Lichen LI

Email: lilichen@stu.pku.edu.cn   Phone: (+86)18201548082

RESEARCH INTERESTS

➢ Multimodal AI, Embodied AI, AI Reasoning, Analysis & Explainability of AI Systems, Efficient AI, Computer Vision.

EDUCATION

Yuanpei College, Peking University (PKU), Beijing, China   Sep 2021 – Jun 2026 (expected)

B.S. in Data Science and Big Data Technology, Minor in Philosophy

Overall GPA: 3.5/4.0 (85/100)

Relevant Coursework: Introduction to Computer Systems (88), Machine Learning (91.5), Introduction to AI (90), Introduction to Computer Vision (93), Introduction to LLMs and Alignment (90.4), Data Structure and Algorithm, Algorithm Design and Analysis, Computer Networks, Computer Architectures, Probability Theory, Mathematical Statistics

Honors & Awards: Peking University Merit Student (2025)

PUBLICATIONS

➢ Qizhe Zhang, Mengzhen Liu, **Lichen Li**, et al. "*Beyond Attention or Similarity: Maximizing Conditional Diversity for Token Pruning in MLLMs*", NeurIPS 2025. arXiv:2506.10967.

RESEARCH EXPERIENCE

Maximizing Conditional Diversity for Token Pruning in MLLMs | PKU, China   Mar 2025 – Jun 2025

Research Assistant | Advisor: Assistant Professor Shanghang Zhang, School of Computer Science

➢ Contributed substantially to a NeurIPS 2025 paper proposing CDPruner, a training-free, model-agnostic visual token pruning method for multimodal LLMs addressing the limitations of attention-based and similarity-based pruning.

➢ Co-designed the core algorithm, reformulating token pruning as conditional diversity maximization, leveraging determinantal point process with a greedy algorithm for efficient subset selection.

➢ Applied multiple pruning baselines to a unified evaluation framework and conducted extensive experiments, achieving performance gains to support paper findings.

Reward Hacking Monitor in Reasoning LLMs | UCLA & University of Maryland   Jul 2025 – Present

Research Assistant | Advisors: Associate Professor Cho-Jui Hsieh and Assistant Professor Tianyi Zhou

➢ Leading a project on reward hacking monitoring in reasoning-based coding tasks, addressing a gap in current misalignment research; planned submission to ACL 2026.

➢ Developed a framework applicable to both inference and training to effectively elicit, amplify, and trace genuine in-the-wild hacking behaviors, overcoming prior reliance on indirect hacking definition/synthetic hacking trajectories.

➢ Developing a robust monitor, trained on trajectories collected in the wild to identify emergent hacking behaviors.

Enhancing Performance of VFL under DP Privacy | Johns Hopkins University, MD   Jun 2024 – Aug 2024

Research Assistant | Advisor: Associate Professor Yinzhi Cao, Department of Computer Science

➢ Led a project adapting optimization techniques from Horizontal to Vertical Federated Learning (VFL). Developed a server-side transformation layer to improve model performance under differential privacy constraints.

➢ Reviewed multiple open-source VFL frameworks and refined one to improve stability and training efficiency, achieving hands-on experience with privacy-preserving FL systems.

Research on Parameter-Efficient Fine-Tuning of Vision Transformer | PKU, China   Mar 2024 – Jun 2024

Research Assistant | Advisor: Assistant Professor Shanghang Zhang, School of Computer Science

➢ Investigated various parameter-efficient fine-tuning methods of ViT, including VPT, AdaptFormer, LoRA, Scaling & Shift Features, Gradient-based Parameter Selection; reviewed various training-free tuning methods.

➢ Analyzed theoretical and empirical factors behind the effectiveness of PEFT methods across diverse applications.

COURSE PROJECT EXPERIENCE

"Introduction to LLMs and Alignment" | PKU, China   Mar 2025 – Jun 2025

Advisor: Boya Assistant Professor Yaodong Yang, Institute for Artificial Intelligence, PKU

➢ Constructed a mini LLM stack from scratch, including BPE tokenizer, transformer, and training pipeline.

➢ Applied LoRA on the Alpaca dataset with hyperparameter analysis and model output evaluation.

➢ Developed preference-based alignment models, implementing reward modeling and DPO on human preference datasets; conducted behavioral analysis and visualization of model responses.

"Introduction to Computer Vision" Course Projects | PKU, China   Mar 2024 – Jun 2024

Advisor: Assistant Professor He Wang, Center on Frontiers of Computing Studies

➢ Constructed and trained a CNN from scratch for CIFAR-10 classification, including hand-coded convolution function, batch normalization (written from scratch) and data augmentation; developed an RNN for image captioning.

➢ Implemented camera calibration for 3D vision, transforming depth images to point clouds, sampling point clouds from meshes and implemented a Marching Cube; implemented and trained a PointNet for classification and segmentation.

VAE for Processing Flower Images | PKU, China   Dec 2023 – Jan 2024

Advisor: Assistant Professor Muhan Zhang, Institute for Artificial Intelligence

➢ Designed a CNN-based VAE for encoding, reconstructing, and generating flower images.

➢ Optimized model architecture, hyperparameters, and optimizer selection based on experimental results.

"Introduction to Computer System" Course Projects | PKU, China   Oct 2022 – Dec 2022

Advisor: Professor Xiangqun Chen, Institute of Software

➢ Designed and implemented a dynamic memory allocator with segregated-fit implicit free lists, a Linux shell, and a caching web proxy supporting concurrency.

➢ Optimized performance through experimental analysis.

Ataxx | PKU, China   Oct 2021 – Jan 2022

Advisor: Professor Houfeng Wang, Institute of Computational Linguistics

➢ Designed an interactive Ataxx bot, ranking in the top 3% among 400 entries in the course tournament.

LEADERSHIP

Class Monitor for Class 4 of 2025, Yuanpei College   Sep 2022 – Sep 2023

➢ Optimized class committee structure and organized academic and social events

➢ Achieved class selection as one of PKU's 37 Demonstration Classes in 2023 (out of ~1000 entries).

SKILLS

Technical Stack: C/C++, Python; Numpy, Pytorch, Tensorflow, Transformers

Research Experience: Multimodal AI, AI Reasoning, Efficient AI, Computer Vision, Analysis & Explainability of AI Systems.

English Language: TOEFL: Total 105 (Reading 26, Listening 27, Speaking 25, Writing 27)