

# YEFAN ZHOU

yefan.zhou.gr@dartmouth.edu | Homepage [↗](#) | Google Scholar [↗](#) | LinkedIn [↗](#) | Hanover, NH, 03755 | 510-809-5378

## EDUCATION

---

### Dartmouth College

*Ph.D. student in Computer Science*

Hanover, NH

*Sep. 2023 – present*

Advisor: Prof. Yaoqing Yang

Research Area: Efficient training, LLM pruning/fine-tuning/mixture-of-expert, model diagnostic

### University of California, Berkeley

Berkeley, CA

*M.Eng in Electrical Engineering and Computer Science; Major GPA: 4.0/4.0*

*Aug. 2021 – Dec. 2022*

Advisor: Prof. Michael Mahoney

Research Area: Pruning for model efficiency

### University of California, Berkeley

Berkeley, CA

*Exchange Student; GPA: 4.0/4.0*

*Jan. 2019 – May. 2019*

### Southeast University

China

*B.Eng in Information Engineering; GPA: 3.7/4*

*Aug. 2016 – Jun. 2020*

## PUBLICATION

---

### *Selected first-author paper:*

- {H. Lu\*, **Y. Zhou\***}, S. Liu, Z. Wang, M. W. Mahoney, Y. Yang “AlphaPruning: Using Heavy-Tailed Self Regularization Theory for Improved Layer-wise Pruning of Large Language Models” (NeurIPS 2024)
- {H. Lu\*, X. Liu\*, **Y. Zhou\***, Q. Li\*}, H. Yang, Y. Yan, K. Keutzer, M. W. Mahoney, Y. Yang “Sharpness-diversity tradeoff: improving flat ensembles with SharpBalance” (NeurIPS 2024)
- {**Y. Zhou\***, J. Chen\*}, Q. Cao, K. Schürholt, Y. Yang “MD tree: a model-diagnostic tree grown on loss landscape” (ICML 2024)
- {**Y. Zhou\***, T. Pang\*}, K. Liu, C. H. Martin, M. W. Mahoney, Y. Yang “Temperature Balancing, Layer-wise Weight Analysis, and Neural Network Training” (NeurIPS 2023 Spotlight)
- **Y. Zhou**, Y. Yang, A. Chang, M. W. Mahoney “A Three-regime model of Network Pruning” (ICML 2023)
- **Y. Zhou**, Y. Shen, Y. Yan, C. Feng, Y. Yang “A Dataset-Dispersion Perspective on Reconstruction Versus Recognition in Single-View 3D Reconstruction Networks” *2021 International Conference on 3D Vision (3DV 2021)*

### *Collaborating or advising paper:*

- {Z. Liu\*, Y. Hu\*}, T. Pang, **Y. Zhou**, P. Ren, Y. Yang “Model Balancing Helps Low-data Training and Fine-tuning” (EMNLP 2024 main)
- P. Qing, C. Gao, **Y. Zhou**, X. Diao, Y. Yang, S. Vosoughi “AlphaExpert: Assigning LoRA Experts Based on Layer Training Quality” (EMNLP 2024 main)
- X. Zhu, **Y. Zhou**, Y. Fan, J. Chen, M. Tomizuka “Learn to Grasp with Less Supervision: A Data-Efficient Maximum Likelihood Grasp Sampling Loss” *2022 International Conference on Robotics and Automation (ICRA 2022)*
- K. Schürholt, L. Meynert, **Y. Zhou**, Y. Yang, D. Borth “A Model Zoo on Phase Transitions in Neural Networks” (Preprint)

## PROFESSIONAL EXPERIENCE

---

### **Research Engineer, International Computer Science Institute**

Berkeley, CA

*supervised by Prof. Michael Mahoney*

*Jan. 2023 – Jun. 2023*

- Researched efficient optimization method for deep neural network.
- Researched ensembling methods for improving the OOD robustness of CV models.
- Developed backdoor detection methods to enhance AI model safety.

### **Graduate Research Assistant, Sky Computing Lab (RISELab), UC Berkeley**

Berkeley, CA

*advised by Prof. Michael Mahoney*

*Aug. 2021 – Dec. 2022*

- Researched neural network pruning for CNNs and Transformers.

## SERVICES AND AWARD

---

**Reviewers:** ICLR 2025-2024, NeurIPS 2023, AAAI 2024, ICML 2024, CVPR 2024, CPAL 2024, IROS 2022, TMLR

### **Talk**

- \* Invited talk at AI-TIME, “Phase transition, loss landscape and model diagnostics”, Jan., 2024.
- \* Invited talk at UC Berkeley/ICSI TrojAI onsite, “Layer-wise Weight Analysis, and Neural Network Training” Oct. 2023
- \* Invited talk at UC Berkeley/ICSI TrojAI onsite, “A Three-regime model of Network Pruning” Mar. 2023

**Award:** ICML 2024 Scholar Award, NeurIPS 2023 Scholar Award

**Teaching (Head TAs):** CS70: Foundations of Applied Computer Science (Dartmouth College Spring 2024)

## SKILLS

---

**Programming Language:** Python, Java, C/C++, CUDA, SQL, MATLAB

**Developer Tools:** PyTorch, Ubuntu, MuJoCo, ROS, PyBullet, Slurm, PyRender, Open3D