Changjiang Li

• changjiang.li@stonybrook.edu • 814-996-8616 • Homepage • Google Scholar • Linkedin

Research Interest

I specialize in research that advances artificial intelligence (AI) technologies while ensuring their secure, private, and trustworthy implementation. My work tackles critical challenges in enhancing the robustness, transparency, and ethical deployment of AI systems, enabling their secure integration into real-world applications.

Education

Stony Brook University • Stony Brook, NY, US — Aug, 2023 – Present
Ph.D. candidate in Computer Science • Advisor: Dr. Ting Wang

Penn State University • State College, PA, US — Jan, 2021 – May, 2023
Ph.D. candidate in Informatics • Advisor: Dr. Ting Wang

Zhejiang University • Hangzhou, Zhejiang, China — Sep, 2017 – Mar, 2020
Master of Engineering in Cybersecurity • Advisor: Dr. Shouling Ji

Tianjin University • Tianjin, China — Sep, 2013 – Jul, 2017
Bachelor of Engineering in Optoelectronic Information Science Engineering

Selected Publications

Peer-reviewed papers:

2025 **RobustKV: Defending Large Language Models against Jailbreak Attacks via KV Eviction**
Tanqiu Jiang, Zian Wang, Jiacheng Liang, Changjiang Li, Yuhui Wang, Ting Wang
*The International Conference on Learning Representations (**ICLR**)*

2025 **RAPID: Retrieval Augmented Training of Differentially Private Diffusion Models**
Tanqiu Jiang, Changjiang Li, Fenglong Ma, Ting Wang
*The International Conference on Learning Representations (**ICLR**)*

2025 **Watch the Watcher! Backdoor Attacks on Security-Enhancing Diffusion Models**
Changjiang Li, Ren Pang, Bochuan Cao, Jinghui Chen, Fenglong Ma, Shouling Ji, Ting Wang.
USENIX Security

2025 **AIA: Autoregression-based Injection Attacks against Text2SQL Models**
Deyin Li, Xiang Ling, Changjiang Li, Xiang Chen, Chunming Wu.
*The AAAI Conference on Artificial Intelligence (**AAAI**)*

2024 **On the Difficulty of Defending Contrastive Learning against Backdoor Attacks**
Changjiang Li, Ren Pang, Bochuan Cao, Jinghui Chen, Shouling Ji, and Ting Wang.
USENIX Security

2024 **Improving the Robustness of Transformer-based Large Language Models with Dynamic Attention**
Lujia Shen, Yuwen Pu, Shouling Ji, Changjiang Li, Xuhong Zhang, Chunpeng Ge, and Ting Wang.
*The Network and Distributed System Security Symposium (**NDSS**)*

2024 **Hijack Vertical Federated Learning Models with Adversarial Embedding**
Pengyu Qiu, Xuhong Zhang, Shouling Ji, Changjiang Li, Yuwen Pu, Xing Yang, Ting Wang.
*IEEE Transactions on Dependable and Secure Computing (**TDSC**)*

2024 **When Large Language Models Confront Repository-Level Automatic Program Repair: How Well They Done?**
Yuxiao Chen, Jingzheng Wu, Xiang Ling, Changjiang Li, Zhiqing Rui, Tianyue Luo, Yanjun Wu.
*The International Conference on Software Engineering (**ICSE-Companion**)*

2024 **Model Extraction Attacks Revisited**
Jiacheng Liang, Ren Pang, Changjiang Li, Ting Wang.
*The ACM Asia Conference on Computer and Communications Security (**Asia-CCS**)*

2024 **Towards Query-Efficient Decision-Based Adversarial Attacks Through Frequency Domain**
Jianhao Fu, Xiang Ling, Yaguan Qian, Changjiang Li, Tianyue Luo, Jingzheng Wu.
*IEEE International Conference on Multimedia and Expo (**ICME**)*

2023 **An Embarrassingly Simple Backdoor Attack on Self-supervised Learning**
Changjiang Li, Ren Pang, Zhaohan Xi, Tianyu Du, Shouling Ji, Yuan Yao, Ting Wang.
*The International Conference on Computer Vision (**ICCV**)*

2023 **On the Security Risks of Knowledge Graph Reasoning**
Zhaohan Xi, Tianyu Du, Changjiang Li, Ren Pang, Shouling Ji, Xiapu Luo, Xusheng Xiao, Fenglong Ma, Ting Wang.
USENIX Security

2023 **Do Imperceptible Perturbations Really Prevent Unauthorized Data Usage in Diffusion-based Image Generation Systems?**
Bochuan Cao, Changjiang Li, Ting Wang, Jinyuan Jia, Bo Li, Jinghui Chen.
*Advances in Neural Information Processing Systems (**NeurIPS**)*

