Kanishk Agarwal

Bangalore, India
+91-8777082770 | agarwalkanishk12345@gmail.com
LinkedIn | GitHub | LeetCode | Portfolio

PROFESSIONAL SUMMARY

Backend Engineer building **high-performance concurrent systems** with **AI-assisted orchestration** and evolving distributed system capabilities. Experienced in designing **event-driven architectures**, optimizing latency by **7–10x**, and enforcing **strong system correctness** through lifecycle invariants and concurrency control. Built **LLM-integrated orchestration systems** across **20+ microservices**, improving operational efficiency by **more than 40%** while maintaining deterministic execution guarantees.

TECHNICAL SKILLS

Languages: C++, C#, Python, SQL
Core Systems: Multithreading, Concurrency Control, Thread Pools, Mutex, Condition Variables, RAII, Rule of Five
Backend: REST APIs, Microservices, Event-Driven Architecture, WebSockets
Distributed Systems: Message Queues, Kafka (Foundations), Producer-Consumer Models, Delivery Semantics
Databases: PostgreSQL, MySQL, MongoDB, Indexing, Query Optimization, Caching
AI Systems: LLM Systems, LangGraph, MCP, RAG, Vector Search
Tools: Docker, Git, CI/CD, AWS

PROFESSIONAL EXPERIENCE

Software Development Engineer I Jan 2025 – Present
Harman International, Bangalore

- Eliminated **more than 90% invalid booking states** by designing **lifecycle-driven invariants**, reducing manual correction effort by **more than 40%**.
- Ensured correctness under **high-concurrency workloads** using transactional boundaries and locking strategies, preventing **race conditions**.
- Reduced booking latency from **6s to 0.8s** by replacing polling-heavy APIs with **event-driven WebSocket architecture**.
- Reduced end-to-end workflow latency from **8s to 0.7s** by isolating **critical paths** and offloading non-blocking operations.
- Prevented request pile-ups under load using a **task-queue execution model**, improving **throughput and responsiveness**.
- Reduced database latency by **approximately 25%** through indexing and **query optimization**.
- Reduced regression issues by **more than 30%** via **decoupling** and service-layer abstractions.
- Improved observability by surfacing failures, reducing debugging time by **approximately 60%**.
- Built **LLM-driven orchestration pipelines** to dynamically route tasks across **microservices**.
- Owned backend workflows for **high-concurrency systems**, ensuring **correctness and reliability**.

Machine Learning Engineer Intern Jan 2024 – May 2024
Assisto Technologies Ltd., Pune

- Designed and deployed **RAG pipelines** integrating LLM inference with retrieval systems.
- Reduced inference latency by **approximately 30%** through pipeline optimization.
- Implemented **vector-based semantic search** improving retrieval relevance.
- Integrated **speech processing pipelines** into production workflows.

PROJECTS

AI-Assisted Task Orchestration Engine (C++)

- Designed a **DAG-based task orchestration engine** enabling dependency-aware execution with **deterministic lifecycle management**.
- Built a **multithreaded execution engine (thread pool)** optimizing parallel task processing under contention.
- Implemented **at-least-once execution semantics** with retry and failure handling.
- Designed **idempotent execution strategies** to ensure correctness under duplicate processing.
- Enabled **crash recovery using persistent storage**, ensuring no task loss across restarts.
- Developed a **producer-consumer queue system** as a foundation for **Kafka-based distributed execution**.
- Built an **orchestration layer** handling dependency resolution, scheduling, and state transitions.
- Integrated an **LLM-assisted advisory layer** with strict **control and execution separation**.

Custom STL-like Vector (C++)

- Implemented a generic vector<T> container with dynamic resizing and contiguous memory layout.
- Designed amortized constant-time growth improving performance and cache locality.
- Implemented full Rule of Five ensuring safe memory management.
- Built iterator support and optimized insertion and deletion.

EDUCATION

B.Tech in Computer Science (AI Specialization) 2020 – 2024
MIT ADT University, Pune
CGPA: 8.0