teaching Assistant:

- Computer Programming with C Language, Fall 2021
- Computational Data Analysis, ISYE 6740, Spring 2023
- Simulation, ISYE 6644, Summer 2023
- Foundations and Applications of Machine Learning, ISYE 4803, Fall 2023
- Regression Analysis, ISYE 6414, Fall 2025

References

Tuo Zhao Associate Professor
H. Milton School of Industrial and Systems Engineering
Georgia Institute of Technology
Email: tourzhao@gatech.edu