Zhongming Zuo

(984) 335-8262 | zz391@duke.edu | linkedin.com/in/zhongming-zuo/

EDUCATION

Duke University   Durham, NC, USA
Master of Science in Computer Science   Aug 2024 - May 2026
GPA: 3.96/4.00
Core Courses: Database Systems, Distributed Systems, Systems for Machine Learning, Computer Vision, Large Language Models, Data Science, Mobile App Development, Advanced Web App Development

Nanjing University   Nanjing, China
Bachelor of Science in Computer Science and Technology   Sep 2020 - Jun 2024
GPA: 4.49/5.00
Core Courses: Data Structures, Algorithm Design and Analysis, Computer Networks, Operating Systems, Digital Logic and Computer Organization, Principles and Techniques of Compilers, Big Data Processing: Comprehensive Experiments, AI Computing Systems, Numerical Method, Probability and Mathematical Statistics
Honors and Awards: Outstanding Graduate, 2024; Academic Scholarship (First Class), 2023; Academic Scholarship (Second Class), 2022; Academic Scholarship (First Class), 2021

SKILLS

Programming Languages: C++, C, Python, Java, Swift, JavaScript, TypeScript
Development Frameworks: React, React Native, Vue.js, Nuxt, Node.js, SwiftUI, Vapor
Databases & Big Data: PostgreSQL, MySQL, MongoDB, Spark, Hadoop MapReduce, Hive
Machine Learning: PyTorch, Pandas, scikit-learn, NumPy

WORK EXPERIENCE

WeRide   San Jose, CA, USA
Software Engineer Intern   May 2025 - Aug 2025
- **Local Pullover Module:**
- Developed a local planning module in C++, enabling autonomous vehicles to safely pull over within the local semantic map range, providing a reliable fallback for global planner failure.
- Designed and implemented mechanisms to detect global planner failure, locate the pullover destination, construct the local MDP, and generate optimal local routes.
- **Routing Service Improvement:**
- Optimized routing and pickup-point locating algorithms to improve the performance of the routing service, achieving $> 2\times$ QPS increase.
- Enhanced the routing and locating pipelines by integrating bidirectional Dijkstra, refining curb intersection detection, and reordering pickup-point evaluation criteria.

RESEARCH EXPERIENCE

Department of Systems Engineering and Engineering Management, CUHK   Jul 2023 - Feb 2024
Junior Research Assistant
- **Key Innovation:** Proposed a novel vertex sampling method for the butterfly counting problem on bipartite graphs. Utilized this method to design algorithms for three different data streaming models, achieving up to 700x lower error rates compared to state-of-the-art algorithms.
- **Theoretical Analysis:** Theoretically analyzed the performance bottlenecks of butterfly counting algorithms for large-scale bipartite graphs through mathematical derivation.
- **Algorithm Design:** Designed three vertex-sampling-based approximate algorithms for butterfly counting under the two-pass, random arrival, and sliding window model, with better theoretical accuracy than existing methods.
- **Implementation:** Implemented all the designed algorithms in C++, evaluated on real-world graphs, benchmarked against established methods, and verified better accuracy performances.
- **Unpublished Manuscript:** Effective Approximation Algorithms for Butterfly Counting in Bipartite Graph Streams. Fangyuan Zhang, Zhongming Zuo, Qintian Guo, Rui Zhang, Sibo Wang.

Department of Computer Science and Technology, Nanjing University   Sep 2022 - Jun 2023
Undergraduate Research Assistant

- Conducted data collection, data processing, and algorithm evaluation, contributing to the construction of a web data retrieval system.
- **Data Collection and Processing:** Used Python (with Beautiful Soup library) for web scraping to build a comprehensive dataset on web page data. Conducted data cleaning with MySQL and Java.
- **Text Classification:** Conducted literature research on existing text classification algorithms and tested different multi-label text classifiers on the collected datasets to classify web pages according to related texts. Applied metrics such as precision, recall, and F1-score for evaluation.

PROJECTS

SyncNote - Collaborative Note Application   Nov 2025 - Dec 2025

- Built a collaborative Markdown editor in a team of two using Nuxt (Vue + Nitro), featuring GitLab OAuth, MongoDB Atlas storage, document CRUD, sharing/co-editing, and live preview/syntax highlighting.
- Implemented real-time multi-user collaboration with Socket.IO, scaled horizontally via the Socket.IO Redis adapter, and ensured WebSocket stability with sticky sessions.
- Deployed on Kubernetes with multi-replica pods behind a LoadBalancer; automated GitLab CI/CD with werf build/deploy plus Playwright E2E tests.

LLM Fine-Tuning & Evaluation   Nov 2025 - Dec 2025

- Built an end-to-end fine-tuning & evaluation pipeline for gemma-3-270m in a team of five, targeting abilities across factual QA, reasoning, and instruction following.
- Implemented supervised fine-tuning with configurable dataset mixing; utilized reinforcement learning (RLOO) to explore alignment gains and analyzed tradeoffs.
- Enhanced inference with RAG based on BM25 and Mixture-of-Experts routing over multiple expert models trained with specialized data.

HiFOSS - Open-Source Forum Application   March 2025 - April 2025

- Developed a cross-platform open-source forum App with React Native in a team of four, featuring post creation, comments, favorites, and private messaging.
- Designed and implemented front-end functions for post display, creation, and commenting, interacting with a Java back-end for data updates and synchronization.

Frontdesk - Assistance Platform for Duke Hospital   Oct 2024 - Dec 2024

- Developed an iOS App in Swift for medical trainees and faculty of Duke Hospital in a team of three, facilitating information entry, communication, evaluation, and scheduling.
- **Front-end Development:** Implemented the complete front-end functions for trainee information entry, trainee evaluation, and surgical team scheduling using SwiftUI.
- **Back-end Development:** Developed the back-end system using Vapor and PostgreSQL, integrating CRUD operations, relationship handling, and availability management. Implemented RESTful APIs to handle different client requests.

Efficient Web Log Analysis with MapReduce   Jun 2023 - Jul 2023

- Designed and implemented a program for analyzing large-scale website logs using Java and Hadoop MapReduce, extracting key website metrics and optimizing data processing efficiency.
- Derived insights on website features such as user demographics and peak access time periods, achieving a clearer understanding of user behavior and discovering potential improvements for the target website.

Compiler Design and Implementation   Mar 2023 - May 2023

- Led a team of two to develop a robust C compiler using C, Flex, and Bison, incorporating lexical analysis, syntax analysis, semantic analysis, intermediate code generation, intermediate code optimization, and code generation.
- Designed the compiler structure and modules, organized team collaboration.
- Completed the whole implementation of lexical analysis and intermediate code optimization, and contributed to the remaining compilation stages.