

# Sen Fang

☎ (+46) 076-6359636 • ✉ senf@kth.se

## Research Interests

---

I have a keen research interest in automated software engineering and deep learning. More specifically, I have dedicated my efforts to constructing deep learning models to automate various software engineering tasks such as code search, code comment generation, and bug report-related software maintenance activities. Currently, my research encompasses three key areas: 1) How different code representations affect the performance of LLMs on the task of program repair. 2) Application of LLMs in the realm of code optimization. 3) Investigate data leakage in large language models (LLMs) and explore its potential impact.

## Education

---

### Central China Normal University

MSc in Electronics and Communication Engineering, GPA:3.44/5.0

Thesis: Research on Machine Translation Model Based on Self-Attention Mechanism

2018-2020

Advisor: Shaocheng Qu

### Wuhan Polytechnic University

BSc in Electronic Information Engineering, GPA: 3.47/5.0

Thesis: Performance Comparison of Direct Sequence Spread Spectrum Systems under Different Modulation Methods

Outstanding Student 2018.

2014-2018

## Research Experience

---

### Completed Research.....

#### Supersonic: Learning to Generate Source Code Optimisations in C/C++

<https://github.com/ASSERT-KTH/Supersonic>

We propose Supersonic, an LLM-based approach for source code-level code optimization, which outperforms GPT-3.5 and GPT-4 in our comparative experiments.

Mar. 2023–Oct. 2023

GitHub

#### RepresentThemAll: A Universal Learning Representation of Bug Reports

<https://github.com/ICSE-2023/RepresentThemALL>

We propose RTA, a pre-trained model that can learn the universal representation of bug reports and can be used for different software maintenance activities.

Mar. 2021–Dec. 2022

GitHub

#### Java Code Clone Detection

Jun. 2021–May 2022

We propose a new approach to identify functional clone code by analyzing the source code at the level of the intermediate code.

#### Automated Pull Request Summarization

<https://github.com/TomasAndersonFang/PRHAN>

We propose PRHAN, a new hybrid attention network based approach that can automatically generate descriptions for pull requests.

Nov. 2020–Jul. 2021

GitHub

#### Bug Priority Prediction

<https://github.com/TanYoushuai123/PPWGNCN>

We propose PPWGNCN, a graph convolutional network-based approach that can automatically perform bug priority inference.

Sep. 2020–May 2021

GitHub

#### Neural Code Search

<https://github.com/TomasAndersonFang/SANCS>

We propose SANCS, a novel neural code search model based on a self-attention network.

Jun. 2020–Feb. 2021

GitHub

### Ongoing Research.....

#### Deep Smart Contract Intent Detection

<https://gitlab.com/web3se/smartintent>

○ Smart contracts are often accompanied by huge economic effects, so their security is a very important issue. We try to define a new task: automated smart contract intent detection. This task can automatically detect risky intent in a smart contract.

○ We design an automated model SMARTINTENTNN, which is a highlight approach based on K-means and Bi-LSTM.

○ We collected the first dataset used for smart contract intent detection.

Jul. 2022–Now

GitLab

#### RepairLLaMA: Fine-Tuning with Efficient Parameters and Representations for Program Repair

Keep Private

Mar. 2023–Now

GitHub

In this work, we explore the impact of various code representations on program repair. Then, we use our best-designed representation to build an LLM-based program repair model, RepairLLaMA.

### Generative AI to Generate Test Data Generators

Mar. 2023–Now

*Keep Private*

*GitHub*

In this work, we explore to what extent an LLM, i.e., ChatGPT, can generate test data generators that can keep consistent with the different cultural backgrounds.

### Data Leakage in LLMs

Mar. 2023–Now

*Keep Private*

*GitHub*

We mainly explore whether the data for automated program repair is leaked in the training phase of LLMs and evaluate the potential impact.

## Publications

---

Published.....

- **Fang, Sen**, Tao Zhang, Youshuai Tan, He Jiang, Xin Xia, and Xiaobing Sun. "RepresentThemAll: A Universal Learning Representation of Bug Reports." In Proceedings of the 45th International Conference on Software Engineering. 2023.
- Yuan, Dawei\*, **Sen Fang\***, Tao Zhang, Zhou Xu, and Xiapu Luo. "Java Code Clone Detection by Exploiting Semantic and Syntax Information From Intermediate Code-Based Graph." IEEE Transactions on Reliability (2022). (\*: Equal contribution)
- **Fang, Sen**, Tao Zhang, You-Shuai Tan, Zhou Xu, Zhi-Xin Yuan, and Ling-Ze Meng. "PRHAN: Automated Pull Request Description Generation Based on Hybrid Attention Network." Journal of Systems and Software 185 (2022): 111160.
- **Fang, Sen\***, You-shuai Tan\*, Tao Zhang, Zhou Xu, and Hui Liu. "Effective prediction of bug-fixing priority via weighted graph convolutional networks." IEEE Transactions on Reliability 70, no. 2 (2021): 563-574.
- **Fang, Sen**, You-Shuai Tan, Tao Zhang, and Yepang Liu. "Self-attention networks for code search." Information and Software Technology 134 (2021): 106542.
- Tan, Youshuai; Chen, JinFu; Shang, WeiYi; Zhang, Tao; **Fang, Sen**; Luo, Xiapu; Chen, Zijie; Qi, Shuhao. "STRE: An Automated Approach to Suggesting App Developers When to Stop Reading Reviews." IEEE Transactions on Software Engineering.
- Li, Yao; Zhang, Tao; Luo, Xiapu; Cai, Haipeng; **Fang, Sen**; Yuan, Dawei. "Do Pre-trained Language Models Indeed Understand Software Engineering Tasks?" IEEE Transactions on Software Engineering.

Submitting.....

- Zimin Chen; **Fang, Sen**; Martin Monperrus. "Supersonic: Learning to Generate Source Code Optimisations in C/C++" IEEE Transactions on Software Engineering. (Under Review)
- Yao Li; **Fang, Sen**, etc. "Enhancing Android Malware Detection: The Influence of ChatGPT on Decision-centric Task" (Under Review)

## Technical skills

---

Programming Languages	Python, C/C++, $\LaTeX$
Libraries	PyTorch, transformers, TensorFlow, JAX, NumPy, Scikit-Learn
Operating Systems	Unix/Linux, Windows
Development Environments	Linux Toolchain, Jupyter, PyCharm, Visual Studio Code