

Aditya Kumar Singh

Bengaluru | adityakumarsingh710@gmail.com | +91-7355553403 | [aditya-singh7101](#) | [AdityaSingh-7](#)

Education

Manipal Institute of Technology

Sep 2022 – Jul 2026

B.Tech in Computer Science and Engineering (Artificial Intelligence) | CGPA: 8.03/10

The Millennium School

Graduated 2021

All India Senior School Certificate Examination (AISSCE), CBSE

Experience

Software Engineering Intern | Eli Lilly and Company

Jan 2026 – Present

- Built and shipped full-stack features (Next.js, React, TypeScript, PostgreSQL, Redis) across an enterprise assistant platform serving 200 internal employees to create and share domain-specific AI assistants.
- Integrated an external knowledge source (Veeva Vault) via the Model Context Protocol with SSE streaming, and built admin governance tooling for model whitelisting and per-assistant access control.
- Resolved the majority of critical/high frontend security vulnerabilities through dependency auditing and npm override strategies.

Research Intern | International Centre for Applied Sciences (ICAS)

Jun 2025 – Jul 2025

- Built a hierarchical multi-head EfficientNet-B0 classification pipeline on DERM12345 (12,345 dermoscopic images, 40 subclasses), achieving 98.32% binary accuracy and 0.994 AUC — establishing early benchmarks on a dataset published in Scientific Data (Nature, 2024) with no prior classification baselines.
- Conducted a 6-model systematic study of loss functions (Dice, Tversky, Focal Tversky) on a 3D U-Net for BraTS 2020 brain tumor segmentation, achieving Dice scores of 0.77 (Tumor Core) and 0.75 (Enhancing Tumor) on 64³ voxel patches.

Software Engineer Intern | Recursive Zero

Jan 2025 – Feb 2025

- Migrated multiple client websites from vanilla HTML/JS to Astro.js, cutting JavaScript bundle size by ~50% and improving Core Web Vitals and SEO across all projects.

Projects

HNSW Vector Search Engine | Python, NumPy, FastAPI

[GitHub](#)

- Implemented HNSW approximate nearest-neighbour search from the original paper (Malkov & Yashunin, 2016); recall curve matches hnswlib/FAISS (95.6% recall@10), with the ~70× QPS gap attributable to pure Python vs. their SIMD-optimized C++.
- Built a reproducible benchmark harness measuring recall-vs-throughput on SIFT-1M set against two industrial C++ libraries, verifying correctness against an exact brute-force baseline.
- Wrapped the index in a FastAPI service with save/load persistence and a semantic-search demo.

Distributed Message Broker (Kafka-style) | Go

[GitHub](#)

- Built an append-only segmented log with per-write fsync durability, verified by crash-recovery tests (kill -9 mid-write → zero data loss on restart).
- Designed a custom binary TCP protocol with length-prefixed framing, serving concurrent clients via a goroutine-per-connection architecture.
- Implemented Kafka's StickyAssignor rebalancing algorithm and heartbeat-based failure detection for consumer group coordination, verified by unit tests.

Self-Healing RAG | Python, FastAPI, Pinecone, Llama 3

[GitHub](#)

- Built a corrective RAG pipeline that validates its own answers via LLM-as-judge and automatically retries with alternative retrieval strategies when confidence drops below threshold.
- Designed graceful degradation: surfaces partial answers with identified knowledge gaps rather than hallucinating when all strategies fail.
- Benchmarked standard vs. self-healing pipelines, characterizing when healing helps — it triggers on low-confidence queries at a ~3× token cost, improving hard cases while adding overhead where the baseline already succeeds.

Technical Skills

Languages: Python, Go, TypeScript, JavaScript, Java, SQL

Frameworks & Tools: Next.js, React, FastAPI, Node.js, Express.js, Tailwind CSS, Astro.js

AI/ML: PyTorch, TensorFlow, scikit-learn, Transformers

Databases & Infra: PostgreSQL, Redis, MongoDB, Pinecone, Docker, Kubernetes, Git