2023 **Defending Pre-trained Language Models as Few-shot Learners against Backdoor Attacks**
Zhaohan Xi, Tianyu Du, Changjiang Li, Ren Pang, Shouling Ji, Jinghui Chen, Fenglong Ma, Ting Wang.
*Advances in Neural Information Processing Systems (**NeurIPS**)*

2022 **The Dark Side of AutoML: Towards Architectural Backdoor Search**
Ren Pang, Changjiang Li, Zhaohan Xi, Shouling Ji, Ting Wang.
*The International Conference on Learning Representations (**ICLR**)*

2022 **Seeing is living? Rethinking the Security of Facial Liveness Verification in the Deepfake Era**
Changjiang Li, Li Wang, Shouling Ji, Xuhong Zhang, Zhaohan Xi, Shanqing Guo, Ting Wang.
USENIX Security

2021 **Towards Certifying the Asymmetric Robustness for Neural Networks: Quantification and Applications**
Changjiang Li, Shouling Ji, Haiqin Weng, Bo Li, Jie Shi, Raheem Beyah, Shanqing Guo, Zonghui Wang, Ting Wang.
*IEEE Transactions on Dependable and Secure Computing (**TDSC**)*

Preprints:

2024 **PRSA: PRompt Stealing Attacks against Large Language Models**
Yong Yang, Changjiang Li, Yi Jiang, Xi Chen, Haoyu Wang, Xuhong Zhang, Zonghui Wang, Shouling Ji.
arXiv preprint

2024 **Your Agent Can Defend Itself against Backdoor Attacks**
Changjiang Li, Jiacheng Liang, Bochuan Cao, Jinghui Chen, Ting Wang.
arXiv preprint

2024 **COPYRIGHTMETER: Revisiting Copyright Protection in Text-to-image Models**
Naen Xu, Changjiang Li, Tianyu Du, Minxi Li, Wenjie Luo, Jiacheng Liang, Yuyuan Li, Xuhong Zhang, Meng Han, Jianwei Yin, Ting Wang.
arXiv preprint

Professional Services

Conference Program Committee Member/Reviewer:

- The Association for the Advancement of Artificial Intelligence (AAAI), 2025.
- The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, 2025.
- The Thirteenth International Conference on Learning Representations (ICLR), 2024, 2025.
- The International Conference on Machine Learning (ICML), 2024.

- Advances in Neural Information Processing Systems (NeurIPS), 2023, 2024, 2025.
- The Information Security Conference (ISC), 2024.
- The Conference on Information and Knowledge Management (CIKM), 2023
- The Workshop on Artificial Intelligence and Security (AISec), 2023

Journal Reviewer:

- Cybersecurity
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- Transactions on Intelligent Systems and Technology (TIST)

Selected Media Coverage

2024 **Artists Are Taking Things into Their Own Hands to Protect Their Work from Generative AI**
The Associated Press

2022 **Deepfakes Can Effectively Fool Many Major Facial 'Liveness' APIs**
United.AI

2022 **Deepfakes Expose Vulnerabilities in Certain Facial Recognition Technology**
Penn State College of IST

2022 **Academic Deepfake Research Paper Suggests Liveness Detection Vulnerable**
Biometric Update

Teaching Experiences

Teaching Assistantship

2018 **Data-driven Security**, Zhejiang University, Instructor: Dr. Shouling Ji.
2021 **Special Topic: Adversarial Machine Learning**, IST 597.006 (Penn State), Instructor: Dr. Ting Wang.
2024 **Adversarial Machine Learning**, CSE 590-02 (Stony Brook), Instructor: Dr. Ting Wang.

Mentoring Experiences

Ph.D. Students

Yuxiao Chen University of Chinese Academy of Sciences, China, mentored from 04/2023 - Now
Yong Yang Zhejiang University, China, co-mentored from 01/2024 - Now
Deyin Li Zhejiang University, China, co-mentored from 03/2023 - Now

Master Students

Naen Xu Zhejiang University, China, co-mentored from 02/2024 - Now
Jianhao Fu University of Chinese Academy of Sciences, China, co-mentored from 01/2024 - Now

Undergraduate Students

Yichi Zhang Zhejiang University, China, mentored from 05/2024 - 08/2024

Awards and Honors

2019 **Graduate of Merit**, Zhejiang University
2020 **Excellent Postgraduate Students**, Zhejiang University
2019 **Kwang-Hua Scholorship**, Zhejiang University