YUE LIU

Ph.D. Candidate ◇ Department of Information Technology ◇ Monash University, Australia
yue.liu1@monash.edu ◇ +61 0479114068 ◇ Github ◇ Google Scholar ◇ Homepage ◇ Linkedin

RESEARCH INTERESTS

My research focuses on **reliable and secure Large Language Models (LLMs) for software development**. I explore how LLMs can enhance developer productivity and software quality by generating, analyzing, and transforming code within Integrated Development Environments (IDEs). I also investigate the challenges and risks of LLMs, such as code quality, security, and explainability, and propose solutions using techniques such as program analysis, adversarial learning, and interpretability methods. My research bridges the gap between natural language processing and software engineering, and aims to create intelligent development tools that are efficient, secure, and reliable.

EDUCATION

Ph.D. in Software Engineering, Monash University 2024
Advised by Chakkrit Tantithamthavorn an Li Li
Thesis Topic: Towards Reliable LLM-based Software Development Tools

Bachelor of Computer Science, Southern University of Science and Technology 2019
Advised by Adam Ghandar and Georgios Theodoropoulos
Thesis Topic: Machine Learning Techniques for Credit Scoring Applications

APPOINTMENTS

Visiting Ph.D. Student, Singapore Management University 2023
Advised by David Lo
Research Assistant, Southern University of Science and Technology 2020
Advised by Yepang Liu

PUBLICATIONS

Published

- Yue Liu, Chakkrit Tantithamthavorn, Yonghui Liu, and Li Li. **On the Reliability and Explainability of Language Models for Program Generation**. In *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 2024, to appear. [pdf] [code] ***(Core A*, CCF A)***
- Yue Liu, Thanh Le-Cong, Ratnadira Widyasari, Chakkrit Tantithamthavorn, Li Li, Xuan-Bach D. Le, and David Lo. **Refining ChatGPT-Generated Code: Characterizing and Mitigating Code Quality Issues**. In *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 2024, to appear. [pdf] [code] ***(Core A*, CCF A)***
- Yue Liu, Chakkrit Tantithamthavorn, Li Li, and Yepang Liu. **Explainable AI for Android Malware Detection: Towards Understanding Why the Models Perform So Well?**. In the *33rd International Symposium on Software Reliability Engineering (ISSRE)*, IEEE, 2022: 169-180. [pdf] [code] ***(Core A, CCF B)***
- Yue Liu, Chakkrit Tantithamthavorn, Li Li, and Yepang Liu. **Deep Learning for Android Malware Defenses: a Systematic Literature Review**. In *ACM Computing Surveys (CSUR)*, 2022, 55(8): 1-36. [pdf] [code] ***(Core A*, SCI-Q1)***
- Yue Liu, Adam Ghandar, and Georgios Theodoropoulos. **Online NEAT for Credit Evaluation–a Dynamic Problem with Sequential Data**. In the *2nd KDD Workshop on Anomaly Detection in Finance*, August 2019. [pdf]
- Yue Liu, Adam Ghandar, and Georgios Theodoropoulos. **Island model genetic algorithm for feature selection in non-traditional credit risk evaluation**. In the *IEEE congress on evolutionary computation (CEC)*, pages 2771-2778, June 2019. [pdf] ***(Core B)***

- Yue Liu, Adam Ghandar, and Georgios Theodoropoulos. **A metaheuristic strategy for feature selection problems: Application to credit risk evaluation in emerging markets**. In the *2019 IEEE Conference on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, pages 1-7, May 2019. [pdf]
- Haonan Hu, Yue Liu, Yanjie Zhao, Yonghui Liu, Xiaoyu Sun, Chakkrit Tantithamthavorn, and Li Li. **Detecting Temporal Inconsistency in Biased Datasets for Android Malware Detection**. In the *2023 38th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, IEEE, 2023: 17-23. [pdf]

Under Review

- Yue Liu, Chakkrit Tantithamthavorn, Yonghui Liu, Patanamon Thongtanunam, and Li Li. **Automatically Recommend Code Updates: Are We There Yet?**. 2024. [pdf]
- Xinyi Hou, Yanjie Zhao, Yue Liu, Zhou Yang, Kailong Wang, Li Li, Xiapu Luo, David Lo, John Grundy, and Haoyu Wang. **Large Language Models for Software Engineering: A Systematic Literature Review**. 2024. [pdf]
- Xinyu She, Yue Liu, Yanjie Zhao, Yiling He, Li Li, Chakkrit Tantithamthavorn, Zhan Qin, and Haoyu Wang. **Pitfalls in Language Models for Code Intelligence: A Taxonomy and Survey**. 2024. [pdf] [code]

