Qian Lou

https://www.qlou.org/ | qian.lou@ucf.edu | 812-558-6704

Research Profile

Qian Lou is a researcher pioneering the co-design of **AI models, computing infrastructure, and security** to build efficient, adaptive, and trustworthy AI systems. His work spans model innovation, dynamic multi-model execution, and secure and verifiable computing, with impact from open-source research to industrial products. He received the **NSF CAREER Award in 2026** and brings experience from both academia and Samsung AI Center.

Working Experience

University of Central Florida — Orlando, FL
Assistant Professor of Computer Science — September 2022 - Present

Samsung AI Center — Mountain View, CA
Senior Research Scientist — March 2021 - August 2022

Samsung AI Center — Mountain View, CA
Research Intern — May 2020 - August 2020

Education

Indiana University Bloomington
Ph.D., Computer Engineering — 2017 - 2021

Indiana University Bloomington
M.S., Computer Engineering — 2017 - 2019

Shandong University
B.S., Computer Science — 2013 - 2017

Grants

2026 **NSF CAREER Award — Principal Investigator; Award #2544047; $600K.**
Scalable, General-Purpose, and Automated Secure Computing for AI Infrastructure

2026 **NSF CyberAI Scholarship for Service (SFS) — Co-Principal Investigator; Award #2622846; $1.96M.**
CyberAI SFS: Educating Florida's Federal CyberAI Workforce for Secure AI in Space, Defense, and Modeling & Simulation

2025 **NSF CSR Small — Principal Investigator; Award #2413232; $600K.**
System Support for Fast, Scalable, and Verifiable Fully Homomorphic Encryption

2025 **NSF CISE Crosscutting Small: SaTC — Principal Investigator; Award #2523407; $125K UCF award.**
Collaborative Research: Privacy-Preserving Synthetic Data Generation

2025 **Department of Defense Pathfinder Program — Principal Investigator; $100K.**
LLM Security Testing

2024 **Arctic Wolf Networks, Inc. — Co-Principal Investigator; $160K.**
Industry-sponsored research on AI and cybersecurity

Selected Research Contributions and Impact

- **Efficient AI models and product translation.** At Samsung AI Center, developed **DictFormer (ICLR 2022)**, a compact Transformer architecture that was patented, demonstrated, and translated into Samsung products.
- **Adaptive AI systems.** Led **R2-Router (ICML 2026)**, which lets an AI system reason about which model and how much computation to use; it became the first method to push RouterArena beyond 70, with code and models released openly.
- **Secure and trustworthy AI. TrojLLM (NeurIPS 2023)** exposed automated prompt-injected Trojan risks in LLMs. **DataSeal (IEEE S&P 2025)** lets users verify that encrypted computation was performed correctly, while **CryptoGen** keeps LLM generation confidential.

Open-Source and Artifacts

Efficient and trustworthy machine learning: github.com/UCF-ML-Research.

Privacy-preserving computing and AI: github.com/UCF-Lou-Lab-PET.

Universal FHE scheme: github.com/UCF-Lou-Lab-PET/Universal-BGV.

Agentic FHE coding benchmark: fhe-coder.github.io.

R2-Router open-source code: github.com/UCF-ML-Research/R2-Router.

Awards

2026 NSF CAREER Award.
2025 Distinguished Speaker, Warren B. Nelms Annual AIoT Conference.
2022 Best Paper Award, Samsung AI Center.
2021 Luddy Outstanding Research Award, Indiana University Bloomington.
2020 Best Paper Finalist, ACM Parallel Architectures and Compilation Techniques (PACT).
2018 Best Paper Finalist, International Conference on Computer-Aided Design (ICCAD).

Patents

[5] Wenbo Li, Zhipeng Mo, Yi Wei, Burak Uzkent, **Qian Lou**, Yilin Shen, Hongxia Jin, Fusion techniques for combining most significant bits and least significant bits of image data in image processing or other applications (US20240080423A1).

[4] Yen-chang Hsu, Ting Hua, Feixuan Wang, **Qian Lou**, Yilin Shen, Hongxia Jin, Machine learning model compression using weighted low-rank factorization (US20230106213A1).

[3] **Qian Lou**, Yen-Chang Hsu, Burak Uzkent, Ting Hua, Yilin Shen, Hongxia Jin, Small and fast transformer model for multi-modal or other tasks (US20230177338A1).

[2] **Qian Lou**, Yilin Shen, Hongxia Jin, Ting Hua, Yen-Chang Hsu, Small and fast transformer with shared dictionary (US20230104491A1).

