Samarth Saxena

+91 8643005535 | samarthsaxena.career@gmail.com | [Linkedin](#) | [GitHub](#) | [Portfolio](#)

Education

Rajiv Gandhi Proudyogiki Vishwavidyalaya (R.G.P.V)   **Indore, India**
B.Tech. in Information Technology   *Nov 2022 – July 2026 (Expected)*

- **Coursework**: Data Structures & Algorithms, Computer Systems Organization, Design & Analysis of Algorithms, Object-Oriented Programming, Operating Systems, Database Management Systems, Software Engineering, Artificial Intelligence

Skills

Programming Languages: C, C++, Python, JavaScript/TypeScript, SQL, Bash
Frameworks & Libraries: React.js, Next.js, FastAPI, Node.js (Express), PyTorch, Scikit-Learn, Pandas, Socket.io
DevOps & Cloud Tools: Docker, Wireshark, Linux Networking (TCP/IP), Git, GitHub Actions, Vercel, Google Cloud Platform
Databases: Pinecone, ChromaDB (Vector DBs), PostgreSQL, MySQL, MongoDB, Firebase, Prisma

Experience

Engineering Intern   Indore, India
Clear-Trail Technologies   *Jan 2026 – Present*

- Architecting a **privacy-centric remote resource sharing framework** as a zero-trust alternative to legacy tools, utilizing **Containerized Resource Isolation** to enable selective data streaming without full-system exposure.
- Analyzing encrypted network traffic patterns using **Wireshark** to extract **Moore's Discriminators**, focusing on identifying behavioral heuristics within secure streams for enhanced security protocols.
- Engineering a specialized **ML training pipeline** using captured network packet data (PCAP) to automate the classification of encrypted traffic based on **ENISA** security standards.
- Developing high-conversion ad-delivery architectures by evaluating **Network-Level DNS Filtering (Pi-hole)** and hybrid-cloud filtering to optimize the balance between user privacy and advertisement relevance.

Projects

AI Microservices Engine | *FastAPI, Docker, Python, ChromaDB, Google GenAI* :

- Built a containerized **FastAPI** inference service with Docker, isolating vector retrieval, embedding, and query routing into separate execution paths for modular deployment.
- Reduced inference memory footprint by **91% (600MB → 50MB)** by offloading vector operations to external embedding endpoints and introducing lazy loading for retrieval components.
- Implemented a semantic search pipeline with **ChromaDB**, including chunk ingestion, batched embedding, and indexed retrieval, achieving **sub-2-second query latency** on test datasets.
- Designed a secure **BYOK authentication layer** with runtime key injection and zero server-side persistence, preventing credential leakage across container boundaries.

NoteWave — AI Brain Platform | *Next.js, TypeScript, Pinecone, Groq (Llama 3.3-70B), ElevenLabs* :

- Built a **RAG pipeline** for PDF ingestion and query answering with document-scoped retrieval using Pinecone vector indexing.
- Implemented a **rate-limit–resilient ingestion workflow** with 1k-token semantic chunking, batched embeddings, and exponential backoff to stabilize processing on constrained APIs.
- Designed a **document-level isolation strategy** using metadata-filtered retrieval to prevent cross-document context leakage during generation.
- Added **streamed token delivery** from the inference endpoint to the client, enabling incremental response rendering without blocking the retrieval pipeline.

Achievements & Certifications

- **Cerebras GLM-4.7 Hackathon**: Winner of the global GLM-4.7 coding challenge hosted by Cerebras.
- **Top 40 Finalist, National Level**: Kriyeta 4.0 Hackathon (Selected among 250+ teams nationwide)
- **Certifications**: NPTEL Computer Graphics (Govt. of India), CS50x Harvard University
- **Senior Diploma in Tabla**: Prayag Sangeet Samiti (Awarded after 5+ years of formal training)