

KARAN SINGH

New Delhi, India, 110078

☎ +91-8512051222 | ✉ karansingh25822@gmail.com | 🔗 www.linkedin.com/in/karansinghdev | 🌐 https://github.com/KaranSinghDev

Experience

IFIC Valencia - LHCb Experiment at CERN

October 2025 – Present

Research Software Engineer

Remote

- Implemented an energy-profiling module in the ACTS C++ framework, discovering the iterative Kalman Fitter to be **95% more efficient** than the Global Chi-Square Fitter (**7.9 kJ vs 15.4 kJ over 1M events**).
- Analyzed the Combinatorial Kalman Filter and uncovered a track-fitting inversion driving up to **15% energy variation**, offering actionable insights for High Low LHC reconstruction optimization.
- Integrated a stable ACTS build from source, addressing **C++20 and Boost conflicts** to create the reproducible environment used for a 1-million-event benchmark.

Google Summer Of Code - CERN-HSF

June 2025 – September 2025

Contributor

Remote

- Architected a Python framework to benchmark classical (PyTorch) and quantum (PennyLane) algorithms, automating a **4-stage workflow** that processes **25 million data points** from raw physics simulations into a final analysis report.
- Assembled a multi-dimensional analysis engine evaluating **8+ sustainability metrics**, including hardware-independent energy models based on **1,100** computational operations per inference (classical) and **19** gates per inference (quantum).
- Delivered a robust, portable solution with cross-platform compatibility (Windows/Linux) and dual-hardware support (GPU/CPU), implementing graceful fallbacks to **reduce setup failures by an estimated 95%**.

Nihin Media K.K.

January 2025 – July 2025

AI Intern

Remote

- Developed an automated evaluation pipeline to rigorously test fine-tuned large language models (e.g., DeepSeek, Gemini), establishing a quality baseline that drove a **92%** improvement in response consistency.
- Built the core information retrieval backend for an interactive AI persona, designed to serve **5,000** users daily, implementing a custom hash-based index that minimized **P99** query latency by **15%** across its **10+ GB** database.
- Reduced inference latency by **15%** by fine-tuning language models (e.g., Phi-3, Qwen) while maintaining accuracy.

Projects

Time Series Database Engine | C++ · Python · FastAPI · Docker

August 2025 – September 2025

- Engineered a **low-level C++ storage engine** implementing custom data compression algorithms (Delta-of-Delta, XOR) to **maximize storage efficiency by 50%** for high-entropy time-series data (**average 8.2 bytes/point**).
- Optimized query performance by architecting a time-sharded layout to minimize disk I/O, achieving **P99 latencies of 1.3 ms** (hot-cache) and **sub-16 ms** for demanding, long-range cold-disk queries spanning **24 hours** of data.

Distributed Fault-Tolerant Cache System | Python · Asyncio · gRPC · Docker

March 2025 – April 2025

- Designed a scalable, clustered caching microservice using Python's asyncio for non-blocking I/O and gRPC as the RPC framework, sustaining a peak throughput of **17,000 ops/sec** from 50 concurrent clients on a **3-node cluster**.
- Ensured fault tolerance by implementing consistent hashing and N-way replication, validated with tests to guarantee **zero data loss** during simulated single-node failure scenarios while preserving **P99 read latency under 6 ms**.

Technical Skills

Programming Languages: C++, Python, SQL

Backend & Databases: Docker, Kubernetes, Kafka, Redis, PostgreSQL, MySQL, FastAPI, Elasticsearch

Cloud & DevOps: AWS (EC2, S3, Lambda, IAM), Azure (Virtual Machines, Blob Storage), CI/CD, GitHub Actions

AI & Machine Learning: PyTorch, Scikit-learn, Pandas, NumPy, Transformers, LangChain, CUDA

Publications & Leadership

Technical Research & Achievements

December 2022 - September 2025

- Co-authored**, Sustainability studies of big data processing in real time for HEP, poster at **ACAT 2025**, Sep 2025
- Co-authored**, Improvements on QAOA for Particle Trajectories at LHCb, poster at **ACAT 2025**, Sep 2025
- Managed my college's AI/ML club mentoring **100+ students** and **organizing 15+ workshops/events**, alongside securing **3rd place** at SRM Builds 5.0 & **6th place** at Intel's Advent of Multimodal AI Hackathon 2024 (100+ teams).

Education

SRM University

September 2021 – June 2025

Bachelor of Technology in Computer Science and Engineering (CGPA: **8.1**)

Delhi-NCR, Sonapat