[1] **Qian Lou**, Yilin Shen, Hongxia Jin, Method and system for secure, accurate and fast neural network inference (US20220121947A1).

Publications

* *Students under my supervision*

[63] Mayank Kumar*, **Qian Lou**, Paulo Barreto, Martine De Cock, and Sikha Pentyala, FHAIM: Fully Homomorphic AIM for Private Synthetic Data Generation. In International Conference on Machine Learning (ICML) 2026.

[62] Nan Yan, **Qian Lou**, and Jiarong Xing, MemIncept: Steering LLM Agents via Cooperative Stealthy Memory Injections. In International Conference on Machine Learning (ICML) 2026.

[61] Jiaqi Xue*, **Qian Lou**, Jiarong Xing, and Heng Huang, R2-Router: A New Paradigm for LLM Routing with Reasoning. In International Conference on Machine Learning (ICML) 2026.

[60] Ruizhi Zhu, Yanan Guo, Huize Li, Weidong Cao, **Qian Lou**, and Xin Xin, HBM-CASO: A Coordinated Approach to HBM System-Level and On-Die ECC. In ACM Annual International Symposium on Computer Architecture (ISCA) 2026.

[59] Nokimul Hasan Arif*, **Qian Lou**, and Mengxin Zheng, Conjunctive Prompt Attacks in Multi-Agent LLM Systems. In Annual Meeting of the Association for Computational Linguistics (ACL) 2026. Main Conference Oral.

[58] Erwin Eko Wahyudi*, Yan Solihin, and **Qian Lou**, Efficient Arithmetic-and-Comparison Homomorphic Encryption with Space Switching. In IEEE Symposium on Security and Privacy (S&P) 2026.

[57] Ahasan Kabir*, Jiaqi Xue, Mengxin Zheng, and **Qian Lou**, HW-Router: Hardware-Aware Routing for Scalable Multi-LLM Serving. In Design Automation Conference (DAC) 2026.

[56] Yukuan Zhang*, Mengxin Zheng, and **Qian Lou**, SecureRoute: Input-Adaptive Encrypted Routing for Efficient and Accurate MPC-Based Transformer Inference. In Design Automation Conference (DAC) 2026.

[55] Yifei Zhao*, **Qian Lou**, and Mengxin Zheng, SIF: Semantically In-Distribution Fingerprints for Large VisionLanguage Models. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2026.

[54] Jiaqi Xue*, Xin Xin, Wei Zhang, Mengxin Zheng, Qianqian Song, Minxuan Zhou, Yushun Dong, Dongjie Wang, Xun Chen, Jiafeng Xie, Liqiang Wang, David Mohaisen, Hongyi Wu, and **Qian Lou**, SoK: Can Fully Homomorphic Encryption Support General AI Computation? A Functional and Cost Analysis. In Proceedings on Privacy Enhancing Technologies (PoPETs) 2026.

[53] Fan Li, Mayank Kumar*, Ruizhi Zhu, Mengxin Zheng, **Qian Lou**, and Xin Xin, ReliaFHE: Resilient Design for Fully Homomorphic Encryption Accelerators. In ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) 2026, pp. 1605-1621.

[52] Jiaqi Xue*, Mengxin Zheng, and **Qian Lou**, RobPI: Robust Private Inference against Malicious Client. In IEEE Conference on Secure and Trustworthy Machine Learning (SaTML) 2026.

[51] Mayank Kumar*, Jiaqi Xue, Mengxin Zheng, and **Qian Lou**, FHE-Coder: Secure Agentic Code Generation for Fully Homomorphic Encryption. In International Conference on Learning Representations (ICLR) 2026.

[50] Chi Zhang*, Mengxin Zheng, **Qian Lou**, Hui Min Leung, and Fan Chen, VQEzy: An Open-Source Dataset for Parameter Initialization in Variational Quantum Eigensolvers. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2026.

[49] Chi Zhang*, Mengxin Zheng, **Qian Lou**, and Fan Chen, DiffQ: Unified Parameter Initialization for Variational Quantum Algorithms via Diffusion Models. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2026.

[48] Cheng Chu*, **Qian Lou**, Fan Chen, Lei Jiang, QNBAD: Quantum Noise-induced Backdoor Attacks against Zero Noise Extrapolation. In Network and Distributed System Security Symposium (NDSS) 2026.

