SHAOKUI WEI

2001 Longxiang Road, Longgang District, Shenzhen, China
Phone: +86 130 5815 8194 ◇ E-mail: shaokui.w@outlook.com ◇ Homepage: shawkui.github.io

EDUCATION BACKGROUND

Ph.D. Candidate in Data Science Sept. 2020 - Present
The Chinese University of Hong Kong, Shenzhen

Bachelor of Engineering in Electronic Information Engineering Sept. 2015 - May. 2020
The Chinese University of Hong Kong, Shenzhen
CGPA: 3.71/4.00, Major GPA: 3.86/4.00, Major Rank: 1^{st}/47

RESEARCH INTEREST

My research interests mainly lie in trustworthy AI, encompassing Security and Fairness, but also include Optimization, Kernel Methods, Reinforcement Learning, and the application of Machine Learning in economics/marketing.

PUBLICATION

- Zihao Zhu, Mingda Zhang, **Shaokui Wei**, Bingzhe Wu, Baoyuan Wu. "VDC: Versatile Data Cleanser for Detecting Dirty Samples via Visual-Linguistic Inconsistency." The Twelfth International Conference on Learning Representations. ICLR 2024.
- **Wei, Shaokui**, Mingda Zhang, Hongyuan Zha, and Baoyuan Wu. "Shared Adversarial Unlearning: Backdoor Mitigation by Unlearning Shared Adversarial Examples." 2023 Conference on Neural Information Processing Systems, NeurIPS 2023.
- Zhu, Mingli*, **Shaokui Wei***, Hongyuan Zha, and Baoyuan Wu. "Neural Polarizer: A Lightweight and Effective Backdoor Defense via Purifying Poisoned Features." 2023 Conference on Neural Information Processing Systems, NeurIPS 2023.
- Zhu, Mingli, **Shaokui Wei**, Li Shen, Yanbo Fan, and Baoyuan Wu. "Enhancing Fine-Tuning Based Backdoor Defense with Sharpness-Aware Minimization." 2023 International Conference on Computer Vision, ICCV 2023.
- **Wei, Shaokui**, Jiayin Liu, Bing Li, and Hongyuan Zha. "Mean Parity Fair Regression in RKHS." In International Conference on Artificial Intelligence and Statistics, pp. 4602-4628. PMLR, AISTATS 2023.
- Wu, Baoyuan, Hongrui Chen, Mingda Zhang, Zihao Zhu, **Shaokui Wei**, Danni Yuan, and Chao Shen. "Backdoorbench: A comprehensive benchmark of backdoor learning." Advances in Neural Information Processing Systems : 10546-10559. NeurIPS 2022.

PRE-PRINT / UNDER REVIEW

- **Wei, Shaokui**, Hongyuan Zha, and Baoyuan Wu. "Mitigating Backdoor Attack by Injecting Proactive Defensive Backdoor." arXiv preprint arXiv:2405.16112 (2024).
- Lin, Weilin, Li Liu, **Shaokui Wei**, Jianze Li, Hui Xiong. "Unveiling and Mitigating Backdoor Vulnerabilities based on Unlearning Weight Changes and Backdoor Activeness." arXiv preprint arXiv:2405.20291 (2024).
- Song, Zhengyao, Yongqiang Li, Danni Yuan, Li Liu, **Shaokui Wei**, and Baoyuan Wu. "WPDA: Frequency-based Backdoor Attack with Wavelet Packet Decomposition." arXiv preprint arXiv:2401.13578 (2024).
- Wu, Baoyuan, Hongrui Chen, Mingda Zhang, Zihao Zhu, **Shaokui Wei**, Danni Yuan, Mingli Zhu, Ruotong Wang, Li Liu, and Chao Shen. "BackdoorBench: A Comprehensive Benchmark and Analysis of Backdoor Learning." arXiv preprint arXiv:2401.15002 (2024).

