Zhenting Wang

Department of Computer Science, Rutgers University, Piscataway, NJ 08854
https://zhentingwang.github.io; zhenting.wang@rutgers.edu
Mobile: +1-201-257-2000; Updated: October 31th, 2023

RESEARCH INTERESTS

My research interests lie at trustworthy machine learning, especially responsible GAI (generative artificial intelligence) and backdoor attacks&defenses.

EDUCATION

Rutgers University — US
PH.D. in Computer Science. — 09/2021 - Present
Advisor: Prof. Shiqing Ma and Prof. Dimitris Metaxas

University of Leeds — UK
B.S. in Computer Science (Highest Honor). — 09/2017 - 07/2021

Southwest Jiaotong University — China
B.E. in Computer Science and Technology. — 09/2017 - 07/2021

INDUSTRY EXPERIENCE

Sony AI — US
Research Scientist Intern. — 05/2023 - Present
Manager: Dr. Lingjuan Lyu

SELECTED HONORS

- ICLR 2023 Financial Assistance Award.
- NeurIPS 2022 Scholar Award.
- NeurIPS 2022 Trojan Detection Challenge (TDC 2022). Our team (PurdueRutgers) ranks the fourth on the Trojan detection track.

PREPRINTS

- **Zhenting Wang**, Chen Chen, Yuchen Liu, Lingjuan Lyu, Dimitris Metaxas, Shiqing Ma, How to Detect Unauthorized Data Usages in Text-to-image Diffusion Models.
- Guanhong Tao, **Zhenting Wang**, Siyuan Cheng, Shiqing Ma, Shengwei An, Yingqi Liu, Guangyu Shen, Zhuo Zhang, Yunshu Mao, Xiangyu Zhang, Backdoor Vulnerabilities in Normally Trained Deep Learning Models.

PUBLICATIONS

- **Zhenting Wang**, Chen Chen, Yi Zeng, Lingjuan Lyu, Shiqing Ma, Where Did I Come From? Origin Attribution of AI-Generated Images. In Proceedings of Neural Information Processing Systems 2023 **(NeurIPS 2023)**.

- Guanhong Tao*, **Zhenting Wang***, Shiwei Feng, Guangyu Shen, Shiqing Ma, Xiangyu Zhang, Distribution Preserving Backdoor Attack in Self-supervised Learning. In IEEE Symposiums on Security and Privacy 2024 **(IEEE S&P 2024, * indicates equal contribution)**.
- **Zhenting Wang**, Kai Mei, Juan Zhai, Shiqing Ma, UNICORN: A Unified Backdoor Trigger Inversion Framework. In International Conference on Learning Representations 2023 **(ICLR 2023, Spotlight)**[code].
- Kai Mei, Zheng Li, **Zhenting Wang**, Yang Zhang, Shiqing Ma, NOTABLE: Transferable Backdoor Attacks Against Prompt-based NLP Models. In Annual Meeting of the Association for Computational Linguistics 2023 **(ACL 2023)**[code].
- **Zhenting Wang**, Kai Mei, Hailun Ding, Juan Zhai, Shiqing Ma, Rethinking the Reverse-engineering of Trojan Triggers. In Proceedings of Neural Information Processing Systems 2022 **(NeurIPS 2022)**[code].
- **Zhenting Wang**, Hailun Ding, Juan Zhai, Shiqing Ma, Training with More Confidence: Mitigating Injected and Natural Backdoors During Training. In Proceedings of Neural Information Processing Systems 2022 **(NeurIPS 2022)**[code].
- **Zhenting Wang**, Juan Zhai, Shiqing Ma, BppAttack: Stealthy and Efficient Trojan Attacks against Deep Neural Networks via Image Quantization and Contrastive Adversarial Learning. In IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022 **(CVPR 2022)**[code].
- Yingqi Liu, Guangyu Shen, Guanhong Tao, **Zhenting Wang**, Shiqing Ma, Xiangyu Zhang, Complex Backdoor Detection by Symmetric Feature Differencing. In IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022 **(CVPR 2022)**[code].
- **Zhenting Wang**, Wei Li, Xiao Wu, Luhan Sheng, Learning Selective Assignment Network for Scene-aware Vehicle Detection. In IEEE International Conference on Image Processing 2022 **(ICIP 2022)**.
- Wei Li, **Zhenting Wang**, Xiao Wu, Ji Zhang, Qiang Peng, and Hongliang Li, CODAN: Counting-driven Attention Network for Vehicle Detection in Congested Scenes. In ACM International Conference on Multimedia 2020 **(MM 2020, Oral)**.

SERVICES

Program Committee/Reviewer:

- International Conference on Learning Representations **(ICLR)**, 2024.
- International Conference on Machine Learning **(ICML)**, 2022, 2023.
- Conference on Neural Information Processing Systems **(NeurIPS)**, 2022, 2023.

- IEEE/CVF Conference on Computer Vision and Pattern Recognition **(CVPR)**, 2023, 2024.
- IEEE/CVF International Conference on Computer Vision **(ICCV)**, 2023.
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining **(KDD)**, 2023.
- International Joint Conference on Artificial Intelligence **(IJCAI)**, 2023.
- SIAM International Conference on Data Mining **(SDM)**, 2024.
- Backdoor Attacks and Defenses in Machine Learning Workshop at ICLR (**BANDS**), 2023.
- Backdoors in Deep Learning - The Good, the Bad, and the Ugly at NeurIPS (**BUGS**), 2023.
- IEEE Transactions on Information Forensics & Security **(IEEE T-IFS)**

Artifact Evaluation Committee:

- USENIX Security Symposium 2022.

TALKS

- Where Did I Come From? Origin Attribution of AI-Generated Images. Peraton Labs, October 2023
- Origin Attribution for Generated Contents and Unauthorized Data Usage Detection in the AIGC Era. Sony AI, September 2023
- UNICORN: A Unified Backdoor Trigger Inversion Framework. ICLR 2023 Spotlight Presentation, May 2023
- Defending against Backdoor Attacks on Deep Neural Networks. Rutgers University, May 2023
- Backdoor Attacks&Defenses on Deep Neural Networks. Rutgers University, November 2022