Yihua Zhang

Ph.D. Student in Computer Science

☎ (+1) 517-980-3880 ✉ zhan1908@msu.edu 🌐 www.yihua-zhang.com NormalUhr G Yihua Zhang

Personal Information

I am a third-year Ph.D. student in computer science at Michigan State University advised by **Dr. Sijia Liu**. I am interested in developing **trustworthy and efficient foundation models** by advancing their optimization foundations, including the optimization theories to improve the robustness, alignment, privacy, and scalability of the current machine learning algorithms.

Education

Doctor of Computer Science 01 2022 — Present
Michigan State University, East Lansing, USA
Advisor: Dr. Sijia Liu
OPTML Lab

Bachelor of Engineering 09 2015 — 06 2019
Huazhong University of Science and Technology, Wuhan, China

Publications

Preprint Paper

[P1] **Y. Zhang**, Y. Zhang, Y. Yao, J. Jia, X. Liu, S. Liu "**UnlearnCanvas: A Stylized Image Dataset to Benchmark Machine Unlearning for Diffusion Models and Beyond**", [PDF], [Code], [Website], [Video], [Dataset], [Benchmark].

[P2] **Y. Zhang***, H. Li*, Y. Yao*, A. Chen, P.-Y. Chen, S. Zhang, M. Wang, S. Liu "**Visual Prompting Reimagined: The Power of Activation Prompts**", [PDF].

[P3] Y. Zhang, J. Jia, X. Chen, A. Chen, **Y. Zhang**, J. Liu, K. Ding, S. Liu "**To Generate or Not? Safety-Driven Unlearned Diffusion Models Are Still Easy To Generate Unsafe Images . . . For Now**", [PDF].

Journal Paper

[J1] **Y. Zhang**, P. Khanduri, I. Tsaknakis, Y. Zhang, M. Hong, S. Liu "**An Introduction to Bi-level Optimization: Foundations and Applications in Signal Processing and Machine Learning**", IEEE Signal Processing Magazine, vol. 41, no. 1, pp. 38-59, 2024 (**Feature Article**), [PDF].

[J2] H. Li, S. Zhang, **Y. Zhang**, M. Wang, S. Liu, P.-Y. Chen, "**How Does Promoting the Minority Fraction Affect Generalization? A Theoretical Study of One-Hidden-Layer Neural Network on Group Imbalance**", IEEE Journal of Selected Topics in Signal Processing, 2024, [PDF].

Conference Papers

(* represents equal contributions and † represents the student mentored by me.)

[C1] **Y. Zhang***, P. Li*, J. Hong*, J. Li, Y. Zhang, W. Zheng, P.-Y. Chen, J. Lee, W. Yin, M. Hong, Z. Wang, S. Liu, and T. Chen "**Revisiting Zeroth-Order Optimization for Memory-Efficient LLM Fine-Tuning: A Benchmark**", The Forty-first International Conference on Machine Learning (ICML'24), [PDF], [Code], [Website].

[C2] **Y. Zhang**, Y. Zhang, A. Chen, J. Jia, J. Liu, G. Liu, S. Chang, M. Hong, S. Liu "**Selectivity Drives Productivity: Efficient Dataset Pruning for Enhanced Transfer Learning**", 37th Conference on Neural Information Processing Systems (NeurIPS'23), [PDF], [Code], [Website].

[C3] **Y. Zhang**, R. Cai, T. Chen, G. Zhang, P. Chen, H. Zhang, S. Chang, W. Zhang, S. Liu "**Robust Mixture-of-Expert Training for Convolutional Neural Networks**", International Conference on Computer Vision 2023 (ICCV'23 - **Oral**), [PDF], [Code], [Poster].

[C4] **Y. Zhang**, P. Sharma, P. Ram, M. Hong, K. R. Varshney, S. Liu "**What Is Missing in IRM Training and Evaluation? Challenges and Solutions**", 11th International Conference on Learning Representations (ICLR'23), [PDF], [Poster].