[47] Jiaqi Xue*, Mayank Kumar*, Yuzhang Shang, Shangqian Gao, Rui Ning, Mengxin Zheng, Xiaoqian Jiang, **Qian Lou**, DictPFL: Efficient and Private Federated Learning on Encrypted Gradients. In Conference on Neural Information Processing Systems (NeurIPS) 2025.

[46] Rochana Prih Hastuti*, Rian Adam Rajagede*, Mansour Al Ghanim*, Mengxin Zheng, **Qian Lou**, Factuality Beyond Coherence: Evaluating LLM Watermarking Methods for Medical Texts. In Conference on Empirical Methods in Natural Language Processing (EMNLP) 2025.

[45] Mansour Al Ghanim*, Jiaqi Xue*, Rochana Prih Hastuti*, Mengxin Zheng, Yan Solihin, **Qian Lou**, Evaluating the Robustness and Accuracy of Text Watermarking Under Real-World Cross-Lingual Manipulations. In Conference on Empirical Methods in Natural Language Processing (EMNLP) 2025.

[44] Yancheng Zhang*, Mengxin Zheng, Xun Chen, Jingtong Hu, Weidong Shi, Lei Ju, Yan Solihin, **Qian Lou**, zkVC: Fast Zero-Knowledge Proof for Private and Verifiable Computing. In Design Automation Conference (DAC) 2025.

[43] Yilan Zhu, Honghui You, Wei Zhang, Jiming Xu, **Qian Lou**, Shoumeng Yan, Lei Ju, DAHE: Parameter-adaptive and memory efficient FPGA acceleration of homomorphic encryption. In IEEE Transactions on Computers 2025.

[42] Yancheng Zhang*, Jiaqi Xue*, Mengxin Zheng, Mimi Xie, Mingzhe Zhang, Lei Jiang, **Qian Lou**, CipherPrune: Efficient and Scalable Private Transformer Inference. In International Conference on Learning Representations (ICLR) 2025.

[41] Rochana Prih Hastuti*, Rian Adam Rajagede, Mengxin Zheng, **Qian Lou**, Clinic-Prompt: Few-Shot Discrete Clinical Prompt Optimization. In Workshop on Large Language Models and Generative AI for Health at the AAAI Conference on Artificial Intelligence (AAAI) 2025.

[40] Muhammad Santriaji*, Jiaqi Xue*, **Qian Lou**, Yan Solihin, DataSeal: Ensuring the Verifiability of Private Computation on Encrypted Data. In IEEE Symposium on Security and Privacy (S&P) 2025 (Oakland). **Qian Lou** is the corresponding author.

[39] Yancheng Zhang*, Yuzhang Shang, Xun Chen, Mengxin Zheng, **Qian Lou**, Encrypted Data Pruning for Confidential Training of Deep Neural Networks. In Conference on Neural Information Processing Systems (NeurIPS) 2024.

[38] Jiaqi Xue*, **Qian Lou**, Mengxin Zheng, BadFair: Backdoored Fairness Attacks with Group-conditioned Triggers. In Conference on Empirical Methods in Natural Language Processing (EMNLP) 2024.

[37] Mansour Al Ghanim*, Saleh Almohaimeed, Mengxin Zheng, Yan Solihin, **Qian Lou**, Jailbreaking LLMs with Arabic Transliteration and Arabizi. In Conference on Empirical Methods in Natural Language Processing (EMNLP) 2024.

[36] Deng, Xianglong, Shengyu Fan, Zhicheng Hu, Zhuoyu Tian, Zihao Yang, Jiangrui Yu, Dingyuan Cao, Dan Meng, Rui Hou, Meng Li, **Qian Lou**, Mingzhe Zhang, Trinity: A General Purpose FHE Accelerator. In IEEE/ACM International Symposium on Microarchitecture (MICRO) 2024.

[35] Yancheng Zhang*, Xun Chen, **Qian Lou**, HEBridge: Connecting Arithmetic and Logic Operations in FV-style HE Schemes. In ACM Conference on Computer and Communications Security (CCS) 2024 Workshop on Encrypted Computing & Applied Homomorphic Cryptography (WAHC). (Theoretical and Foundational Work).

[34] Kunbei Cai, Zhenkai Zhang, **Qian Lou**, Fan Yao, WBP: Training-time Backdoor Attacks through Weight Bit Poisoning. In European Conference on Computer Vision (ECCV) 2024.

