Linghao Zhang

✉ starrywhuzhang@gmail.com | ⌂ github.com/StarryZhang-whu | 🔗 szhang.me

Research Interests

The intersection of Software Engineering and Machine Learning, focusing on ① **ML for SE:** using ML/AI to address various challenges in SE and ② **SE for ML:** leveraging best practices from SE to guide software systems with ML components.

Education

Wuhan University — Wuhan, China
Bachelor in Software Engineering, School of Computer Science — Sep. 2021 – Expect Jun. 2025
- GPA: **3.71/4**, TOEFL: **104/120**

Publications

Preprint:

[1] **BEYOND CODE: Evaluating Large Language Models on Dependency Inference**
Linghao Zhang, Junhao Wang, Shilin He, Chaoyun Zhang, Bowen Li, Yu Kang, and other 4 authors.
arXiv, 2025. *[code] [dataset]* *(to be public)*

[2] **An Empirical Study of Bugs in MLOps Frameworks**
Linghao Zhang*, Vishesh Prasad*, Shreya Rao*, Saikat Dutta.
ISSTA, 2025. (in submission).

[3] **RAG-Enhanced Commit Message Generation**
Linghao Zhang, Hongyi Zhang, Chong Wang, Peng Liang.
TOSEM, 2025. (in submission). [arXiv]

Conference Proceeding:

[4] **Using Large Language Models for Commit Message Generation: A Preliminary Study**
Linghao Zhang, Jingshu Zhao, Chong Wang, Peng Liang.
SANER (ERA Track), 2024. [paper]

Research Experience

Microsoft Research
Research Intern at **DKI Group, MSRA Shanghai**, advised by Shilin He and Yu Kang — Aug. 2024 – Present
- **Research Topics**: AI for Software Engineering, Repo-level Tasks, CI/CD.
- Built a repository-level automated software engineering testing infrastructure. Curated a large-scale, multi-language benchmark comprising 800 real-world repositories through a CI-based, fully automated pipeline [1].
- Exploring pathways toward repository intelligence, addressing SE challenges in repo/project scope.

Cornell University
Research Intern at **SEAL**, advised by Prof. Saikat Dutta — May. 2024 – Present
- **Research Topics**: MLOps, SE4ML, Software Testing, OSS.
- Focus on researching foundational software systems for machine learning in production, including open-source MLOps frameworks. Mining bugs from GitHub repositories' PRs and issue tracker, identifying patterns in root causes and fixes for these bugs across different MLOps frameworks. To uncover technical debt in developing and operating ML models [2].

Wuhan University
Research Assistant at **Software Architecture Group**, advised by Prof. Peng Liang — Jun. 2022 – May. 2024
- **Research Topics**: Code Language Models, LLMs, IR, and their applications in SE.
- Leverage information retrieval and language models with code capabilities to handle natural language elements in software development process. Fine-tune pre-trained models with augmented inputs on downstream tasks; Play LLMs with structured output and efficient serving techniques, covered 7 models for comprehensive experiments [3][4].

Presentations

Using Large Language Models for Commit Message Generation Rovaniemi, Finland
SANER. Mar, 2024. [slides]

Awards

2025   **Travel Grant for ICSE'25**, SMeW
2024   **Luojia Excellence Fellowship**, Wuhan University
2021-2023   **Outstanding Student Scholarship**, Wuhan University

Skills

Programming: <u>Python</u> *(preferred)*, Rust *(favored)*, C/C++, C#, Java, JavaScript, Linux, SQL, Swift, LaTeX
Framework: ⏻ PyTorch, 🦜🔗 LangChain, DSPy, Git, vLLM, 🦙 Ollama, 🤗 Transformers, 🐳 Docker, Sysbox, K8s

Trying to contribute to open-source every now and then.

Last update: 2024-12