Yuhao lu

Richmond, BC, Canada | (+1) 6043669214 | lu.yuhao@northeastern.edu

EDUCATION

University of Northeastern Sept 2023 – May 2026
Master of Computer Science Vancouver, Canada

SKILLS

Language: Java, C, C#, C++, Python, JavaScript, TypeScript, SQL, Shell, HTML, CSS, Bash, Pytorch
Frameworks/Databases: Spring Boot, MyBatis, Express, .Net, MySQL, PostgreSQL, MongoDB, Redis
DevOps/Cloud: Docker, Kubernetes, GitLab, CI/CD, AWS, GCP, Zookeeper, Kafka, RabbitMQ, RocketMQ
Frontend/Testing/Tools: React, Redux, Git, Linux, Maven, IDEA

EXPERIENCE

Northeastern University Sep 2024 – Dec 2024
Research Assistant Vancouver, Canada

- Processed over **500** audiofiles by trimming to **30** seconds and augmented data using noise addition, pitch shifting, and time stretching, improving dataset robustness by **20%**.
- Incorporated **WavLM-Large** for audio processing and **TinyLlama** for language understanding(LLM), reducing computational resource usage by **30-50%** compared to larger models.
- **Fine-tuned** the linear projector across 50 epochs with 2.5k steps, achieving a **10%** decrease invalidation loss across experiments.
- Designed and evaluated three experimental setups—audio-only, text-only, and **multimodal**—demonstrating a **5.7%** improvement in F1-score with multimodal analysis over single-modality approaches.
- Engineered the SLAM-HATE-SPEECH model, integrating audio and text data for hate speech detection on social VR platforms, achieving an **F1-score** of **0.91** and **accuracy** of **91.0%**. Conducted **2-fold cross-validation** to validate model performance, focusing on precision, recall, and F1-score to balance false positives and false negatives.

North Carolina State University Jul 2023 – Aug 2023
Research Assistant NC, USA

- Collect data from PV rebate in China and PV installed capacity in China and use **NumPy** and **Pandas** to process the 50k+ data injupyter notebook, including data cleansing, processing missing values and outliers. And draw two pictures to express the date in charts.
- Use time series analysis model to do predict. According the **gray prediction model**.
- Executed a data storage solution by transferring image and video data to **AWS S3**, which significantly reduced storage costs by 70% and employed data lifecycle policies and version control for automated data management.
- Leveraged **Prometheus** and **Grafana** to build a monitoring platform that swiftly detects and responds to system failures within minutes, ensuring business continuity.

Zhonghui Certified Public Accountants LLP Jul 2022 – Aug 2022
Risk Consulting Department, Internship Hangzhou, China

- Used **SQL** to import and extract large amounts ofdata, including company financial statements, billing flows, contracts, and various meeting minutes, and wrote SQL scripts to complete data validation. Completed data cleansing, processing, including **missing values and outliers**, and detailed data analysis pipeline for Erie's data.

PROJECTS

Distributed File Storage System Jun 2024 – Aug 2024

- Developed a distributed file storage system with high availability and strong consistency. The stress test showed **20,000** QPS for 4KB file mixed read and write, and the P99 delay is **800** milliseconds. Implemented a file client that encapsulates file content into KV requests and sends them to the backend, with support for **zlib** for lossless content compression.
- Implemented the **Raft** consensus algorithm, with core functions such as Leader election, log replication, and snapshot update. Based on the consistent hashing architecture, the data is partitioned into Shards and can be migrated in multiple Raft Groups.
- Implemented support for storage engines such as **RocksDB**, B-tree, and hash tables to adapt to scenarios with different IO models. Based on asynchronous Apply, **ReadIndex**, and **FollowerRead** to optimize read performance, and based on **Prevote** to avoid frequent master switching.

High Performance Gaming Platform Feb 2024 – Apr 2024

- Built a high-performance gaming web platform with a backend built on **SpringBoot**, **MyBatis**, **Redis**, **AWS RDS**, and **Kafka**, and a frontend built with **React**. Supported up to 10,000 concurrent users, handled over **3,000** transactions per second, and maintained a P99 latency ofless than 1 second.
- Implemented databasesharding and table partitioning using **Amazon RDS** to effectively model user data, game resources, and transaction records, and enhanced query performance by using indexing and partitioning techniques.
- Cached data in Amazon ElastiCache for **Redis**, reducing query latency from an average of200ms to an average of40ms. Configured TTL to manage data expiration.Optimized **Kafka** partition and replica configurations, enhancing the speed of asynchronous message processing and eliminate traffic peaks, and increased throughput from 300 messages per second to 1200 messages per second.