- Wu, Baoyuan, **Shaokui Wei**, Mingli Zhu, Meixi Zheng, Zihao Zhu, Mingda Zhang, Hongrui Chen, Danni Yuan, Li Liu, and Qingshan Liu. "Defenses in adversarial machine learning: A survey." arXiv preprint arXiv:2312.08890 (2023).
- Zhu, Zihao, Mingda Zhang, **Shaokui Wei**, Li Shen, Yanbo Fan, and Baoyuan Wu. "Boosting backdoor attack with a learnable poisoning sample selection strategy." arXiv preprint arXiv:2307.07328 (2023).
- Yuan, Danni, **Shaokui Wei**, Mingda Zhang, Li Liu, and Baoyuan Wu. "Activation Gradient based Poisoned Sample Detection Against Backdoor Attacks." arXiv preprint arXiv:2312.06230 (2023).

PATENT

- 魏少魁, 吴保元, 张明达, 查宏远. 一种后门防御中消除共享对抗样本的方法及后门防御系统. 中国. 专利号 (Patent No.) CN117390622A. 2024.01.12.
- 朱明丽, 魏少魁, 吴保元. 通过神经偏振器净化有毒特征的后门防御方法及系统. 中国. 专利号 (Patent No.) CN116629319A. 2023.08.22.
- 吴保元, 朱明丽, 魏少魁, 沈力, 樊艳波. 后门防御方法、终端设备及计算机可读存储介质. 中国. 专利号 (Patent No.) CN116578974A. 2023.08.11.

FUNDING AND PROJECT

- Daoyuan Youth Fund Project - Class I (道远 I 类青年基金项目)

ACADEMIC ACTIVITY

- **Guest speaker** for the tutorial Backdoor Learning: Recent Advances and Future Trends in ICCV 2023.
- **Reviewer** for top-tier conferences such as ICML, NeurIPS, ICLR, CVPR, and AISTATS.

TEACHING EXPERIENCE

Teaching Assistant Sep. 2023 - Dec. 2023
The Chinese University of Hong Kong, Shenzhen
Course Title: STA 4010 Causal Inference

· This course is designed to study causal inference. Topics include discussions of observational studies, propensity score analysis, and double machine learning. Additionally, the course covers topics such as causal graphs, structural causal models, and causal discovery.

Teaching Assistant Jan. 2022 - May 2022
The Chinese University of Hong Kong, Shenzhen
Course Title: STA 3006 Design and Analysis of Experiments *Score: 5.85/6.00*

· This course is designed to study various statistical aspects of models in the analysis of variance. Topics include randomization, replication and blocking, randomized blocks, Latin squares and related designs, missing values, incomplete block designs, factorial designs, nested designs and nested-factorial designs, and 2k factorial designs. The use of statistical packages are demonstrated.

Teaching Assistant Sep. 2021 - Dec. 2021
The Chinese University of Hong Kong, Shenzhen
Course Title: STA 4030 Categorical Data Analysis *Score: 6.00/6.00*

· This course deals with major statistical techniques in analysing categorical data. Topics include measures of association, inference for two-way contingency tables, loglinear models, logit models and models for ordinal variables. The use of related statistical packages are demonstrated.

Teaching Assistant Jan. 2021 - May 2021
The Chinese University of Hong Kong, Shenzhen
Course Title: DDA 4230 Reinforcement Learning *Score: 5.86/6.00*

· This course is a basic introduction to reinforcement learning algorithms and their applications. Topics include multi-armed bandits; finite Markov decision processes; dynamic programming; Monte Carlo methods; temporal-difference learning; actor-critic methods; off-policy learning; and introduction to approximation methods.

Teaching Assistant Sep. 2020 - Dec. 2020
The Chinese University of Hong Kong, Shenzhen
Course Title: STA3050 Statistical Software *Score: 5.61/6.00*

· This course aims at providing students with basic knowledge of programming in R. A problem-solving approach is employed. Algorithm development and implementation with emphasis on examples and applications in statistics are discussed.

HONOR AND SCHOLARSHIP

Best Poster Award in The 3rd Daoyuan Academic Forum	2024
2023 Guo Tai Jun An Scholarship	2023
2023 Duan Yong Ping Travel Award	2023
AIRS Talent of Ph.D.	2020
Master's List (Top 10%)	2019, 2018
Dean's List (Top 10%)	2016-2019
National Endeavor Scholarship	2018, 2017
Undergraduate Research Award	2018, 2017, 2016
Academic Performance Scholarship	2018, 2017
Meritorious Winner in MCM (Top 10%)	2018
2^{nd} **Prize in CUMCM (Top 5%)**	2017
1^{st} **Prize in CUMCM, Guangdong Division (Top 5%)**	2017