

# QI LUO

 [Github@rocky-lq](#)  [12232440@mail.sustech.edu.cn](mailto:12232440@mail.sustech.edu.cn)  [Google Scholar](#)  +86-188-9859-1764

## EDUCATION

### Southern University of Science and Technology

Shenzhen, China

M.Eng. in Computer Science and Technology, GPA: 3.62 / 4.0

Sep.2022 - Jun.2025 (exp.)

Courses: Advanced Algorithms, Advanced Artificial Intelligence, Numerical Computing, Reinforcement Learning, etc.

### Xidian University

Xi'an, China

B.Eng. in Computer Science and Technology, GPA: 3.6 / 4.0

Sep.2015 - Jun.2019

Courses: Operating Systems, Computer Networks, Advanced Mathematics, Linear Algebra, Data Structures and Algorithms, etc.

## RESEARCH PAPERS

- [TOSEM'2025, CCF-A] **Prompt-based Code Completion via Multi-Retrieval Augmented Generation.**  
Hanzhuo Tan\*, Qi Luo\*, Ling Jiang, Zizheng Zhan, Jing Li, Haotian Zhang, Yuqun Zhang [Paper]
- [EMNLP'2024, CCF-B] **LLM4Decompile: Decompile Binary Code with Large Language Models.**  
Hanzhuo Tan, Qi Luo, Jing Li, Yuqun Zhang. [Paper] [Code] [Models]
- [FSE'2023, CCF-A] **Enhancing Coverage-Guided Fuzzing via Phantom Program.**  
Mingyuan Wu, Kunqiu Chen, Qi Luo, Jiahong Xiang, Ji Qi, Junjie Chen, Heming Cui, Yuqun Zhang. [Paper]
- [arXiv preprint] **Condor: A Code Discriminator Integrating General Semantics with Code Details.**  
Qingyuan Liang, Zhao Zhang, Chen Liu, Zeyu Sun, Wenjie Zhang, Yizhou Chen, Zixiao Zhao, Qi Luo, Wentao Wang, Yanjie Jiang, Yingfei Xiong, Lu Zhang. [Paper]
- [arXiv preprint] **Grammar-Based Code Representation: Is It a Worthy Pursuit for LLMs? .**  
Qingyuan Liang, Zhao Zhang, Zeyu Sun, Zheng Lin, Qi Luo, Yueyi Xiao, Yizhou Chen, Yuqun Zhang, Haotian Zhang, Lu Zhang, Bin Chen, Yingfei Xiong. [Paper]

## RESEARCH EXPERIENCE

### SUSTech ARiSE Lab

Supervised by Prof. *Yuqun Zhang*

May.2022 - Aug.2022

- Worked on the effectiveness of fuzz testing for both binary applications and deep learning models.

## WORK EXPERIENCE

### Kwai

LLM Algorithm Engineer Intern

Beijing, China

Jun.2023 - Dec.2024

- Contributed to *Kwaipilot*, an LLM-powered code assistant plugin with code completion and Q&A capabilities.
- Conducted data collection, cleaning, classification, filtering, and processing of over 100TB domain code data.
- Conducted the evaluation of code LLMs by constructing diverse code completion benchmarks through extracting abstract syntax trees from code and human labeling, supporting over 10 programming languages.
- Conducted Continue Pre-Training and Post-Training on various base models, including StarCoder, DeepSeek, Gemma, Llama, etc., ranging in size from 1B to 70B parameters.
- Explored and designed the Block-FIM to enhance the model's ability to automatically stop completions. ‘
- Built a Retrieval-Augmented Generation system, boosting the code completion performance of LLMs.
- Conducted training of user behavior analysis models utilizing real auto-completion usage data, focusing on contextual filtering, to enhance the capabilities of LLMs on edge devices.
- Contributed to the data platform's AI development assistant, including SQL completion, SQL repair, and Text2SQL features, achieving over 2,000 DAU with an apply rate of 30%+.
- Provided intelligent code completion services to Xiaomi.
- During the internship, DAU increased from 60 to 4,000+, and the apply rate grew from 5% to 40%.

- Contributed to ADAS (Advanced Driver Assistance Systems), including FCW (Forward Collision Warning), LDW (Lane Departure Warning), and PCW (Pedestrian Collision Warning).
- Contributed to FCW. FCW utilizes object detection and regression algorithms to precisely locate vehicle position information in the driving area precisely. FCW can monitor collision risks in real-time during driving, integrated with ranging, modeling systems, and lane perception information. FCW can issue timely warning signals 2.7 seconds before danger occurs, effectively preventing traffic accidents such as rear-end collisions.
- Conducted an automated simulation testing tool based on Qt Creator. This tool can extract frame information from H.264 videos of real-world driving scenarios and use it as input for the algorithm pipeline, enabling the simulation of real-world testing and significantly improving algorithm testing efficiency.

- Developed back-end for a build pipeline client to streamline the microservices deployment process.

## OPEN-SOURCE CONTRIBUTIONS

---

- **LLM4Decompile:** Decompiling Binary Code with Large Language Models.
  - Accumulated over 5400 stars on GitHub.
  - Featured on Hacker News.
  - LLM4Decompile models have been downloaded over 100k times on Hugging Face.
  - Implemented at ByteDance as an IDA Pro/Ghidra plugin, serving reverse engineers.

## SKILLS

---

**General:** Python, C/C++, Java, Shell, SQL, Git, Matlab, LaTeX, etc.

**DevOps:** ElasticSearch, gRPC, Docker, Kubernetes, Qt Creator, Ghidra, faiss, Spark, vLLM, etc.

**MLOps:** Pytorch, WanDB, TensorFlow, LangChain, DeepSpeed, Megatron etc.

## SELECTED HONORS & AWARDS

---

- |   |                         |
|---|-------------------------|
| • Excellence Award, GBA CS Institutes Joint Poster Competition.             | <i>SUSTech, 2025</i>    |
| • SF Technology Scholarship ( $\sim 1\%$ ).                                 | <i>XDU, 2018</i>        |
| • Honorable Mention in the American Mathematical Contest in Modeling (MCM). | <i>XDU, 2018</i>        |
| • Second Prize in the National College Student Computer Design Competition. | <i>XDU, 2018</i>        |
| • Outstanding Student Award.  | <i>XDU, 2017</i>        |
| • Third Prize in ACM Collegiate Programming Contest, Shaanxi Province.      | <i>XDU, 2017</i>        |
| • Second Prize in the Provincial College Student Mathematics Competition.   | <i>XDU, 2016</i>        |
| • First-Class & Third-Class Merit Student Scholarship.                      | <i>XDU, 2015 - 2017</i> |

## TEACHING ASSISTANT

---

- **Object-oriented Analysis and Design:** Teaching Assistant for SUSTech Computer Science. *SUSTech, 2023*