[33] Mengxin Zheng, Jiaqi Xue*, Zihao Wang, Xun Chen, **Qian Lou**, Lei Jiang, Xiaofeng Wang, SSL-Cleanse: Trojan detection and mitigation in self-supervised learning. In European Conference on Computer Vision (ECCV) 2024.

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

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

[30] Mengxin Zheng, Cheng Chu, **Qian Lou**, Nathan Youngblood, Mo Li, Sajjad Moazeni, Lei Jiang, OptoTFHE: An Electro-Optical Accelerator for Discretized TFHE. In IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED) 2024.

[29] Tho Nguyen, Ladda Thiamwong, **Qian Lou**, Rui Xie, Unveiling Fall Triggers in Older Adults: A Machine Learning Graphical Model Analysis. In Mathematics 2024.

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

[27] Jiaqi Xue*, Mengxin Zheng, Ting Hua, Yilin Shen, Yepeng Liu, Ladislau Boloni, **Qian Lou**, TrojLLM: A Blackbox Trojan Prompt Attack on Large Language Models. In Conference on Neural Information Processing Systems (NeurIPS) 2023.

[26] Mansour Al Ghanim*, Muhammad Santriaji*, **Qian Lou**, Yan Solihin, TrojBits: A Hardware Aware Inference-Time Attack on Transformer-Based Language Models. In European Conference on Artificial Intelligence (ECAI) 2023.

[25] Mengxin Zheng, Fan Chen, Lei Jiang, **Qian Lou**, PriML: An Electro-Optical Accelerator for Private Machine Learning on Encrypted Data. In International Symposium on Quality Electronic Design (ISQED) 2023.

[24] **Qian Lou**, Yepeng Liu*, Bo Feng, " TrojText: Invisible Test-time Trojan Insertion". In International Conference on Learning Representations (ICLR) 2023.

[23] Mengxin Zheng, **Qian Lou**, and Lei Jiang. " TrojViT: Trojan Insertion in Vision Transformers". In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2023.

[22] Mengxin Zheng, **Qian Lou**, Lei Jiang. "Primer: Privacy-Preserving Transformer on Encrypted Data". In Design Automation Conference (DAC) 2023.

[21] **Qian Lou**, Yen-Chang Hsu, Burak Uzkent, Ting Hua, Yilin Shen, and Hongxia Jin. " Lite-MDETR: A Lightweight Multi-Modal Detector ". In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2022.

[20] Lei Jiang, **Qian Lou**, Nrushad Joshi. "MATCHA: A Fast and Energy-Efficient Accelerator for Fully Homomorphic Encryption over the Torus". In Design Automation Conference (DAC) 2022.

[19] Yen-Chang Hsu, Ting Hua, Sungen Chang, **Qian Lou**, Yilin Shen, and Hongxia Jin. " Language model compression with weighted low-rank factorization ". In International Conference on Learning Representations (ICLR) 2022.

[18] **Qian Lou**, Ting Hua, Yen-Chang Hsu, Yilin Shen, and Hongxia Jin. "DictFormer: Tiny Transformer with Shared Dictionary". In International Conference on Learning Representations (ICLR) 2022.

[17] Mingqin Han, Yilian Zhu, **Qian Lou**, Zimeng Zhou, Shanqing Guo, Lei Ju. "coxHE: A software hardware co-design framework for FPGA acceleration of homomorphic computation". In Design, Automation and Test in Europe Conference (DATE) 2022.

[16] Bo Feng, **Qian Lou**, Geoffrey C. Fox, and Lei Jiang. "Low Latency Privacy-Preserving Text Analysis With GRU". In Conference on Empirical Methods in Natural Language Processing (EMNLP) 2021.

[15] **Qian Lou**, Lei Jiang "HEMET: A Homomorphic-Encryption-Friendly Privacy-Preserving Mobile Neural Network Architecture", In International Conference on Machine Learning (ICML) 2021

[14] Changsheng Zhao, Ting Hua, Yilin Shen, **Qian Lou**, and Hongxia Jin. "Automatic Mixed-Precision Quantization Search of BERT". In International Joint Conference on Artificial Intelligence (IJCAI) 2021.

[13] **Qian Lou**, Yilin Shen, Hongxia Jin, and Lei Jiang. "SAFENet: A Secure, Accurate, and Fast Neural Network Inference". In International Conference on Learning Representations (ICLR) 2021.

[12] **Qian Lou**, Wenjie Lu, Cheng Hong, and Lei Jiang. "Falcon: Fast Spectral Inference on Encrypted Data". In Conference on Neural Information Processing Systems (NeurIPS) 2020.

