Muxi Lyu

muxilyulucy.github.io | muxi_lyu@berkeley.edu

Research Interests

My research lies at the intersection of **AI Security** and **AI for Security**, including:

- **Secure Web Agents**: Designing web-browsing agents that are resilient against attacks such as prompt injection, malicious content manipulation, and unsafe action execution.
- **Secure Code Generation**: Leveraging large language models to synthesize code that not only meets functional requirements but also avoids introducing security vulnerabilities.
- **AI for Software Testing**: Applying AI techniques to enhance fuzzing, vulnerability detection, and automated test generation for large-scale software systems.

Education

University of California, Berkeley 08/2025 – Present
Ph.D. in Computer Science Advisors: David Wagner, Koushik Sen

Johns Hopkins University 08/2021 – 12/2024
B.S. in Computer Science & Applied Mathematics and Statistics; M.S.E. in Computer Science
GPA: 3.91/4.00 Advisor: Yinzhi Cao

Publications

[1] Zifeng Kang, **Muxi Lyu**, Zhengyu Liu, Jianjia Yu, Runqi Fan, Song Li, and Yinzhi Cao. *Follow My Flow: Unveiling Client-Side Prototype Pollution Gadgets from One Million Real-World Websites*. In the Proceedings of the IEEE Symposium on Security and Privacy (Oakland), 2025. [Link] **Distinguished Paper Award**.

[2] Zhengyu Liu, Jiacheng Zhong, Jianjia Yu, **Muxi Lyu**, Zifeng Kang, and Yinzhi Cao. *The First Large-Scale Systematic Study of Python Class Pollution Vulnerability*. To appear in the Proceedings of the IEEE Symposium on Security and Privacy (Oakland), 2026.

Project Experience

Client-side JavaScript Dynamic Analysis Across 1M Websites Sep 2023 – May 2024

– Modified Chrome's V8 JavaScript engine to track data flow and enable runtime value injections.
– Designed a high-throughput database to manage concurrent data processing for large-scale web crawling.
– Discovered **133 zero-day vulnerabilities**, including critical ones in **Meta software** and **Vue**, leading to CVE assignments and bug bounty rewards.
– Identified **23 websites** where prototype pollution vulnerabilities, previously deemed inconsequential, caused real-world security impacts through the discovered gadgets.

Agentic Java Fuzzing Pipeline with Structured Input Generator May 2024 – Jul 2024

– Designed and developed an automated Java fuzzing pipeline that iteratively generated and refined structural input generators, improving testing efficiency by increasing input validity rates.
– Automated dependency rebuilding and code generation using a feedback-driven ReAct agent.
– Collaborated within a ∼**50-person** multi-lab team to integrate the pipeline into the final DARPA AIxCC competition submission, contributing to recognition as one of the **top seven winners** and receiving a **$2M award**.

Taxonomy and Prevalence Study of Python Class Pollution April 2025 – June 2025

– Defined the first taxonomy of Python class pollution, introducing **5 new vulnerability types**.

– Characterized pollution **primitives, targets, and consequences**, providing a foundation for triage, detection, and exploit development.

– Conducted a large-scale measurement study on over **50K popular PyPI packages**, revealing the prevalence of diverse "get" and "set" primitives across Python's ecosystem.

– Identified critical security consequences including **remote code execution (RCE)**, **authentication bypass**, and **denial-of-service (DoS)**, expanding the recognized threat landscape.

Honors

Distinguished Paper Award, IEEE Symposium on Security and Privacy (Oakland)	2025
Honorable Mention in the NSF GRFP competition	2025
Michael J. Muuss Research Award ($3000 Award)	2024
Finalist – DARPA AI Cyber Challenge (AIxCC), with Team 42-b3yond-6ug ($2,000,000 Award)	2024

Skills

Programming	Python, Java, C/C++, JavaScript
Security & Analysis	Fuzzing, Dynamic & Static Analysis, Vulnerability Detection
Systems & Tools	Agentic Systems, Chrome V8 Engine Modification