Aryan Ghobadi

Chicago, IL | U.S. Permanent Resident | 317-998-6315 | aryanghbd@gmail.com
www.aryanghobadi.com | linkedin.com/in/aryan-ghobadi | github.com/aryanghbd

Education

Imperial College London Sept. 2019 – June 2023

MEng, Electronic & Information Engineering (Integrated Master's) *London, UK*

U.S. GPA equivalents: 3.5 (undergraduate years 2–3), 3.8 (master's year).

Selected coursework: Advanced Computer Architecture, Instruction Architectures and Compilers, Software Systems, Embedded Systems, Discrete Mathematics, Optimisation, Coding Theory, Principles of Distributed Ledgers, Deep Learning.

Experience

Certara Dec. 2023 – Present

Software Engineer, ML Infrastructure *US (Remote)*

- Serve as one of the **lead backend engineers and a technical architecture team member** for Nexus, Certara's enterprise AI platform; drive cross-service design and major backend refactors across five projects spanning JVM services, Python inference and data infrastructure.
- Re-architected asynchronous agent and workflow execution in Spring Boot/Kotlin, replacing process-local generation state with durable jobs, idempotent submissions and cancellation to support recovery across worker restarts and replicas.
- Led the refactor of scientific data ingestion spanning **tens of millions of documents** and several hundred thousand daily ingestions from PubMed, PMC Open Access and USPTO; rebuilt metadata extraction and Elasticsearch indexing, cutting full-source indexing from weeks to hours.
- Raised document-ingestion throughput by **35**× through stage-level profiling and targeted replacement of CPU-bound preprocessing components.
- Expanded JVM orchestration engines with parallel fan-out, agent tool loops and composable model/data operations; implemented Temporal-backed model execution. Cut average map-reduce summarization runtime by **50%** on 100+ document sets through decoupled preprocessing and parallel inference.
- Built a **multi-provider model control plane** across AWS Bedrock, Gemini, Anthropic and OpenAI, unifying model discovery, metadata and native tool calling; optimized self-hosted vLLM inference through GPU-memory tuning and quantization.
- Strengthened service reliability across Kubernetes, RabbitMQ and Redis/KeyDB through dependency health checks and recovery mechanisms; resolved **30+ production escalations** as a designated backend engineer for releases.
- Mentored a junior engineer to independent ownership of multiple repositories; recognized company-wide as *Rockstar of the Month*.

Huawei R&D UK Apr. 2022 – Oct. 2022

Software Engineer Intern *Cambridge, UK*

- Developed and tested RISC-V acceleration modules using C/C++ and QEMU for Secure Element OS workloads.
- Profiled and optimized latency-critical C/C++ paths in a Secure Element OS, cutting key runtimes by **10×+**; migrated manual tests into CI/CD, reducing test effort by an estimated **80%**.

Certara Aug. 2021 – Sept. 2021

Software Engineer Intern *Sheffield, UK*

- Designed a Python automation framework for pharmaceutical trial simulations, achieving a **90× latency reduction**.

Selected Projects

- **Itchy** *(C++20):* Built a NASDAQ TotalView-ITCH 5.0 decoder and L3 order-book replay engine; increased replay throughput by **5.9**× through profiling, pooled allocation and intrusive FIFO price levels, cutting replay time for 269M messages from roughly 10 minutes to under 2 minutes.
- **The Rusty Vault** *(Rust/Tokio):* Built a concurrent cryptocurrency market-data gateway ingesting Coinbase, Kraken and Binance WebSocket feeds; normalized trades, quotes and L2 book events into a shared schema using bounded asynchronous channels to decouple exchange adapters from downstream processing.

Technical Skills

Languages: Python, C/C++, Rust, Java, Kotlin, Groovy, SQL, TypeScript, Bash

Backend & Systems: Spring Boot, Grails, Temporal, FastAPI, REST/SSE, RabbitMQ, Redis/KeyDB, performance profiling

Data & Infrastructure: PostgreSQL, Trino, Elasticsearch, Weaviate, Kubernetes, Docker, Helm, CircleCI, AWS, Azure, GCP