[C5] C. Fan†, J. Liu, **Y. Zhang**, E. Wong, D. Wei, S. Liu "**Salun : Empowering Machine Unlearning via Gradient-based Weight Saliency in Both Image Classification and Generation**", 12th International Conference on Learning Representations (ICLR'24 - **Spotlight**), [PDF], [Code], [Poster].

[C6] A. Chen†, Y. Zhang, J. Jia, J. Diffenderfer, J. Liu, K. Parasyris, **Y. Zhang**, Z. Zhang, B. Kailkhura, S. Liu "**DeepZero: Scaling up Zeroth-Order Optimization for Deep Model Training**", 12th International Conference on Learning Representations (ICLR'24), [PDF], [Code], [Poster].

[C7] B. Hou, **Y. Zhang**, J. Jia, G. Zhang, Y. Zhang, S. Liu, S. Chang "**TextGrad: Advancing Robustness Evaluation in NLP by Gradient-Driven Optimization**", 11th International Conference on Learning Representations (ICLR'23), [PDF], [Code].

[C8] P. Khanduri, I. Tsaknakis, **Y. Zhang**, J. Liu, S. Liu, J. Zhang, M. Hong "**Linearly Constrained Bilevel Optimization: A Smoothed Implicit Gradient Approach**", 40th International Conference on Machine Learning (*ICML'23*), [PDF].

[C9] A. Chen†, Y. Yao, P.-Y. Chen, **Y. Zhang**, S. Liu, "**Understanding and Improving Visual Prompting: A Label-Mapping Perspective**", 2023 Conference on Computer Vision and Pattern Recognition (CVPR'23), [PDF], [Code].

[C10] H. Zhuang†, **Y. Zhang**, S. Liu, "**A Pilot Study of Query-Free Adversarial Attack against Stable Diffusion**", 2023, Conference on Computer Vision and Pattern Recognition (CVPR'23), [PDF], [Code].

[C11] **Y. Zhang***, Y. Yao*, P. Ram, P. Zhao, T. Chen, M. Hong, Y. Wang, S. Liu, "**Advancing Model Pruning via Bi-level Optimization**", 36th Conference on Neural Information Processing Systems (NeurIPS'22), [PDF], [Code], [Poster], [Project Website].

[C12] **Y. Zhang***, G. Zhang*, Y. Zhang, W. Fan, Q. Li, S. Liu, S. Chang, "**Fairness Reprogramming**", 36th Conference on Neural Information Processing Systems (NeurIPS'22), [PDF], [Code], [Poster], [Project Website].

[C13] G. Zhang*, S. Lu*, **Y. Zhang**, X. Chen, P. Chen, Q. Fan, L. Martie, M. Hong , S. Liu, "**Distributed Adversarial Training to Robustify Deep Neural Networks at Scale**", 38th Conference on Uncertainty in Artificial Intelligence (*UAI'22 - **Oral, Best Paper Runner-up Award***), [PDF], [Code], [Poster], [Slides], [Award].

[C14] **Y. Zhang***, G. Zhang*, P. Khanduri, M. Hong, S. Chang, S. Liu, "**Fast-BAT: Revisiting and Advancing Fast Adversarial Training through the Lens of Bi-level Optimization**", 39th International Conference on Machine Learning (*ICML'22*), [PDF], [Code], [Poster], [Slides], [Talk].

[C15] T. Chen*, Z. Zhang*, **Y. Zhang***, S. Chang , S. Liu , Z. Wang "**Quarantine: Sparsity Can Uncover the Trojan Attack Trigger for Free**", Computer Vision and Pattern Recognition Conference 2022 (*CVPR'22*), [PDF], [Code], [Poster], [Project Website].

Industrial Experience

Research Scientist Intern 12/2023 – Present
Cisco Research *Remote*

- Project: A general machine unlearning solution for foundation models: Large Language Models (LLMs), Diffusion Models (DMs), and Mixture-of-Experts (MoEs).

