

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

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**, School of Computing and Information Systems
Ph.D. Candidate in Computer Science

- **22 Conference/Journal Publications**
- **OpenAI Researcher Access Program Grant** (2025)
- **SMU SCIS Dean's List Award** (2023)
- **SMU Presidential Doctoral Fellowship** (2023)
- **ACM SIGSOFT Distinguished Paper Award Nomination** (ASE 2022)
- **Honorable Mention Award** (ACSAC 2022)

Supervisor: ACM/IEEE/ASE Fellow, OUB Chair Professor [David Lo](#)

Aug 2016 – Jun 2020 **Yangzhou University**, School of Information Engineering
B.Eng. in Electronic Engineering

- **Outstanding Undergraduate Thesis Award** (Top 0.2% in province)
- **Outstanding Graduate** (Top 5%)

EMPLOYMENT

Jun 2021 – Present **Singapore Management University**, Research Engineer
Working as lead research staff under the AI.SG Programme on the testing of reinforcement learning-enabled software. Led to multiple publications, including [J.5, J.6, C.5, C.8, C.9].

Jul 2020 – Apr 2021 **Chinese Academy of Sciences, Institute of Computing**, Research Assistant Intern
Developed a prototype system for generating faithful natural language text from structured data.

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 Conference Paper (Total: 11) [J]: Full Journal Paper (Total: 7) [S]: Short Paper (Total: 4)

[As Lead Author \(1st / 🏆 Co-1st / 📧 Corresponding\)](#)

- [J.7] Zhou Yang, **Jieke Shi** 📧, Premkumar Devanbu, David Lo. “Ecosystem of Large Language Models for Code”. ACM Transactions on Software Engineering and Methodology (**TOSEM 2025**). [[PDF](#)]
- [J.6] **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 2025**). [[PDF](#)] [[Code](#)]

- [J.5] **Jieke Shi**, Junda He, Zhou Yang, Đorđe Žikelić, David Lo. “*Synthesizing Efficient and Permissive Programmatic Runtime Shields for Neural Policies*”. ACM Transactions on Software Engineering and Methodology (**TOSEM 2025**, Under Revision). [[PDF](#)]
- [J.4] **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 2024**). [[PDF](#)]
- [C.11] **Jieke Shi**, Zhou Yang, Hong Jin Kang, Bowen Xu, Junda He, and David Lo. “*Greening Large Language Models of Code*”. In Proceedings of the 46th IEEE/ACM International Conference on Software Engineering (**ICSE 2024**). [[PDF](#)] [[Code](#)]
- [C.10] **Jieke Shi**, Zhou Yang, Bowen Xu, Hong Jin Kang, and David Lo. “*Compressing Pre-trained Models of Code into 3 MB*”. In Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering (**ASE 2022**, 🏆 **ACM SIGSOFT Distinguished Paper Award Nomination**). [[PDF](#)] [[Poster](#)] [[Code](#)]
- [S.4] **Jieke Shi**, Zhou Yang, Junda He, Bowen Xu, and David Lo. “*Can Identifier Splitting Improve Open-Vocabulary Language Model of Code?*”. In Proceedings of the 29th IEEE International Conference on Software Analysis, Evolution and Reengineering (**SANER 2022**, ERA Track). [[PDF](#)] [[Poster](#)] [[Code](#)]
- [S.3] Zhou Yang, **Jieke Shi** 📄, Shaowei Wang, and David Lo. “*IncBL: Incremental Bug Localization*”. In Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering (**ASE 2021**, Tool Demonstrations). [[PDF](#)] [[Poster](#)] [[Code](#)]

As Co-Author

- [J.3] 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 2024**). [[PDF](#)] [[Code](#)]
- [J.2] 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 2024**). [[PDF](#)]
- [C.9] Chen Gong, Zhou Yang, Yunpeng Bai, Junda He, **Jieke Shi**, Kecen Li, Arunesh Sinha, Bowen Xu, Xinwen Hou, David Lo, and Tianhao Wang. “*BAFFLE: Hiding Backdoors in Offline Reinforcement Learning Datasets*”. In Proceedings of the 45th IEEE Symposium on Security and Privacy (**S&P 2024**). [[PDF](#)] [[Code](#)]
- [C.8] Junda He, Zhou Yang, **Jieke Shi**, Chengran Yang, Kisub Kim, Bowen Xu, Xin Zhou, and David Lo. “*Curiosity-Driven Testing for Sequential Decision-Making Process*”. In Proceedings of the 46th IEEE/ACM International Conference on Software Engineering (**ICSE 2024**). [[PDF](#)] [[Code](#)]
- [C.7] Zhou Yang, Zhipeng Zhao, Chenyu Wang, **Jieke Shi**, Dongsun Kim, DongGyun Han, and David Lo. “*Unveiling Memorization in Code Models*”. In Proceedings of the 46th IEEE/ACM International Conference on Software Engineering (**ICSE 2024**). [[PDF](#)] [[Code](#)]
- [J.1] Zhou Yang, Bowen Xu, Jie M. Zhang, Hong Jin Kang, **Jieke Shi**, Junda He, and David Lo. “*Stealthy Backdoor Attack for Code Models*”. IEEE Transactions on Software Engineering (**TSE 2024**). [[PDF](#)] [[Code](#)]
- [C.6] 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*”. In Proceedings of the 20th IEEE/ACM International Conference on Mining Software Repositories (**MSR 2023**). [[PDF](#)] [[Code](#)]
- [C.5] 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*”. In Proceedings of the 39th Annual Computer Security Applications Conference (**ACSAC 2022**, 🏆 **Honorable Mention Award**). [[PDF](#)] [[Code](#)]

