Lichen Li

Room 341, Building 35#, Peking University, Beijing, 100871, P. R. China
(+86)18201548082 lilichen@stu.pku.edu.cn

RESEARCH INTERESTS

- Multimodal AI, Embodied AI, AI Reasoning, Reinforcement Learning, Analysis & Explainability of AI Systems, Efficient Training & Deployment, Computer Vision.

EDUCATION

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

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

- **GPA:** 3.5/4.0
- **Core Courses:** 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, etc.
- **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 | Research Assistant Mar 2025 – Jun 2025

Advisor: Shanghang Zhang, Assistant Professor at the School of Computer Science, PKU

- Made substantial contribution 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 dynamic programming for efficient subset selection.
- Integrated multiple pruning baselines into a unified evaluation framework and conducted extensive experiments, validating performance gains to support paper findings.

Reward Hacking Monitor in Reasoning LLMs (ongoing) **|** Research Assistant Jul 2025 – Present

Advisor: Cho-Jui Hsieh, Associate professor, UCLA; Tianyi Zhou, Assistant Professor, University of Maryland

- Leading a project on reward hacking monitoring in reasoning-based coding tasks, filling a gap in current misalignment research; planned submission to ACL 2026.
- Introduced a more intuitive and behaviorally grounded criterion for identifying reward hacking, together with an automated framework for tracing such behaviors, addressing the limitations of prior indirect approaches.
- Built a framework for both inference and training to elicit in-the-wild hacking behaviors under this new criterion, and are developing a CoT-based monitor trained on collected trajectories to identify emergent hacking behaviors.

Enhancing Performance of VFL under DP Privacy | Research Assistant. Jun 2024 – Aug 2024

Advisor: Yinzhi Cao, Associate Professor at the Department of Computer Science, Johns Hopkins University

- Led a project adapting optimization techniques from Horizontal to Vertical Federated Learning (VFL) and developed a server-side transformation layer to improve model performance under Differential Privacy constraints.
- Surveyed multiple open-source VFL frameworks and refined one to improve stability and training efficiency, gaining

hands-on experience with privacy-preserving FL systems.

Research on Parameter-efficient Fine-tuning of Vision Transformer | Research Assistant Mar 2024 – Jun 2024

Advisor: Shanghang Zhang, Assistant Professor at the School of Computer Science, PKU

- Investigated various parameter-efficient fine-tuning methods of ViT, including VPT, AdaptFormer, LoRA, Scaling & Shift Features, Gradient-based Parameter Selection, etc. Investigated 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" Course Projects Mar 2025 – Jun 2025

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

- **Built 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 Mar 2024 – Jun 2024

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

- **NN**: Built and trained a CNN from scratch for CIFAR-10 classification, including hand-coded convolution function, batch normalization (written from scratch) and data augmentation. Built and trained a RNN for image captioning.
- **3D Vision**: Implemented camera calibration, transforming depth images to point clouds, sampling point clouds from meshes and implemented Marching Cube. Implemented and trained a PointNet for classification and segmentation.

VAE for Processing Flower Images | Course Project Dec 2023 – Jan 2024

Advisor: Muhan Zhang, Assistant Professor at the Institute for Artificial Intelligence, PKU

- 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 Oct 2022 – Dec 2022

Advisor: Xiangqun Chen, Professor at the Institute of Software, PKU

- 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 | Course Project Oct 2021 – Jan 2022

Advisor: Houfeng Wang, Professor at the Institute of Computational Linguistics, PKU

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

LEADERSHIP & ACTIVITIES

Class 4 of 2025, Yuanpei College, PKU **|** Class Monitor Sep 2022 – Sep 2023

- Optimized class committee structure and organized academic and social events; the class was selected as one of PKU's 37 Demonstration Classes in 2023 (out of ~1000 classes).

SKILLS

Programming Languages: C/C++, Python.

Frameworks and Libraries: Numpy, Pytorch, Tensorflow, Transformers.

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