

# PUFAN ZUO

ResearchGate ◇ GitHub ◇ LinkedIn

Phone: +86 155-9684-8180 ◇ Email: [z3253345336@gmail.com](mailto:z3253345336@gmail.com)

Homepage: [zac-saodiseng.github.io](https://zac-saodiseng.github.io)

## EDUCATION

---

**Northeastern University** (985/211 Project, Double First-Class)  
B.S. in Computer Science and Technology

*Sep 2022 - Jul 2026*  
*China*

**Core Courses:**

C++ Programming (A/92), Data Structures (A/90), Network Fundamentals (A/91)

## RESEARCH INTERESTS

---

**Data Systems:** Privacy-preserving Data Systems, Machine Learning for Systems, Cloud Database, Streaming Database, Vector Database, Data Integration and Cleaning, Distributed OLTP/OLAP, Query Optimization

**Data Science:** Anomaly Detection, Intelligent Data Systems, Big Data Analytics, ETL, Interpretable Machine Learning

## PUBLICATIONS AND PREPRINTS

---

Yu, S., Lin, S., Gong, S., Xie, Y., Liu, R., Zhou, Y., **Zuo, P.**, Zhang, Y., Sun, J., & Yu, G. (2025). *Updating Graph-based Index with Fine-grained Blocks for Large-scale Streaming High-dimensional Vectors*. Preprint. [arXiv:2503.00402v1](https://arxiv.org/abs/2503.00402v1)

## PROFESSIONAL EXPERIENCE

---

**CelerData (StarRocks)**

*SaaS Team R&D Engineer*

Feb 2025 - Present

Menlo Park, CA

- Designed ecosystem-specific data integration solutions with kernel-level optimizations for StarRocks database engine
- Led Apache Arrow & RPC framework integration achieving 20× general query speedup (peak 160×), with open-source contributions : [github.com/StarRocks/starrocks/pull/57956](https://github.com/StarRocks/starrocks/pull/57956)
- Developed official BI connectors: StarRocks Tableau JDBC Connector and dbt-starrocks adapter

**Kuaishou Technology**

*Infrastructure R&D Engineer*

Jul 2024 - Dec 2024

Beijing, China

- Contributed to the development of KDev, Kuaishou's in-house API testing platform, including task scheduling for KAPI (Kuaishou's internal API platform) to support automated data migration between versions and the cleanup of empty directories generated by the platform.
- Participated in the design of the HTTP API Swagger parsing framework, enabling the parsing of Swagger 3.0 specifications via OpenAPI to extract low-level data types and enumerated values.
- Implemented support for file upload types in the API mocking and proxy forwarding module, and introduced dynamic Kconf configuration with upload size limits of up to 200MB.

**Meituan**

*Backend Engineer, Resource Scheduling*

Mar 2024 - Jul 2024

Beijing, China

- Participated in the foundational development of Strategy Analysis Tool 2.0, a Meituan in-house ToB platform for business strategy evaluation and optimization.
- Utilized Lion, Meituan's distributed configuration management system, to dynamically adjust the UI layout of the 1.0 version and conduct A/B testing for strategy optimization, comparing different matching strategies across adjacent data slices to improve overall service quality.

- Mitigated the risk of large-scale data growth (250GB per table) by implementing sharding and tiered storage using MySQL for hot data and Blade (an internal storage system built on TiDB) for cold data management.

**TutorEva**  
*Algorithm Engineer, Mathematical Reasoning*

Jan 2024 - Mar 2024  
Beijing, China

- Developed syntax parsing and semantic mapping algorithms for middle and high school mathematical expressions, enabling accurate interpretation and transformation of mathematical formulas.
- Improved and optimized the core framework and underlying algorithms for symbolic algebra computation, enhancing the system’s accuracy, performance, and extensibility.
- Participated in the design and implementation of an automated problem-solving and logical reasoning framework based on Drools.

**Yonyou Hong Kong**  
*Chanjet Team R&D, DevOps*

Nov 2023 - Jan 2024  
Hong Kong, China

- Participate in the quality assurance tasks of the marketing system, participate in project requirements review, write complete test cases and conduct testing.
- Read the source code of project, find and submit potential defects and loopholes, and assist the RD team in solving project BUG.
- Participate in the development of automated api testing framework and improve the testing scenarios.

RESEARCH EXPERIENCE

**iDC-NEU Research Group**  
*Research Intern*

Jul 2024 - Mar 2025  
Prof. Yanfeng Zhang and Prof. Shufeng Gong

- Software-Hardware Cooperative Optimization
- Researched SPANN and SPFresh, and optimized vector database system I/O using SPDK.

**Artificial Intelligence Lab of NEU**  
*Research Intern*

Mar 2023 - Jul 2023  
Prof. Miao Fang

- Learned deep learning fundamentals such as logistic regression, regularization, and CNN.

ACHIEVEMENTS

**StarRocks Active Contributor**  
Core maintainer for 10K+ star MPP database project

Feb 2025 - Present

**HUAWEI Student Developers(HSD)**

Mar 2023 - Present

**National Third Prize**  
Intelligent Computing Innovation Design, System Development Competition

Jul 2024 - Nov 2024

TECHNICAL SKILLS

Certifications	TOEFL 103, GRE 332 (Q:168, V:164, AW:5.0)
Languages	C/C++, Python, Java, Golang, SQL, CUDA, C#, MATLAB, HTML, CSS, JavaScript
Databases	StarRocks, MySQL, TiDB, MongoDB, Redis, Databricks
Frameworks	PyTorch, Spring (MVC/Boot/Cloud/Security) , Flask, Apache Arrow, MyBatis, Dubbo
Systems	Linux/Unix, Docker, Distributed Systems, MPP
Tools	Git, LaTeX, Spark, Kafka, Selenium, DBT, Tabular, FIO, SPDK, Maven, Nginx, Drools