Linghao Zhang

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

Research Interests

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

Education

Wuhan University   Wuhan, China
Bachelor in Software Engineering, School of Computer Science   Sep. 2021 – Expect Jun. 2025
- GPA: **3.71/4**, weighted avg. score: **89**.

Publications

Preprint:

[C3] **BigBuild: Can LLMs Identify Dependencies of Your Repo?**
Linghao Zhang*, Junhao Wang*, Shilin He, Yu Kang.
ACL, 2025. (in submission)

[C2] **An Empirical Study of Bugs in MLOps Frameworks**
Linghao Zhang*, Vishesh Prasad*, Shreya Rao*, Saikat Dutta.
ISSTA, 2025. (in submission)

[J1] **RAG-Enhanced Commit Message Generation**
Linghao Zhang, Hongyi Zhang, Chong Wang, Peng Liang.
TOSEM, 2025. (in submission). [arXiv]

Conference Proceeding:

[C1] **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 Yu Kang   Aug. 2024 – Present
- **Research Topics**: AI for Software Engineering, Repo-level Tasks, CI/CD.
- Ongoing

UIUC+ Summer Research Program
Summer Intern at **SEAL**, advised by Saikat Dutta *(Cornell)*   May. 2024 – Oct. 2024
- **Research Topics**: MLOps, SE4ML, Software Testing, OSS.
- Focused on researching foundational software infrastructure for machine learning in production, including open-source MLOps frameworks. Mining bugs from GitHub repositories' PRs and issues, and identifying common patterns in root causes and fixes across different MLOps from a high-level perspective.

Wuhan University
Research Intern at **Software Architecture Group**, advised by Peng Liang and Chong Wang   Jun. 2022 – May. 2024
- **Research Topics**: Code Language Models (CodeT5, CodeLlama, StarCoder, etc.), LLMs, and their applications in SE.
- During the research, I focused on using language models with code capabilities to generate natural language text (commit message) in software engineering. I fully led the research, completing the entire process from literature review, idea come up, experimental work, to paper writing. Eventually, I published two papers [1][2] as the first author.

Wuhan University

Research Intern at **Sigma Lab**, advised by Juhua Liu and Bo Du   Mar. 2023 – Dec. 2023
- **Research Topics**: Natural Language Processing, BERT, Transformer, the life-cycle of training a LLM.

Presentations

[1] **Using Large Language Models for Commit Message Generation**   Rovaniemi, Finland
SANER, 2024, Main Conference. [slides]

Skills

Programming: Python *(preferred)*, C/C++, C#, Rust *(favored)*, Java, JavaScript, Linux, SQL, Swift

Framework: ⭕ PyTorch, 🔝 TensorFlow, 🦜🔗 LangChain, Git, vLLM, 🦙 Ollama, 🤗 Transformers, 🐳 Docker, K8s

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

Last update: 2024-10