+61 421436108 — email: Ericsgc@outlook.com

Guocheng Song

🌐 Portfolio

Summary

Results-oriented Software Engineer specializing in full-stack applications with Golang and TypeScript, and researching database optimization and distributed storage systems with C++ and Rust. Demonstrates strong execution skills, teamwork, and high learning efficiency.

Work Experience

WuXi AppTec, Shanghai, China

Laboratory Assistant (Intern)   *2022.6 - 2023.1*

- Supported **AI-driven drug discovery** programmes, merging deep-learning QSAR models with high-throughput reaction data to accelerate candidate screening.
- Built predictive **computational chemistry** pipelines for reaction yield and property forecasting, utilising PyTorch Geometric on 100k+ curated reactions.
- Synthesised and purified novel organic intermediates, validating in-silico predictions through bench-scale experiments.
- Automated data capture and lab reporting, boosting experimental turnaround by 30% .

Seetrum, Shanghai, China

AI Developer Engineer, Part-time   *2023.3 - 2023.11*

- Built full-stack **web portal** for spectroscopy-chip products: **React** front-end and **Typescript** backend exposing RESTful APIs.
- Trained deep-learning models (CNN & Transformer) for **spectral fingerprint recognition** and chemical composition prediction, boosting accuracy by 20% on internal dataset.
- Automated data pipeline and CI/CD (Python, SQL, GitHub Actions) for continual model retraining and deployment to ARM edge devices.
- Integrated on-chip inference API with hardware team, achieving real-time classification under 150 ms latency.

Datap.ai, Sydney, Australia

Full-stack Developer, Part-time   *2024.7 - 2025.01*

- Led backend development using *Go* and *Python* to build an AI-powered data pipeline system.
- Implemented ML-based ETL automation with *Airbyte* and *DBT*, and developed agents for natural language to code conversion.
- Enhanced front-end performance using *React* for improved data visualization.
- Collaborated with cross-functional teams to integrate AI/ML features, data pipelines, and front-end components.

UNSW Data and Knowledge Research Group, Sydney, Australia

Research Assistant, Part-time   *2025.2 - Present*

- Researching database optimization techniques, focusing on SQL query optimization and low-level performance enhancements.
- Participated in research on **SIMD-accelerated set intersection algorithms** to improve performance of database query operations.
- Participated in research exploring **JIT compilation techniques** to enhance analytical query execution efficiency.

Projects

Financial AI – Multi-Agent Investment Platform ⏻ GitHub   *Co-Lead Developer*

- Co-led an open-source initiative to deliver an **AI-driven financial analytics suite** that automates market insight, risk scoring and anti-fraud monitoring.
- Architected a *multi-agent system* with Spring AI: *RecordRetriever*, *StockAnalyzer* and *WeatherAnalyzer* collaborate through explicit **function calls**, sharing vectorised context to drive collective reasoning and **real-time decision making**.
- Implemented a **Retrieval-Augmented Generation** chatbot using Chroma DB for vector storage, enabling natural–language investment queries and on-the-fly portfolio advice.
- Exposed a clean set of **RESTful APIs** (Spring Boot 3) that orchestrate analysis → prediction → action pipelines; secured with Spring Security 6 + JWT.
- Designed data plane with MySQL → TiDB migration path, **Redis** caching and **RabbitMQ** event bus; leveraged Elasticsearch + Kibana for end-to-end observability of agent decisions.
- Deployed **streaming anomaly detection** to flag suspicious transfers within 200 ms SLA, reducing false positives by 35% in simulated benchmarks.

NoKV - Key-Value Store Engine ⏻ GitHub

- Designed and implemented a high-performance key-value storage system in *Go*, achieving 30% improvement in data retrieval speed through optimized LSM tree architecture and *parallel compaction*.
- Designed and implemented an *LSM tree storage* engine with multi-level caching, BloomFilter optimization, and WAL for crash recovery.
- Implemented *Whisckey*, a key-value separated storage system, utilizing a lock-free skiplist as the structure for the memtable.
- Implemented *MVCC transaction* support with snapshot isolation.
- Key features: KV separation, parallel compaction, W-TinyLFU cache.

QuillSQL - SQL Database ⏻ GitHub

- Developed a minimalist **relational SQL database** in **Rust**, inspired by CMU 15-445's BusTub project.
- Designed a well-structured **three-layer architecture**:
  - **SQL Layer**: Parses queries into Abstract Syntax Trees (AST), constructs logical plans, applies rule-based optimisation, and generates physical execution plans.
  - **Execution Layer**: Implements a *Volcano-style* iterator model to execute physical plans.
  - **Storage Layer**: Supports modular storage engines, including *B+Tree* and *Bitcask*, enabling snapshot isolation and full ACID compliance.
- Provides an interactive command-line interface for executing SQL queries and managing data.

GoChat - Distributed IM system ⏻ GitHub

- Designed a distributed instant messaging platform in **Go**, with a lightweight, console-based user interface and robust client-server model.
- Applied **Domain-Driven Design (DDD)** to modularise services and isolate responsibilities across the system.
- Implemented **end-to-end message reliability** guarantees, ensuring safe and complete delivery.
- Built **resilient network handling**, featuring intelligent access point selection for unstable or low-connectivity environments.
- Enabled **zero-downtime hot deployment** for backend services, with long-lived client connections remaining unaffected.

Skills

Languages: Python, **Go**, Rust, **TypeScript**, C/C++
AI/ML: PyTorch, TensorFlow, CUDA, cuDNN, Hugging Face Transformers
Distributed/HPC: MPI, OpenMP, Ray, Dask, Spark
Frontend: React, Next.js
Backend: Gin, Node.js, Spring Boot, gRPC
Databases: MySQL, PostgreSQL, Redis, Elasticsearch, TiDB
DevOps/Cloud: Docker, Kubernetes, GitHub Actions, AWS, Linux

Certificates and Awards

Kaggle Competition: Chatbot Arena Human Preference Predictions (silver medal)   2024.8

Education

University of Sydney   **Australia**
Master of Computer Science (Advanced Entry)   *February 2024 - Present*
Specialisation: Software Engineering

University College London   **United Kingdom**
Bachelor of Science in Chemistry   *September 2020 - June 2023*