Muyang Zheng

sandyyyzheng33@gmail.com | sandyyyzheng.github.io | Google Scholar

Education

Hefei University of Technology — Hefei, China
B.E. in Information Security — *Sept 2021 - June 2025*

University of California, Davis — Davis, CA, USA
M.S. in Computer Science — *Sept 2025 - Present*

Research Experience

Research Intern, Binjiang Institute of Zhejiang University – Hangzhou, Zhejiang — Aug 2024 – Apr 2025

- Developed core algorithms for a Large Language Model Safety Evaluation Platform, integrating **15 jailbreak attack algorithms** (e.g., GCG, AutoDAN, GPTFuzzer) and **2 evaluation algorithms**
- Conducted in-depth study and reproduction of open-source LLM safety frameworks like OpenCompass from Shanghai AI Lab and PyRIT from Microsoft
- **Conducted performance testing** of in-house LLMs with BitNet and LLMPerf; analyzed throughput and latency, and produced comprehensive reports

Publication

MIST: Jailbreaking Black-box Large Language Models via Iterative Semantic Tuning (Under Review) — May 2025

Muyang Zheng, Yuanzhi Yao*, Changting Lin, Rui Wang, Meng Han

Preprint: `arxiv.org/abs/2506.16792`

Projects

Jailbreak System — `github.com/SandyyyZheng/JailbreakSystem` — May 2025

- Awarded **University-Level Outstanding Graduation Project**
- Developed a **comprehensive platform** for testing, evaluating, visualizing, and analyzing jailbreak attacks against closed-source large language models
- Implemented our proposed jailbreak method *MIST* within the platform, demonstrating a **high attack success rate** in empirical evaluations

Decentralized Threshold Signatures with Dynamically Private Accountability — Aug 2024

- Awarded **Second Prize** in the 17^{th} National College Student Information Security Contest
- Implemented frontend for Web and Android; designed interfaces with **Figma** and Element Plus
- Led the design and writing of an **84-page technical report in LaTeX**, including methodology, experimental design, and results visualization

Mouse Maze — `github.com/SandyyyZheng/Mouse-Maze` — June 2022

- **Graded A** in the curriculum design for data structure
- Designed and implemented an interactive maze game with a polished UI, where players guide a mouse to find cheese under time constraints, using **Depth-First Search** (DFS) for pathfinding

Industry Experience

Data Analysis Intern, Wuhan Cloud Computing Technology Co., Ltd. — June 2024 - Aug 2024

- Participated in the development of Wuhan urban digital public infrastructure platform and data governance
- Applied Python and SQL to process, analyze and integrate data, formed **120+** categories of effective quality inspection rules and completed **1,000+** integration governance tasks

Software Development Intern, Wuhan Municipal Healthcare Security Administration — June 2023 - Aug 2023

- Extracted and anonymized (with SHA-256 hashing and character masking) personal data of **10,000+** insured residents in Wuhan
- Implemented Wuhan health insurance policy rules in Python to calculate annual premiums for **10,000+** residents; generated detailed result tables via SQL for reporting and analysis

Extracurricular Activities

Member, Technology Innovation Center of Student Union, HFUT — Oct 2021 - May 2023

- Designed 10+ posters and wrote press releases and thank-you notes for the university's various competitions and activities
- Organized and planned the 21^{st} Radio Orienteering Contest, responsible for equipment debugging and radio transmitter deployment

Skills

Coding: Python, Java, SQL, JavaScript, C++, C

Frameworks & Tools: Flask, Vue.js, Spring Boot, Git, LaTeX, Visio, Figma

Honors

Third-class Scholarship & Merit Student in 2021-2022 Academic Year

Outstanding Member of the Student Union in 2021-2022 Academic Year