Ryan Zhao

ryanzhao@stanford.edu | 203-859-0643 | github.com/KnightAsterial

Education

Stanford University | *Class of 2025*
MS in Computer Science, GPA: 4.077

University of California, Berkeley | *Class of 2023*
BA in Computer Science, GPA: 4.00

Skills

Languages: Java, Go, C, Python, C++, JavaScript, SQL

Relevant Tools: PyTorch, Pandas, LangChain, Amazon Web Services, MapReduce, PostgreSQL, Cassandra, MongoDB, gRPC, Docker, SLURM

Research Experience

Kundaje Lab, Stanford | Graduate Student Researcher **October 2023 - Present**

- Leveraging ChromBPNet models and 3D heart organoids to decipher transcription factor motifs related to cardiogenesis and identify causal, non-coding mutations linked with congenital heart disease
- Applying contrastive learning with protein language models to predict T-cell receptor/epitope binding and improve generalization to previously unseen epitopes
- Designing multi-agent LLM system capable of utilizing bioinformatics tools to improve the speed of discovery of causal mechanisms between rare genetic variants and disease phenotypes
- Refined pipeline for peak calling, bigWig generation, and model training at scale from scATAC-seq data

UC Berkeley RISE Lab | Research Assistant **August 2021 - March 2022**

- Designed TAOBench benchmark to simulate industry query distributions for distributed databases
- Developed drivers in C++ to run workload on 4 distributed database systems and evaluated performance
- Published in VLDB 2022 alongside advisor Audrey Cheng, supervised by Ion Stoica and Natacha Crooks

Industry Experience

Cockroach Labs | Software Engineer Intern **May 2022 - August 2022**

- Researched bottlenecks in CockroachDB's snapshot process, a mechanism that allows the distributed database to replicate and transfer data across all nodes in the cluster
- Designed and implemented experiments to determine the primary hardware resources used by snapshots and how snapshot transfers affect the tail latency of foreground user traffic
- Identified future area of focus and directly informed next steps on admission control for snapshots
- Used Golang and gRPC for database code and leveraged Prometheus and Grafana to run experiments

Amazon (AWS) | Software Development Intern **May 2021 - August 2021**

- Built IOPS canary for AWS Elastic File System (EFS) to detect regressions in promised IO throughput
- Designed a multithreaded poller to perform block-level file operations and push loads of up to 300 MiB/s
- Managed proper inode metadata to durably track allocated blocks and prevent memory leakage
- Engineered custom build and deployment process using Apache Ant and internal Amazon tooling

DataStax | Developer through Berkeley Codebase **January 2020 - May 2020**

- Built fault-tolerant database migration service to transfer existing Cassandra databases to the cloud
- Designed and created a multithreaded Go proxy that parsed Cassandra's binary protocol to direct reads/writes during migration for zero user downtime
- Implemented custom communication channel between migration and proxy to coordinate operations

Projects

Synthetic Free-Text EHR Generation with LLM Fine-tuning **June 2024**

- Fine-tuned BioMistral-7B using the MIMIC-III dataset to generate synthetic nurse's notes conditioned on patient ICD codes. Evaluated model's potential to improve downstream model training through data augmentation. Implemented training with 4-bit quantization (QLoRA) and performed comprehensive hyperparameter sweep to achieve training stability on 40GB A100 GPU