Jiahao Zhang

Nashville, TN, USA

Homepage | Google Scholar | Email

Education

Vanderbilt University — **TN, USA**
Ph.D. in Computer Science — Aug. 2024 – Present
Advisor: Prof. Yu Huang

Southern University of Science and Technology — **Guangdong, China**
B.E. in Computer Science and Technology — Sept. 2020 – June 2024
Highest Honors in Computer Science and Engineering
Advisor: Prof. Yepang Liu

University of Notre Dame — **IN, USA**
Exchange Student — Aug. 2023 – Dec. 2023
Advisor: Prof. Joanna C.S. Santos

University of California, Irvine — **CA, USA**
UCInspire Spring Research Program — Mar. 2023 – July 2023
Advisor: Prof. Marco Levorato

Publications(*equal contribution)

[C1] Re(gEx|DoS)Eval: Evaluating Generated Regular Expressions and their Proneness to DoS Attacks
ML Siddiq, **Jiahao Zhang**, Lindsay Roney, and Joanna C.S. Santos [URL]
Proceedings of the 46th International Conference on Software Engineering
ICSE *(NIER Track)*, 2024

[C2] Understanding ReDoS: Insights from LLM-Generated Regexes and Developer Forums
ML Siddiq*, **Jiahao Zhang***, and Joanna C.S. Santos [URL]
Proceedings of the 32nd International Conference on Program Comprehension
ICPC *(Research Track)*, 2024

[C3] Quality Assessment of ChatGPT Generated Code and their Use by Developers
ML Siddiq, Lindsay Roney, **Jiahao Zhang**, and Joanna C.S. Santos [URL]
Proceedings of the 21st International Conference on Mining Software Repositories
MSR *(Mining Challenge Track)*, 2024

Research Experiences

University of Notre Dame — Aug. 2023 – Dec. 2023
Supervised by Prof. **Joanna C.S. Santos**
Research Topics: *Software Security, Mining Software Repositories, AI4SE, LLM*

- Designed Re(gEx|DoS)Eval, a framework for systematically evaluating the correctness and security against ReDoS attacks of regexes generated by large language models. Included a high-quality dataset, refined prompts, test cases, and novel metrics for functional correctness and vulnerability assessment [C1].
- Addressed critical gaps in understanding the capabilities of LLMs in generating non-vulnerable and functionally correct regular expressions. Empirically evaluated three LLMs for correct and secure regex generation, compared characteristics of LLM-generated ReDoS vulnerabilities with real-world cases, and analyzed the developer community's discussion on ReDoS issues with data collected from GitHub and StackOverflow [C2].
- Analyzed the quality and usage of ChatGPT-generated code in software development, examining its integration into software repositories and its broader applications in learning new frameworks and development kits [C3].

Southern University of Science and Technology Aug. 2022 – July. 2023
Supervised by Prof. **Yepang Liu**
Research Topics: *Program Repair, Software Evolution, Mining Software Repositories, AI4SE, LLM*

– Proposed a novel "history-driven" approach for Just-In-Time (JIT) method renaming, leveraging the extensive history of code changes combined with the advanced capabilities of large language models (LLMs). This bolstered software reliability and understandability by providing more accurate method names, informed by the most similar code change in history.

University of California, Irvine Mar. 2023 – July 2023
Supervised by Prof. **Marco Levorato**
Research Topics: *Computational Slicing, Resource Allocation, Traffic Shaping*

– Introduced a computational slicing framework that leveraged traffic shaping to indirectly control application processing time by managing byte flow into the CPU. This offered an abstraction over direct CPU cycle control, beneficial for testing delay-sensitive applications across diverse hardware and firmware setups.

Teaching Experience

Teaching Assistant of CS201: Discrete Mathematics Fall 2022, Spring 2023

Skills

Programming Languages	Skilled: Java, Python, LaTeX Familiar: C/C++, SQL, HTML/CSS, JavaScript/TypeScript, Scheme, Verilog
Frameworks	PyTorch, Langchain, React, Next.js, Spring Boot, Django, CherryPy, Node.js
Language Proficiency	Mandarin - native, English - fluent (IELTS: 7.5, CET-6: 634)