- [C.4] Chengran Yang, Bowen Xu, Ferdian Thung, Yucen Shi, Ting Zhang, Zhou Yang, Xin Zhou, **Jieke Shi**, Junda He, DongGyun Han, and David Lo. “*Answer Summarization for Technical Queries: Benchmark and New Approach*”. In Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering (**ASE 2022**). [[PDF](#)] [[Code](#)]
- [C.3] Zhou Yang, **Jieke Shi**, Muhammad Hilmi Asyrofi, and David Lo. “*Revisiting Neuron Coverage Metrics and Quality of Deep Neural Networks*”. In Proceedings of the 29th IEEE International Conference on Software Analysis, Evolution and Reengineering (**SANER 2022**). [[PDF](#)] [[Code](#)]
- [C.2] Ratnadira Widayarsi, Stefanus Agus Haryono, Ferdian Thung, **Jieke Shi**, Constance Tan, Fiona Wee, Jack Phan, and David Lo. “*On the Influence of Biases in Bug Localization: Evaluation and Benchmark*”. In Proceedings of the 29th IEEE International Conference on Software Analysis, Evolution and Reengineering (**SANER 2022**). [[PDF](#)] [[Code](#)] [[Dataset](#)]
- [C.1] Zhou Yang, **Jieke Shi**, Junda He, and David Lo. “*Natural Attack for Pre-trained Models of Code*”. In Proceedings of the 44th IEEE/ACM International Conference on Software Engineering (**ICSE 2022**). [[PDF](#)] [[Code](#)]
- [S.2] Zhou Yang, Harshit Jain, **Jieke Shi**, Muhammad Hilmi Asyrofi, and David Lo. “*BiasHeal: On-the-Fly Black-Box Healing of Bias in Sentiment Analysis Systems*”. In Proceedings of the 37th IEEE International Conference on Software Maintenance and Evolution (**ICSME 2021**, NIER Track). [[PDF](#)] [[Code](#)]
- [S.1] Muhammad Hilmi Asyrofi, Zhou Yang, **Jieke Shi**, Chu Wei Quan, and David Lo. “*Can Differential Testing Improve Automatic Speech Recognition Systems?*”. In Proceedings of the 37th IEEE International Conference on Software Maintenance and Evolution (**ICSME 2021**, NIER Track). [[PDF](#)] [[Code](#)]

HONORS & GRANTS

- 2025 **OpenAI Researcher Access Program Grant** (\$5,000)
- 2023 **Presidential Doctoral Fellowship**, Singapore Management University
SCIS Dean’s List Award, Singapore Management University
- 2022 **Honorable Mention Award**, ACSAC 2022
ACM SIGSOFT Distinguished Paper Award Nomination, ASE 2022
- 2020 **Outstanding Undergraduate Thesis** (Top 0.2% in province), Jiangsu Province, China
Outstanding Graduate (Top 5%), Yangzhou University
- 2019 & Before **National Encouragement Scholarship**, Ministry of Education of China
2nd Class Presidential Scholarship, Yangzhou University
1st Class Presidential Scholarship, Yangzhou University

ACADEMIC SERVICE

Program Committee / Reviewer (Reviewed 43 Submissions)

- 2024–2025 3× Full Papers, Shadow Program Committee Member, ICSE
- 2021–2025 3× Full Papers, External Reviewer, ICSE
- 2023–2025 4× Full Papers, External Reviewer, FSE
- 2022–2024 6× Full Papers, External Reviewer, ASE
- 2025 4× Full Papers, Program Committee Member, COMPSAC
- 2024–2025 5× Full/Short Papers, Junior Program Committee Member, MSR
- 2024–2025 2× Short Papers, Program Committee Member, GREENS@ICSE
- 2024 3× Full Papers, Reviewer, ACM Transactions on Software Engineering and Methodology
- 2022–2023 3× Full Papers, External Reviewer, IJCAI

2023 1× Full Paper, External Reviewer, ICSME
2023 1× Full Paper, External Reviewer, LLM4Code@ICSE
2022 2× Full Papers, External Reviewer, SANER
2022 2× Full Paper, External Reviewer, CAIN
2022 1× Full Paper, External Reviewer, SE4RAI@ICSE
2022 1× Full Paper, External Reviewer, ICECCS
2022 1× Full Paper, Robotics and Autonomous Systems Journal
2021 1× Full Paper, Reviewer, Neurocomputing Journal (Software Section)

Organizing Committee

2024 Web Chair, SEA4DQ 2024 Workshop at FSE 2024

INVITED TALKS

Nov 2023 **Singapore Management University**, Cluster Seminar, Invited by David Lo
Towards Compact and Efficient Language Models of Code via Compression
Oct 2023 **North Carolina State University**, Guest Lecture, Invited by Bowen Xu
Enhancing Efficiency of Large Language Models of Code via Model Compression [[Link](#)]

REFEREES

David Lo, ACM/IEEE/ASE Fellow, OUB Chair Professor
School of Computing and Information Systems, Singapore Management University
davidlo@smu.edu.sg