



Aaryan Mahajan

📍 DEL, India | 📞 +91-9810381171 | ✉ aaryanmhjn@gmail.com | 🌐 aaryanmahajan | 📁 Portfolio

🔗 LinkedIn | 🐙 Github | 🏆 LeetCode | 📝 Blog

EDUCATION

Course	Institute	Year	CGPA%	Rank
B.Tech (Computer Science & Engineering)	Netaji Subhas University of Technology	Nov 2022 - May 2026	8.53	40/250

PROFESSIONAL EXPERIENCE

- **Software Engineering Intern | Editorialist | Gurugram** Jan 2026 - Present
Engineered backend API systems and concurrency infrastructure for a production app serving 3k+ users.
 - Engineered backend API contracts and synchronization workflows consumed by iOS client across **20+ endpoints**.
 - Migrated entire codebase and all submodules from **Swift 5** → **Swift 6**, introducing actor-based concurrency model.
 - Optimized request lifecycle via async processing and response caching, reducing redundant API calls by **35%**.
 - Reduced client-server response latency by **25%** through API contract redesign and payload optimization.
- **AI Agent Developer Intern | Fast Retailing | Japan (ONSITE)** May 2025 - Jul 2025
Designed and deployed backend knowledge systems handling large-scale enterprise documentation workflows.
 - Architected an agentic RAG backend on GCP, owning the retrieval pipeline end-to-end over **100k+ enterprise docs**.
 - Built ingestion, chunking, embedding, and indexing pipelines with async workers, cutting retrieval latency by **70%** (217s → 65s).
 - Migrated org-wide documentation from Confluence to SharePoint via **MS Graph API**, improving data consistency.
 - Delivered system optimizations saving **120+ engineering man-days/month** across multiple internal teams.
- **Software Development Engineering Intern | Zimyo | Gurugram (ONSITE)** May 2024 - Jul 2024
Built and optimized backend services powering payroll and HR workflows at scale.
 - Designed query layer and snapshot-based aggregation view over **sharded distributed payroll backend**.
 - Built async background job pipelines for email and payroll processing, improving throughput by **30%**.
 - Owned end-to-end feature delivery across 3 modules including **UI completion logic** and flag-based render control.
 - Shipped features via CI/CD pipelines with weekly releases and **<5% regression rate**.

PROJECTS

- **Aqueduct: Distributed Durable Task Queue & Scheduler with Lease-Based Recovery** | GitHub
 - Built a durable job queue in **Go + PostgreSQL** using **SELECT FOR UPDATE SKIP LOCKED**, **686 jobs/sec** single-node.
 - Engineered **at-least-once** delivery via lease-based claims and a heartbeat **reaper** reclaiming jobs from dead workers.
 - Validated via a **custom chaos harness**: 4 process replicas, **100k jobs**, 117 SIGKILLs, zero loss, 0.10% redelivery.
 - Exposed **Prometheus** metrics, structured logging, and a live **SSE** dashboard fed from the engine.
- **ContextOS: Learned Context Selection Policy that Beats SOTA Token Compression** | GitHub
 - Trained a **GradientBoosting selection policy** (13 signals, 777 params) on HotpotQA + MuSiQe with **GroupKFold CV**.
 - Outperformed **LLMLingua-2 (SOTA)** at every reduction level — **+5.5 to +17.4 F1 points** across 55–85% compression.
 - Built a **FastAPI + Qdrant** RAG backend with hybrid retrieval, ROI cross-encoder scoring, and sub-ms learned inference.
 - Achieved **56.8% mean token reduction** across 10k scenarios (92.7M tokens saved) with 0.863 gold-fact recall and 0.92 answer similarity.
- **Multi-Tenant AI Console: Sandboxed Code Execution & Universal Connector System** | GitHub
 - Built a multi-tenant console executing untrusted user code in isolated **E2B Firecracker** microVMs on Hono + Supabase.
 - Engineered an agentic workflow engine with LLM intent routing, streaming live execution via **AsyncGenerator SSE**.
 - Built a universal connector discovery system — LLM infers API auth type, encrypts credentials at rest with **AES-256**.
 - Implemented per-user model quotas, **RBAC** middleware, token-bucket rate limiting across **8 concurrent sandboxes**.
- **Open Source Contribution** | GitHub
 - Fixed Gantt tooltip reporting incorrect **End Date** on queued/scheduled task segments in **Apache Airflow (38k+ stars)** — carried `end_when` across all segments of a try, mirroring the `start_when` fix from #68176. tests added in `Gantt/utils.test.ts`.

SKILLS

- **Languages:** C++, Go, Python, TypeScript, JavaScript, Swift, SQL
- **Databases:** PostgreSQL, MongoDB, Supabase (RLS + Migrations), Redis
- **Backend & AI/ML:** REST APIs, WebSockets, System Design, RAG Pipelines, LLM Integration, SSE Streaming, Agentic AI
- **Systems & Infra:** Linux (Ubuntu Server), Docker, Prometheus, GCP, Distributed Systems, Concurrency, Caching, CI/CD
- **Tools:** Git, Hono, Next.js, Convex, FastAPI, Qdrant
- **Problem Solving:** Solved **700+ DSA problems** on LeetCode (Rating: **1627**) | **95 Hard** problems solved.

ACHIEVEMENTS

- **Amazon ML School '24** — Selected for Amazon's ML training program; cohort drawn from top engineering institutions.
- **IQC (International Quant Championship)** - Qualified for Round 2 and achieved Gold level researcher status.
- **Technical Writing (Medium)** — Published ML research blog on context optimization and MCP discovery engine.