Applied Scientist Intern 05/2023 – 08/2023
AWS AI Lab *Seattle, US*

- Project: In-context learning for vision generative models: design, training, and generalization study.
- Mentor: Zhou (Joe) Ren

Research Intern 01 2021 – 08 2021
JD AI Research (JD Explore Academy) *Beijing, China*

- Project: Model robustness, fairness, and explanability co-design.
- Mentor: Dr. Jinfeng Yi.

Awards

Scholarly Awards

• ML and Systems Rising Stars sponsored by NVIDIA	2024
• CVPR Outstanding Reviewer	2023
• Best Paper Runner-up Award of UAI 2022	2022
• NeurIPS Top Reviewer	2022
• NeurIPS Top Reviewer	2023
• UAI Student Scholarship	2022

Conference Scholar Award

- NeurIPS Scholar Award 2022, 2023
- AAAI 2023 Travel Award 2023
- Travel Grant Award of ICML 2022 2022

Undergraduate Award

- National Scholarship, by Ministry of Education of China (Top 1%, highest undergraduate honor) 2017
- National Scholarship, by Ministry of Education of China (Top 1%, highest undergraduate honor) 2016

Tutorials/Talks

- **02/2024**: “Zeroth-Order Machine Learning: Fundamental Principles and Emerging Applications in Foundation Models”, **AAAI 2024 (Tutorial)**
- **02/2023**: “Bi-level Optimization in Machine Learning: Foundations and Applications”, **AAAI 2023 (Tutorial)**
- **11/2022**: “Invariant Risk Minimization through Bi-level Optimization and Beyond”, **Invited Talk at UMN**
- **10/2022**: “Revisiting and Advancing Fast Adversarial Training through the Lens of Bi-level Optimization”, **Invited Talk at INFORMS Annual Meeting (2022)**
- **04/2022**: “Adversarial Training via Bi-level Optimization”, **Invited Talk at UCSB**.

Professional Activities

- **Volunteer**: AAAI’23, ICLR’23
- **Reviewer**: NeurIPS, ICML, AISTATS, ICLR, ICASSP, ICCV, CVPR, UAI, T-PAMI, T-IFS, TMRL
- **Student Chair** for the ICML Workshop AdvML: New Frontiers in Adversarial Machine Learning in 2022 and 2023.

Mentorship

- **Mentee: Haomin Zhuang**:
  - *Role:* Undergraduate at South China University of Technology, China
  - *Mentoring Period:* Oct. 2022 - Mar. 2023
  - *Project:* A Pilot Study of Query-Free Adversarial Attack against Stable Diffusion (Paper)
  - *Conference:* The 3rd Workshop of Adversarial Machine Learning on Computer Vision@CVPR’23 (Website)
  - *Current Position:* Ph.D. student at University of Notre Dame
- **Mentee: Aochuan Chen**:
  - *Role:* Undergraduate at Tsinghua University, China
  - *Mentoring Period:* Aug. 2022 - Nov. 2023
  - *Projects:*
    - Understanding and Improving Visual Prompting: A Label-Mapping Perspective (CVPR 2023 Paper)
    - DeepZero: Scaling up Zeroth-Order Optimization for Deep Model Training (ICLR 2024 Paper)
  - *Current Position:* Ph.D. student at Hong Kong University of Science and Technology
- **Mentee: Chongyu Fan**:
  - *Role:* Undergraduate at Huazhong University of Science and Technology, China

  - *Mentoring Period:* May 2023 - Present
  - *Project:* SalUn: Empowering Machine Unlearning via Gradient-based Weight Saliency in Both Image Classification and Generation
  - *Conference:* ICLR 2024 Spotlight
  - *Current Position:* Ph.D. student at MSU OPTML Group
- **Mentee: Mohammad Jafari**:
  - *Role:* Undergraduate at Sharif University of Technology, Iran
  - *Mentoring Period:* May 2023 - Oct. 2023
  - *Project:* The Power of Few: Accelerating and Enhancing Data Reweighting with Coreset Selection (ICASSP 2024 Paper)

Last updated: June 6, 2024.