[11] **Qian Lou**, Bian Song, and Lei Jiang. "AutoPrivacy: Automated Layer-wise Parameter Selection for Secure Neural Network Inference". In Conference on Neural Information Processing Systems (NeurIPS) 2020.

[10] **Qian Lou**, Bo Feng, Geoffrey C. Fox, and Lei Jiang. "Glyph: Fast and Accurately Training Deep Neural Networks on Encrypted Data". In Conference on Neural Information Processing Systems (NeurIPS) 2020.

[9] **Qian Lou**, Sarath Janga, and Lei Jiang. "Helix: Algorithm/Architecture Co-design for Accelerating Nanopore Genome Base-calling." In ACM International Conference on Parallel Architectures and Compilation Techniques (PACT) 2020. Best Paper Nomination (4/197).

[8] **Qian Lou**, Feng Guo, Minje Kim, Lantao Liu, and Lei Jiang. "AutoQ: Automated Kernel-Wise Neural Network Quantization." In International Conference on Learning Representations (ICLR) 2020.

[7] Farzinah Zokaee, **Qian Lou**, N. Youngblood, Weichen Liu, and Lei Jiang. "LightBulb: A Photonic-NonvolatileMemory-based Accelerator for Binarized Convolutional Neural Networks." In Design, Automation and Test in Europe Conference (DATE) 2020.

[6] **Qian Lou**, Weichen Liu, Wenyang Liu, Feng Guo, and Lei Jiang, "MindReading: An Ultra-Low-Power Photonic Accelerator for EEG-based Human Intention Recognition," In Asia and South Pacific Design Automation Conference (ASP-DAC) 2020.

[5] **Qian Lou** and Lei Jiang. "SHE: A Fast and Accurate Deep Neural Network for Encrypted Data." In Conference on Neural Information Processing Systems (NeurIPS) 2019.

[4] Weichen Liu, Wenyang Liu, Yichen Ye, **Qian Lou**, Yiyuan Xie, and Lei Jiang. "HolyLight: A Nanophotonic Accelerator for Deep Learning in Data Centers." In Design, Automation and Test in Europe Conference (DATE) 2019.

[3] **Qian Lou**, Wujie Wen, and Lei Jiang. "3DICT: a reliable and QoS capable mobile process-in-memory architecture for lookup-based CNNs in 3D XPoint ReRAMs". In IEEE/ACM International Conference on Computer-Aided Design (ICCAD) 2018.

[2] **Qian Lou**, and Lei Jiang. "BRAWL: A Spintronics-Based Portable Basecalling-in-Memory Architecture for Nanopore Genome Sequencing". IEEE Computer Architecture Letters, 2018.

[1] **Qian Lou**, Mengying Zhao, Lei Ju, Chun Xue, Jingtong Hu, and Zhiping Jia. "Runtime and reconfiguration dual-aware placement for SRAM-NVM hybrid FPGAs." IEEE Non-Volatile Memory Systems and Applications Symposium (NVMSA) 2017.

Mentoring

Mentoring 12 Ph.D. students, including one recent graduate who accepted a tenure-track faculty position. Student research published at IEEE S&P, NeurIPS, ICLR, ICCV, DAC, ACL, EMNLP, and MICRO.

Teaching

University of Central Florida: CDA 5106 Advanced Computer Architecture (Fall 2025, Fall 2023, Fall 2022); CAP 6614 Current Topics in Machine Learning (Spring 2025, Spring 2024, Spring 2023).
Indiana University Bloomington: IFT 4030/7030 Machine Learning for Signal Processing (Spring 2020); ENGR-E599 Introduction to Intelligent Systems (Fall 2019).

Invited Talks

2025 Distinguished Speaker, Warren B. Nelms Annual AIoT Conference at University of Florida.
2025 NSF CSR PI Meetings.
2024 Distinguished Speaker, Florida State University.

Professional Service

NSF Panelist (2023-2025).
Area Chair: ACL, NeurIPS, ICLR.
Program Committee: CCS, NDSS, ISCA, DAC, ICCAD.
Track Chair: Hardware Security, GLSVLSI 2025-2026.
Senior/Program Committee: AAAI, CVPR, ACL, EMNLP, ICML, TMLR, PETs, TIFS.
Additional service: Session Chair, PACT 2024; Local Area Chair, ISCA 2023; UCF CS Graduate Committee Member; ESWEEK 2023 Design Contest Organizer.