Jiaqi Xue

+1 689-837-6702 | jiaqi.xue@ucf.edu | linkedin.com/in/jiaqi | jqxue1999.github.io

Education

University of Central Florida Orlando, FL
Ph.D. candidate in Computer Science *Jan. 2023 – Present*

Chongqing University Chongqing, CHN
B.S. in Computer Science *Sep. 2018 – Jun. 2022*

Research Area

- Adversarial Attacks and Trojan Attacks on Machine Learning [1, 4, 6, 8, 10, 12]
- Privacy-Preserving Machine Learning [2, 3, 5]
- Secure and Robust Machine Learning [1, 7, 9, 11]

Working Experience

Samsung Research America Mountain View, CA
Research Intern, supervised by Dr. Xun Chen *May. 2024 – Aug. 2024*

Working on research projects on adversarial attacks against Large Language Models (LLM) and Retrieval Augmented Generation (RAG) [12].

University of Central Florida Orlando, FL
Graduate Research Assistant, advised by Dr. Qian Lou *Jan. 2023 – Present*

Working on research projects of private machine learning [2, 3, 5], adversarial machine learning [1, 4, 6, 8, 10, 11], defense against backdoor/trojan attacks on ML [1, 7, 9] and other AI related tasks [11].

Y-tech, Kuaishou Technology Beijing, CHN
Research Intern, supervised by Dr. Shenkun Xu *Mar. 2022 – May. 2022*

Design Recommendation Algorithms for smart shooting assistant, a function for Kwai APP.

Honors and Awards

NeurIPS Top Reviewer Award	2024
NeurIPS Scholar Award	2023

Reviewer Services

- International Joint Conference on Artificial Intelligence (IJCAI)
- Neural Information Processing Systems (NeurIPS)
- International Conference on Learning Representations (ICLR)

Publications (∗ indicates equal contribution)

[12] **Jiaqi Xue**, Mengxin Zheng, Yebowen Hu, Fei Liu and Qian Lou. BadRAG: Identifying Vulnerabilities in Retrieval Augmented Generation of Large Language Models. *Under Review*

[11] Muhammad Husni Santriaji, **Jiaqi Xue**, Yancheng Zhang, Qian Lou and Yan Solihin. DataSeal: Ensuring the Verifiability of Private Computation on Encrypted Data. *The 45th IEEE Symposium on Security and Privacy, Oakland 2025*

[10] **Jiaqi Xue**, Qian Lou and Mengxin Zheng. BadFair: Backdoored Fairness Attacks with Group-conditioned Triggers. *Findings of the Empirical Methods in Natural Language Processing EMNLP 2024*

[9] **Jiaqi Xue***, Mengxin Zheng*, Zihao Wang, Xun Chen, Qian Lou, Lei Jiang and Xiaofeng Wang. SSL-Cleanse: Trojan Detection and Mitigation in Self-Supervised Learning. *The 18th European Conference on Computer Vision, ECCV 2024*

[8] Mengxin Zheng, **Jiaqi Xue**, Xun Chen, Yanshan Wang, Qian Lou and Lei Jiang. TrojFSP: Trojan Insertion in Few-shot Prompt Tuning. *2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL 2024 (Oral)*

[7] Qian Lou, **Jiaqi Xue***, Xin Liang*, Yancheng Zhang, Rui Xie and Mengxin Zheng. CR-UTP: Certified Robustness against Universal Text Perturbations on Large Language Models. *Findings of the Association for Computational Linguistics ACL 2024*

[6] **Jiaqi Xue**, Mengxin Zheng, Ting Hua, Yilin Shen, Yepeng Liu, Ladislau Boloni and Qian Lou. TrojLLM: A Black-box Trojan Prompt Attack on Large Language Models. *Thirty-seventh Conference on Neural Information Processing Systems, NeurIPS 2023*

[5] Ardhi Wiratama Baskara Yudha, **Jiaqi Xue**, Qian Lou, Huiyang Zhou and Yan Solihin. BoostCom: Towards Efficient Universal Fully Homomorphic Encryption by Boosting the Word-wise Comparisons. *Proceedings of the 2024 International Conference on Parallel Architectures and Compilation Techniques, PACT 2024*

[4] **Jiaqi Xue**, Mengxin Zheng, Yi Sheng, Lei Yang, Qian Lou and Lei Jiang. TrojFair: Trojan Fairness Attacks. *1st ACM Workshop on Large AI Systems and Models with Privacy and Safety Analysis, CCS 2024*

[3] **Jiaqi Xue**, Yancheng Zhang, Yanshan Wang, Xueqiang Wang, Hao Zheng and Qian Lou. CryptoTrain: Fast Secure Training on Encrypted Dataset. *1st ACM Workshop on Large AI Systems and Models with Privacy and Safety Analysis, CCS 2024*

[2] Yancheng Zhang, **Jiaqi Xue**, Mengxin Zheng, Mimi Xie, Mingzhe Zhang, Lei Jiang and Qian Lou. CipherPrune: Efficient and Scalable Private Transformer Inference. *Under Review*

[1] **Jiaqi Xue**, Lei Xu, Lin Chen, Weidong Shi, Kaidi Xu and Qian Lou. Audit and Improve Robustness of Private Neural Networks on Encrypted Data. *Under Review*