Yuqi Jia

Email: yuqi.jia@duke.edu
Phone: (+1)9842784308
Google Scholar: link
Citizenship: China

Education

Duke University — Durham, US
Ph.D. student in ECE — Aug. 2022 – Present
Advisor: **Prof. Neil Gong**

Shanghai Jiao Tong University — Shanghai, China
Undergraduate in Computer Science — Sept. 2018 – June. 2022
Member of **ACM Honors Class**, which is an elite CS program for top 5% talented students.

Research Interests

AI Security, Trustworthy AI,
Large Language Model, Prompt Injection Attacks

Honors and Scholarships

Scholarship
Zhiyuan Honorary Scholarship (**Top 5%** in SJTU) 2018-2021
Excellent School-level Scholarship (**Top 3%** in SJTU) 2018-2021

Competition
Honorable Mention, Interdisciplinary Contest In Modeling(ICM) 2020
First Prize, Mathematics competition of Chinese College Students(CMC) 2019
Silver Medal (rank 1st), The 33rd Chinese Mathematical Olympiad(CMO) 2017

Research Experience

Research Intern, ByteDance
Security Flow Team — May. 2025 – Sep. 2025
Research topic: Redteaming against Multi-agent Systems

Research Intern, ByteDance
Security Flow Team — June. 2024 – August. 2024
Research topic: Prompt Injection Attacks

Research Assistant, Duke University
advised by **Prof. Neil Gong** — Aug. 2022 – Present
Research topic: Large Language Model, Federated Learning

Undergraduate researcher, Mila-Quebec AI Institute
advised by **Prof. Jian Tang** — May. 2021 – Nov. 2021
Research topic: Graph Structure Learning

Undergraduate Researcher, ThinkLab, Shanghai Jiao Tong University

advised by **Prof. Junchi Yan** Aug. 2020 – July 2021
Research topic: Adversarial Attack

Publications

Competitive Advantage Attacks to Decentralized Federated Learning. PDF
Y. Jia, M. Fang and N. Gong
Conference on Neural Information Processing Systems (NeurIPS) 2025.

Tracing Back the Malicious Clients in Poisoning Attacks to Federated Learning. PDF
Y. Jia, M. Fang, H. Liu, J. Zhang and N. Gong
Conference on Neural Information Processing Systems (NeurIPS) 2025.

PromptLocate: Localizing Prompt Injection Attacks.
Y Jia, Y Liu, Z Shao, J Jia, and NZ Gong
IEEE Symposium on Security and Privacy 2026.

DataSentinel: A Game-Theoretic Detection of Prompt Injection Attacks. PDF
Y Liu, **Y Jia**, J Jia, D Song and NZ Gong
IEEE Symposium on Security and Privacy 2025. **Distinguished Paper Award**

Evaluating Large Language Model based Personal Information Extraction and Countermeasures. PDF
Y Liu, **Y Jia**, J Jia and NZ Gong
USENIX Security Symposium 2025.

Formalizing and benchmarking prompt injection attacks and defenses. PDF
Y. Liu, **Y. Jia**, R. Geng, J. Jia, N. Gong
USENIX Security Symposium 2024.

Unlocking the Potential of Federated Learning: The Symphony of Dataset Distillation via Deep Generative Latents. PDF
Y. Jia*, S. Vahidian*, J. Sun, J. Zhang, V. Kungurtsev, N. Gong and Y. Chen
European Conference on Computer Vision (ECCV) 2024.

Academic Activities

Conference Reviewer
USENIX Security 2026; ACM CCS 2025; ICLR 2026,2024

Teaching Experience

Teaching Assistant, Duke University Spring 2025
Generative AI: Foundations, Applications, and Safety, ECE590

Teaching Assistant, Duke University Fall 2024
Natural Language Processing, ECE684

Teaching Assistant, Shanghai Jiao Tong University Fall 2020
Mathematical Logic, CS301-1