

# YIHUA ZHANG

## Ph.D. Student in Computer Science

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

### PERSONAL INFORMATION

I am a first-year Ph.D. student in computer science at Michigan State University, where I am advised by **Dr. Sijia Liu**. I am interested in the optimization foundation of **trustworthy and scalable machine learning**, including the optimization theories to improve the robustness, explainability, fairness, and scalability of current machine learning algorithms.

### EDUCATION

#### Doctor of Computer Science

Michigan State University, East Lansing, USA

Advisor: Dr. Sijia Liu

OPTML Lab

01 2022 — Present

#### Bachelor of Engineering

Huazhong University of Science and Technology, Wuhan, China

09 2015 — 06 2019

### AWARDS

#### Research Awards and Scholarships

- Best Paper Runner-up Award of UAI 2022

2022

#### Travel Grants

- AAAI 2023 Travel Award
- NeurIPS Scholar Award
- NeurIPS Top Reviewer
- UAI Student Scholarship
- Travel Grant Award of ICML 2022

2023

2022

2022

2022

2022

#### Undergraduate Award

- National Scholarship, by Ministry of Education of China (Top 2%, 5/300)
- National Scholarship, by Ministry of Education of China (Top 2%, 5/300)

2017

2016

### PUBLICATIONS

#### Conference Papers

- [1] **Yihua Zhang\***, Yuguang Yao\*, Parikshit Ram, Pu Zhao, Tianlong Chen, Mingyi Hong, Yanzhi Wang, Sijia Liu, "[Advancing Model Pruning via Bi-level Optimization](#)", 36th Conference on Neural Information Processing Systems (NeurIPS'22), [PDF], [Code].
- [2] Guanhua Zhang\*, **Yihua Zhang\***, Yang Zhang, Wenqi Fan, Qing Li, Sijia Liu, Shiyu Chang "[Fairness Reprogramming](#)", 36th Conference on Neural Information Processing Systems (NeurIPS'22), [PDF].
- [3] 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].
- [4] **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].
- [5] 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].

#### Papers under Submission

- [6] **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](#)", submitted to CVPR 2023.
- [7] **Y. Zhang**, P. Sharma, P. Ram, M. Hong, K. R. Varshney, S. Liu "[What Is Missing in IRM Training and Evaluation? Challenges and Solutions](#)", [link], submitted to ICLR 2023 (rating 6.67).

- [8] B. Hou, J. Jia, **Y. Zhang**, G. Zhang, Y. Zhang, S. Liu, S. Chang "[TextGrad: Advancing Robustness Evaluation in NLP by Gradient-Driven Optimization](#)", [link], submitted to ICLR 2023 (rating 6.25).
- [9] P. Khanduri, I. Tsaknakis, **Y. Zhang**, J. Liu, S. Liu, J. Zhang, M. Hong "[Linearly Constrained Bilevel Optimization: A Smoothed Implicit Gradient Approach](#)", [link], submitted to ICLR 2023 (rating 6.75).
- [10] H. Li, S. Zhang, M. Wang, **Y. Zhang**, P. Chen, S. Liu "[Theoretical Characterization of Neural Network Generalization with Group Imbalance](#)", [link], submitted to ICLR 2023 (rating 6.6).

## RESEARCH OF INTEREST

---

### **Bilevel Optimization in Deep Learning: Theory, Algorithm, and Application**

02 2019 - Present

Bi-level optimization (BLO) is a challenging mathematical problem, while many of the deep learning tasks can be naturally formulated as a BLO and thus, the effective and efficient algorithms to solve BLO is cherished by the research community. My research in this direction are as follows:

- Summarize different BLO formulations and corresponding theories/algorithms in deep learning. Develop a ToolBox for BLO in Python (current work) .
- Design effective and efficient BLO algorithms for specific deep learning tasks, such as pruning [1] and adversarial training [4, 8].
- Provide new perspectives to interpret the current deep learning tasks and possible existing algorithms from the lens of BLO.

Related publications/submissions: [1, 4, 9]

### **Trustworthy Machine Learning: Robust, Interpretable, and Fair**

02 2019 - Present

The robustness of the deep learning models have become a research hotspot in the last decade. However, to build a trustworthy machine learning algorithm requires more than robustness. My research interest in this topic is summarized as follows:

- Design effective, efficient, and scalable robust training algorithm [3, 4, 8] to improve the adversarial robustness.
- Improve the fairness of the model [2].
- Design defense strategy against backdoor attacks [5].

Related publications/submissions: [2, 3, 4, 5, 6, 8]

## TUTORIALS/INVITED TALKS

---

- **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 in UMN**
- **10/2022:** "Revisiting and Advancing Fast Adversarial Training through the Lens of Bi-level Optimization", **INFORMS Annual Meeting (2022)**
- **04/2022:** "Adversarial Training via Bi-level Optimization", **Invited Talk in UCSB.**

## PROFESSIONAL ACTIVITIES

---

- **Reviewer:** NeurIPS'22, AISTATS'23, ICLR'23, ICASSP'23, CVPR'23, TMRL
- **TPC** for KDD'22 Workshop 4th Workshop on Adversarial Learning Methods for Machine Learning and Data Mining.
- **Student Chair** for ICML'22 Workshop AdvML: New Frontiers in Adversarial Machine Learning.
- **TPC** for NeurIPS'21 Workshop NFFL: New Frontiers in Federated Learning: Privacy, Fairness, Robustness, Personalization and Data Ownership.

## SKILLS

---

**Programming Languages** Python, C++, Java, C

**Libraries** PyTorch, OpenCV, NumPy, Matplotlib.