Shuangning Yang

github.com/VodkaVortex | vodkavortex.github.io | realyangshuangning@gmail.com | +86 133-0866-2608

Research Interests

I am interested in both securing AI systems and using AI to improve software security. Specifically, my research focuses on improving the reliability and robustness of LLM-based systems, while exploring how LLMs can support security tasks, such as vulnerability detection.

Education

Anhui University, M.Eng. in Artificial Intelligence — Sep. 2024 – May. 2027

- Average Scores: 92/100 Track: LLM Security & AI for Vulnerability Detection
- Selected Achievements: NDSS 26, ICASSP 26, 1 under submission

Hefei University of Technology, B.S. in Logistics Management — Sep. 2020 – May. 2024

- Average Scores: 88/100 Track: Data-Driven Modeling for Logistics Systems
- Selected Coursework: C, Java, Computer Networks, Database Systems

Publications

- **S. Yang**, H. Yang, G. Zhao, M. Zhou, J. Tai, and X. Jia, "What is the Risk? Evaluating the Impact of Knowledge Distillation on LLM Vulnerabilities," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2026.
- H. Yang, J. Guo, **S. Yang**, G. Zhao, Q. Liu, C. Zhang, Z. Tan et al., "IoTBec: An Accurate and Efficient Recurring Vulnerability Detection Framework for Black Box IoT Devices," in *Proc. Network and Distributed System Security. (NDSS)*, 2026.
- **S. Yang**, H. Yang, J. Peng et al., "DetFuzz: A Semantic-Guided Fuzzing Framework via Operation Logic Inference for Black-box IoT Devices," *Under Submission*.
- J. Tai, **S. Yang**, J. Wang, Y. Li, Q. Liu, and X. Jia, "Survey of Adversarial Attacks and Defenses for Large Language Models," *Journal of Computer Research and Development*, 2025.
- Q. Liu, H. Yang, **S. Yang**, G. Zhao, J. Peng, J. Guo, and W. Zhang, "FirmSV: Detecting Stored Vulnerabilities in IoT Firmware using Static Taint Analysis," in *Proc. Conf. Computer Supported Cooperative Work in Design (CSCWD)*, 2026.

Research Experience

DetFuzz: LLM-Driven Semantic Fuzzing for Black-box IoT Vulnerability Detection Oct.2025 – Feb.2026

Lead Researcher

- Designed an LLM-driven fuzzing pipeline that combines CoT-guided operation logic inference with front-end static analysis to automatically identify high-risk parameters and satisfy operation prerequisites, enabling valid test case generation against black-box IoT devices without firmware or source code.
- Detected **137 vulnerabilities** (122 CVE IDs assigned; **40 previously unknown**, avg. CVSS **8.54**); outperformed SOTA black-box and grey-box fuzzers by **7**×, achieving 100% precision and 94.48% recall.

What is the Risk: Security Risks of Knowledge Distillation in LLMs Apr.2025 – Sep.2025

Lead Researcher

- Designed and implemented an evaluation framework across five distillation configurations (SeqKD, PPO/DPO-based RL/RO and their combinations) with dual-judge assessment (human + GPT-4), demonstrating that joint knowledge learning increases jailbreak ASR by up to 18.38% and induces alignment failure in safety-robust models.
- Published as first author at **ICASSP 2026**.

IoTReaper: An End-to-End LLM Agent for IoT Vulnerability Detection Mar.2026 – Apr.2026

Independent Developer

- Designed and built an open-source multi-agent system that automates the full 1-day discovery pipeline—from CVE patch analysis to PoC generation and emulator-based verification—targeting commodity IoT devices (Tenda, TP-Link, TOTOLink).
- **Outcome:** Discovered 34 1-day vulnerabilities across 19 firmware versions.

Skills

- **Programming:** Agent Engineering, Python, C
- **Security Tools:** IDA, Binwalk, QEMU
- **Vulnerabilities**: Submitting **20+ independent CVEs**
- **Language:** Mandarin (Native), English (Conversational)
- **Machine Learning:** PyTorch, HuggingFace Transformers (fine-tuning, RL-based training)
- **Certifications:** CISP-PTE (Certified Information Security Professional – Penetration Testing Engineer), China Information Technology Security Evaluation Center