

# Jieke Shi

CS Ph.D. Candidate | Research Engineer  
Singapore Management University

**Address:** College of Integrative Studies, B2,  
RISE Center, 10 Canning Rise, Singapore 179873

**E-Mail:** [jiekeshi@smu.edu.sg](mailto:jiekeshi@smu.edu.sg)

**Website:** <https://jiekeshi.github.io>

## RESEARCH INTERESTS

Software Engineering, Large Language Models of Code, Program Synthesis, AI Testing

I design and implement **automated synthesis**-assisted methods for validating the **safety** of AI-enabled software and enhancing the **efficiency** of large language models for code.

## EDUCATION

Aug 2022 - Present **Singapore Management University**, Ph.D. Candidate in Computer Science  
Supervisor: **ACM/IEEE/ASE Fellow, OUB Chair Professor David Lo**

Aug 2016 - Jun 2020 **Yangzhou University**, B.Eng. in Electronic Engineering

## EMPLOYMENT

Jun 2021 - Present **Singapore Management University**, Research Engineer  
Working on the research and development of

Jul 2020 - Apr 2021 **Chinese Academy of Sciences, Institute of Computing**, Research Assistant Intern  
Working on the research and development of Data2Text, a system that generates text from data using a neural network model.

## PUBLICATIONS

Note: In software engineering, ICSE and ASE are recognized as top-tier conferences, and TOSEM and TSE are top-tier journals (CORE A\*); MSR, ICSME, and SANER are prominent conferences covering software mining, analysis, and evolution (CORE A). In security, S&P is a top-tier conference (CORE A\*), while ACSAC is a leading venue with a focus on security applications (CORE A).

[C]: Full Paper in Conference

[J]: Full Paper in Journal

[S]: Short Paper

- [J.6] Zhou Yang, **Jieke Shi** (✉ Corr.), Premkumar Devanbu, David Lo. (2025) “Ecosystem of Large Language Models for Code”. **TOSEM** - ACM Transactions on Software Engineering and Methodology. [PDF]
- [J.2] **Jieke Shi**, Zhou Yang, Junda He, Bowen Xu, Dongsun Kim, DongGyun Han, David Lo. “Finding Safety Violations of AI-Enabled Control Systems through the Lens of Synthesized Proxy Programs.” *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 33 pages, CORE A\*. [PDF] [Code]
- [J.1] **Jieke Shi**, Zhou Yang, David Lo. “Efficient and Green Large Language Models for Software Engineering: Literature Review, Vision, and the Road Ahead.” *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 23 pages, CORE A\*. [PDF]
- [C.8] Zhou Yang, **Jieke Shi**, Muhammad Hilmi Asyrofi, Bowen Xu, Xin Zhou, DongGyun Han, David Lo. “Prioritizing Speech Test Cases.” *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 28 pages, CORE A\*. [PDF] [Code]
- [J.0] Zhou Yang, Zhipeng Zhao, Chenyu Wang, **Jieke Shi**, Dongsun Kim, DongGyun Han, David Lo. “Gotcha! This Model Uses My Code! Evaluating Membership Leakage Risks in Code Models.” *IEEE Transactions on Software Engineering (TSE)*, 16 pages, CORE A\*. [PDF]
- [C.7] **Jieke Shi**, Zhou Yang, Hong Jin Kang, Bowen Xu, Junda He, David Lo. “Greening Large Language Models of Code.” *2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE 2024, SEIS Track)*, 12 pages, CORE A\*. [PDF] [Code]

- [C.6] Junda He, Zhou Yang, **Jieke Shi**, Chengran Yang, Kisub Kim, Bowen Xu, Xin Zhou, David Lo. “Curiosity-Driven Testing for Sequential Decision-Making Process.” *2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE 2024, Research Track, 13 pages)*, CORE A\*. [[PDF](#)] [[Code](#)]
- [C.5] Zhou Yang, Zhipeng Zhao, Chenyu Wang, **Jieke Shi**, Dongsun Kim, DongGyun Han, David Lo. “Unveiling Memorization in Code Models.” *2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE 2024, Research Track, 13 pages)*, CORE A\*. [[PDF](#)] [[Code](#)]
- [C.4] Zhou Yang, Chenyu Wang, **Jieke Shi**, Thong Hoang, Pavneet Kochhar, Qinghua Lu, Zhenchang Xing, David Lo. “What Do Users Ask in Open-Source AI Repositories? An Empirical Study of GitHub Issues.” *2023 IEEE/ACM 20th International Conference on Mining Software Repositories (MSR 2023, Technical Track, 12 pages)*, CORE A. [[PDF](#)] [[Code](#)]
- [C.3] Chen Gong, Zhou Yang, Yunpeng Bai, **Jieke Shi**, Arunesh Sinha, Bowen Xu, David Lo, Xinwen Hou, Guoliang Fan. “Curiosity-Driven and Victim-Aware Adversarial Policies.” *2022 39th Annual Computer Security Applications Conference (ACSAC 2022, Technical Paper, Honorable Mention Award, 15 pages)*, CORE A. [[PDF](#)] [[Code](#)]
- [C.2] **Jieke Shi**, Zhou Yang, Bowen Xu, Hong Jin Kang, David Lo. “Compressing Pre-trained Models of Code into 3 MB.” *2022 IEEE/ACM 37th International Conference on Automated Software Engineering (ASE 2022, Research Paper, Nominated for ACM SIGSOFT Distinguished Paper Award, 12 pages)*, CORE A\*. [[PDF](#)] [[Poster](#)] [[Code](#)]
- [C.1] **Jieke Shi**, Zhou Yang, Junda He, Bowen Xu, David Lo. “Can Identifier Splitting Improve Open-Vocabulary Language Model of Code?” *2022 IEEE 29th International Conference on Software Analysis, Evolution and Reengineering (SANER 2022, ERA Track, 5 pages)*, CORE A. [[PDF](#)] [[Poster](#)] [[Code](#)] [[Video](#)]