Yuhe Liu

🇨🇳 Shenzhen (+86) 18800131229 junetheriver@gmail.com JuneTheRiver

EDUCATION

Tsinghua University	Computer Science and Technology (M.Eng.)	3.92/4.0	2023.09 - 2026.06
Tsinghua University	Computer Science and Technology (B.Eng.)	3.83/4.0	2019.09 - 2023.06

PROJECT EXPERIENCE

AIOps Automatic Evaluation System Construction *@ Huawei, CAS, CAICT* 2024.08 - 2025.08

- **AIOps Evaluation Framework:** Conducted a comprehensive study of AIOps-related LLM research and proposed a taxonomy for AIOps foundation models. Established a three-dimensional evaluation system based on "Language Ability × Operations Task × Data Modality", filling the gap in AIOps LLM benchmarking.
- **Automatic QA Generation:** Built an automatic benchmark generation pipeline using RAG and Agent techniques, generating domain-specific QA data from product documentation to enhance efficiency, diversity, and accuracy.
- **Deployment & Release:** Delivered and deployed the QA generation and model evaluation systems within Huawei. The evaluation framework was officially released on the CAICT benchmarking platform.

OpsEval: Benchmarking AIOps Capabilities of LLMs *@ NetMan Lab* 2023.09 - 2024.07

- **Dataset Construction:** Collected and curated 9,000+ AIOps QA pairs from public corpora and industry partners.
- **Leaderboard Development:** Designed and maintained the OpsEval leaderboard (https://opseval.cstcloud.cn/v2/)
- **Model Evaluation:** Evaluated 20+ mainstream LLMs using multiple paradigms, including LLM-as-a-Judge, RAGAS, and frequency-based analysis.
- **Publication:** First-author paper OpsEval accepted at FSE 2025 (CCF-A, Software Engineering top-tier conference).

Fine-tuning Large Language Models for the Telecom Domain *@ ZTE* 2023.11 - 2024.07

- **Project Background:** Addressed the performance gap of general-purpose LLMs in telecom-related QA tasks through domain-specific fine-tuning.
- **Data Preparation:** Collected and cleaned telecom-domain text data, and automatically constructed high-quality domain-specific QA pairs.
- **Fine-tuning Experiments:** Conducted fine-tuning using DeepSpeed and Megatron frameworks; compared parameter-efficient methods (LoRA, Freeze) and full fine-tuning under PT, SFT, and DPO training paradigms.

INTERNSHIP EXPERIENCE

Tencent TEG LLM-based Log Analysis for CI Pipeline *Algorithm Intern* 2024.08 - 2024.11

- **Project Background:** The department's CI/CD pipeline generates a large volume of error logs daily, which are analyzed by LLMs to identify root causes and propose solutions.
- **Log Compression:** Optimized the LLM-based abnormal log retrieval algorithm (keyword table + context window matching) to accurately extract core log entries.
- **Answer Evaluation:** Designed an evaluation framework (keyword matching, TF-IDF, and LLM-as-a-Judge) to automatically assess the accuracy and readability of model-generated analyses.

Huawei Terminal BG Android Malicious App Detection Based on User Feedback *Algorithm Intern* 2022.07 - 2022.09

- **Task:** Cleaned and analyzed tens of millions of user feedback entries to extract static and behavioral features.
- **Method:** Built a decision-tree baseline and experimented with deep neural networks, achieving a 12% improvement in F1 score for malicious app detection.

PUBLICATIONS

1. Eagle: A Comprehensive LLM Benchmarking Framework for Operations Capability

FSE 2026 (CCF-A) Lead Author

2. OpsEval: A Comprehensive Benchmark Suite for Evaluating LLMs' Capability in IT Operations Domain

FSE 2025 (CCF-A) Lead Author

3. Skeleton-CutMix: Mixing Up Skeleton with Probabilistic Bone Exchange for Supervised Domain Adaptation

TIP 2023 (CCF-B) Second Author

4. TechSupportEval: An Automated Evaluation Framework for Technical Support Question Answering

IJCNN 2025 Third Author

TECHNICAL SKILLS

- **Programming Languages**: Proficient in Python, C/C++, TypeScript, Golang, and Rust.
- **Algorithmic Foundations**: Deep understanding of common data structures and algorithms; familiar with mainstream machine learning methods (e.g., XGBoost, Logistic Regression, SVM) and deep learning architectures (e.g., Transformer, BERT, Diffusion models).
- **Large Language Models (LLMs)**: Familiar with LLM-related technologies; well-versed in the paradigms and principles of LLM-based RAG, Agent, and Evaluation frameworks. Experienced in building LLM evaluation benchmarks and deploying/optimizing LLM applications in vertical domains.
- **Language Proficiency**: English CET-6 (Score: 657). Strong academic writing and reading skills; capable of conducting fluent technical communication and academic presentations in English (presented at FSE 2025).

CAMPUS EXPERIENCE

Teaching Assistant, "Software Engineering" Course, Dept. of Computer Science 2023, 2024, 2025

- **Tech Stack:** Django, Docker, MySQL
- **Responsibilities:** Guided student teams in completing course projects, including an e-commerce platform, GIF-sharing platform, and paper annotation website.

Cluster Management, NetMan Lab 2023 – Present

- Responsible for internal network account and CPU/VM resource management.

Research Intern, Scene Reconstruction Lab 2022.06 – 2023.02

- Proposed a cross-domain skeleton augmentation framework that improved action recognition accuracy across age groups.
- Co-authored TIP 2023 (CCF-A) paper as second author.

Member, Network Department, Student Association of the CS Department 2020.09 – 2021.09

- Participated in maintenance of the departmental server room and the scholarship evaluation website.
- Served as lecturer during the summer training program, teaching "Introduction to Linux."

Academic Excellence Scholarship 2021.12, 2022.12

COURSE PROJECTS

Computer Graphics: Ray Tracing 2021.03 - 2021.06

- Implemented Progressive Photon Mapping (PPM) rendering algorithm

Computer Networks Laboratory 2021.09 - 2021.12

- Implemented RIPng protocol for IPv6 (Documentation Link)

Computer Architecture: rCore Operating System 2022.03 - 2022.06

- Developed an operating system based on the Rust-based rCore framework (Documentation Link)