
Education Background

- Aug. 2021 - **Ph.D. Candidate**, Department of Computer Science, School of Computing, National University of Singapore, Singapore.
Present
 - Supervisor: Prof. Mohan Kankanhalli.
 - Research field: safety in machine learning [1, 2, 3, 4, 5].
- Sep. 2017 - **Undergraduate**, Taishan College, Shandong University, Jinan, China.
- Jun. 2021
 - Advisor: Prof. Lizhen Cui and Prof. Jingfeng Zhang.
 - Research field: adversarial machine learning [6, 7, 8].

Research Work

- [1] Xilie Xu, Keyi Kong, Ning Liu, Lizhen Cui, Di Wang, Jingfeng Zhang, and Mohan Kankanhalli. An LLM can fool itself: A prompt-based adversarial attack. *arXiv*, 2023.
- [2] Xilie Xu, Jingfeng Zhang, and Mohan Kankanhalli. Autolora: A parameter-free automated robust fine-tuning framework. *arXiv preprint arXiv:2310.01818*, 2023.
- [3] Xilie Xu*, Jingfeng Zhang*, Feng Liu, Masashi Sugiyama, and Mohan Kankanhalli. Efficient adversarial contrastive learning via robustness-aware coreset selection. In *Neural Information Processing Systems Annual Conference (NeurIPS)*, 2023.
- [4] Xilie Xu*, Jingfeng Zhang*, Feng Liu, Masashi Sugiyama, and Mohan Kankanhalli. Enhancing adversarial contrastive learning via adversarial invariant regularization. In *Neural Information Processing Systems Annual Conference (NeurIPS)*, 2023.
- [5] Xilie Xu*, Jingfeng Zhang*, Feng Liu, Masashi Sugiyama, and Mohan Kankanhalli. Adversarial attack and defense for non-parametric two-sample tests. In *International Conference on Machine Learning (ICML)*, 2022.
- [6] Jingfeng Zhang*, Xilie Xu*, Bo Han, Tongliang Liu, Gang Niu, Lizhen Cui, and Masashi Sugiyama. Noilin: Improving adversarial training and correcting stereotype of noisy labels. *Transactions on Machine Learning Research*, 2022.
- [7] Chen Chen*, Jingfeng Zhang*, Xilie Xu, Lingjuan Lyu, Chaochao Chen, Tianlei Hu, and Gang Chen. Decision boundary-aware data augmentation for adversarial training. *IEEE Transactions on Dependable and Secure Computing*, 2022.

⁰An asterisk (*) beside authors' names indicates equal contributions.

- [8] Jingfeng Zhang*, Xilie Xu*, Bo Han, Gang Niu, Lizhen Cui, Masashi Sugiyama, and Mohan Kankanhalli. Attacks which do not kill training make adversarial learning stronger. In *International Conference on Machine Learning (ICML)*, 2020.

Professional Service

- 2021-Present **Serve as a conference reviewer at NeurIPS'[21-23], ICML'[22-23], ICLR'[22-24].**
- 2022-Present **Serve as a journal reviewer at TAI, TMLR, IPL.**
- Apr. 2023 **Create the website of N-CRiPT Technical Workshop 2023 [link] and present a poster at the workshop,** National University of Singapore.
- Oct. 2022 **Serve as a member of the executive group of TrustML Young Scientist Seminars,** RIEKN-AIP, Tokyo.
- Apr. 2022 **Serve as a student reviewing member of the Master of Computing admission,** School of Computing, National University of Singapore.

Award

- Aug. 2022 **Research Achievement Award,** School of Computing, National University of Singapore.
- Jun. 2022 **ICML 2022 Participation Grant,** ICML 2022.
- Oct. 2021 **Outstanding Reviewer Award,** NeurIPS 2021.
- Jun. 2021 **Outstanding Undergraduate Thesis Award,** Shandong University.
- Sep. 2020 **Specialty Scholarship (Research Innovation Award), First Prize,** Shandong University.
- Sep. 2018 - **Outstanding Student Scholarship,** Shandong University.
Sep. 2020
- Oct. 2018 **First Prize at the 10th Mathematics Competition of Chinese College Student,** Chinese Mathematical Society.

Internship

- Jun. 2021 - **Research Intern,** Department of Ant Group-CRO Line-Security and Risk
Jul. 2021 Management, Ant Z Space, Hangzhou, China.
- Mentor: Dr. Lingjuan Lyu.
 - Research topic: adversarial machine learning and privacy.
 - Result: Proposed an innovative patent to protect model intellectual property and data privacy.

Teaching

- Aug. 2023 - **Serve as a teaching assistant for CS3244 Machine Learning,** School of
Dec. 2023 Computing, National University of Singapore.
- Jan. 2022 - **Serve as a teaching assistant for CS5242 Deep Learning and Neural
May. 2022 Networks,** School of Computing, National University of Singapore.