PROJECTS

Towards Reliable Language Models for Code Intelligence

- Built a taxonomy to summarize the existing reliability issues of large language models (LLMs) for code generation, analysis, and transformation tasks. [project]
- Evaluated the current benchmarks for LLMs and found severe data quality issues that inflated the performance of LLMs. [pdf] [code]
- Identified code smells and bugs in code generated by state-of-the-art AI models like OpenAI's ChatGPT. Proposed strategies for automated prompt design to improve code generation. [pdf] [code]

Towards Robust and Secure AI-based IDEs

- Analyzed the security risks of VSCode, the most popular IDE, and its extension system, and found several attack vectors and vulnerabilities that could compromise the user's data and code.
- Proposed an automated static and dynamic program analysis tool for evaluating the security of VSCode extensions.
- Discovered new attack vulnerabilities in popular AI-driven software like Copilot and Tabnine, which could lead to data and API key theft.

TECHNICAL SKILLS

Programming Languages:

- Python: Proficient in Python programming for data analysis, machine learning, natural language processing, web scraping, and software engineering research
- Java: Experienced in Java development for teaching undergraduate computer science courses
- JavaScript: Knowledgeable in JavaScript for developing VSCode extensions and IDE plugins

Large Language Models (LLMs):

- Familiar with using LLMs such as GPT, T5, and BERT for natural language processing tasks.
- Expertise in evaluating the reliability, security, and explainability of LLMs for code generation, analysis, and transformation
- Skilled in prompt engineering and fine-tuning LLMs for software engineering tasks

- Skilled in feature engineering, model selection, and hyperparameter tuning for machine learning
- Skilled in using AI frameworks and tools such as TensorFlow, PyTorch, and Scikit-learn.

Software Engineering:

- Strong background in software testing, debugging, and code review
- Knowledgeable in software security analysis, including static and dynamic program analysis for detecting vulnerabilities
- Experienced in mining and analyzing large-scale software repositories and datasets

Research Skills:

- Strong ability to conduct literature reviews, gap analysis, and problem formulation
- Experienced in designing and conducting empirical studies, user studies, and experiments
- Skilled in statistical analysis and data visualization for research insights
- Proficient in academic writing and presentation for communicating research findings

PROFESSIONAL SERVICE ACTIVITIES

Reviewer:

- IEEE Transactions on Software Engineering (TSE, Core A*)
- The International World Wide Web Conference (WWW 2024, Core A*)
- The 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023, Core A)
- The 20th International Conference on Mining Software Repositories (MSR 2023, Core A)

Sub-Reviewer:

- International Conference on Software Engineering (ICSE)
- ACM Computing Surveys (CSUR)
- IEEE Transactions on Dependable and Secure Computing
- IEEE Transactions on Information Forensics and Security

AWARDS AND HONORS

- Monash Ph.D. Scholarship (Full Time), 2019-2024.
- Excellent Undergraduate Thesis, Southern University of Science and Technology(SUSTech), 2019
- Chinese National Encouragement scholarship, 2017-2019

ADVISING

- Xinyu She (Master Student at Huazhong University of Science and Technology), Improving the reliability of large language models for code
- Haonan Hu (Bachelor Student at Southern University of Science and Technology), The impacts of API evolution on ML-based Android malware detection

TEACHING EXPERIENCE

- Teaching Assistant, FIT3077 Software Engineering: Architecture and Design, Monash University (Spring 2022)
- Teaching Assistant, FIT2099 Object oriented Design and Implementation, Monash University (Spring 2022)

- Teaching Assistant, FIT1051 Programming fundamentals in Java, Monash University (Fall 2022)
- Teaching Assistant, CS203 Data Structure and Algorithm Analysis, SUSTech (Fall 2018)
- Teaching Assistant, CS209 Computer System Design and Application, SUSTech (Spring 2018, Fall 2017)
- Teaching Assistant, CS201 Computer Organization Principle, SUSTech (Spring 2018)