

Mingjie Shen

PhD Student, Electrical and Computer Engineering

✉ shen497@purdue.edu 📍 West Lafayette, IN, USA 🔗 szsam.github.io 🌐 szsam 📄 mingjie-shen-92554818a

🎓 Google Scholar

Summary

- PhD student in ECE at Purdue working on software security, static program analysis, and AI for software engineering. Built Aithos, a tool-augmented LLM-agent framework that filters 78% of SAST false positives while keeping 97.5% of true positives (RAID 2026); applied CodeQL to 258 open-source embedded projects, finding 709 defects (ISSTA 2025); currently building LLM agents that synthesize CodeQL queries and program-transformation rules for EDK2 UEFI firmware. Over 100 of my patches have been [merged upstream](#) into projects including Apache NuttX, Contiki-NG, Mbed OS, RIOT, SDL, zstd, and util-linux. Seeking full-time software security or systems engineering roles in industry.

Education

Aug 2019 – present **Purdue University**, PhD in Electrical and Computer Engineering – West Lafayette, IN, USA

- Advisor: Prof. Aravind Machiry
- GPA: Plan-of-Study 3.87, Overall 3.58

Sept 2015 – June 2019 **Nanjing University**, BSc in Computer Science and Technology – Nanjing, China

- GPA: 4.58/5.0 (top 4%), Major GPA: 4.66/5.0

Experience

Jan 2023 – present **Research Assistant**, Purdue University – West Lafayette, IN

Static analysis and LLM-agent research under Prof. Aravind Machiry; responsible disclosure and upstream patching of defects in open-source software.

June 2024 – Dec 2024 **Visiting Student Intern (full-time)**, Dartmouth College – Hanover, NH

Worked with Prof. Christophe Hauser on Aithos, an LLM-agent SAST triage framework (RAID 2026).

Projects

Jan 2025 – present **LLM Agents for Static Analysis of EDK2 UEFI Firmware** – Purdue

- Building LLM agents that generate EDK2-specific CodeQL queries for UEFI vulnerability discovery.
- Learning symbolic program-transformation rules to automate collateral evolution across EDK2.

June 2024 – Oct 2026 **Aithos: LLM-Agent Triage of Static Analysis Alerts** – Purdue & Dartmouth

- Designed an LLM agentic framework that classifies SAST alerts from natural-language guidelines, dynamically retrieving program context through LLVM-based analysis primitives.
- Added a feedback loop that turns developer corrections into reusable, repository-specific examples.
- Preserved 97.5% of true positives while filtering 78.0% of false positives across four SAST tools on real-world C/C++ projects, outperforming a state-of-the-art baseline.
- Uncovered 90 unknown bugs in 28 popular Debian packages; fixes merged in zstd, util-linux, pciutils.

Mar 2023 – Dec 2024 **Detecting 709 Defects in Open-Source Embedded Software with CodeQL** – Purdue

- Ran CodeQL across 258 GitHub projects; found 709 true defects at a 34% false positive rate, 535 (75%) likely security vulnerabilities (e.g., null pointer dereference, buffer overflow).
- Got 376 (53%) defects confirmed by maintainers, mainly by submitting and merging [pull requests](#) to projects including those by Microsoft, Amazon, and the Apache Foundation; 2 CVEs issued.
- Opened pull requests adding CodeQL scans to CI pipelines; 37 (71% of active repositories) merged.

- July 2022 – **Detecting Layering Violations in Embedded Systems** – Purdue
Mar 2023
- Developed a static analysis tool to detect MMIO accesses bypassing the HAL.
 - Analyzed 988 applications across 5 RTOSes; identified 369 violations.

Publications

- 2026 **Democratizing False Positives Filtering Through Learning Assisted Reasoning**
Mingjie Shen, Sai Ritvik Tanksalkar, Christophe Hauser, Aravind Machiry
29th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2026), to appear
- July 2025 **Finding 709 Defects in 258 Projects: An Experience Report on Applying CodeQL to Open-Source Embedded Software (Experience Paper)**
Mingjie Shen, Akul Abhilash Pillai, Brian A. Yuan, James C. Davis, Aravind Machiry
Proc. ACM Softw. Eng. 2, ISSTA 2025, 1077–1100
doi.org/10.1145/3728923
- Dec 2024 **Insights from Running 24 Static Analysis Tools on Open Source Software Repositories**
Fabiha Hashmat, Zeyad Alwaleed Aljaali, **Mingjie Shen**, Aravind Machiry
International Conference on Information Systems Security (ICISS 2024), LNCS 15416, 225–245
doi.org/10.1007/978-3-031-80020-7_13
- June 2023 **Towards Automated Identification of Layering Violations in Embedded Applications (WIP)**
Mingjie Shen, James C. Davis, Aravind Machiry
Languages, Compilers, and Tools for Embedded Systems (LCTES 2023), 143–147
doi.org/10.1145/3589610.3596271

Skills

Languages: C, C++, Python, Bash

Frameworks: CodeQL, LLVM, LangChain, GitHub Actions

Areas: Static Analysis, Software Security, LLM Agents, Embedded Systems