

Manit Kumar

8800-688-629 | manit17102006@gmail.com | linkedin.com/manit-kumar | github.com/manit989 | codolio.com/anonusetux

EXPERIENCE

Indian Army

June 2026 – August 2026

Intern

Delhi Cantt.

- Engineered a fully local RAG system using **FAISS** and **HuggingFace** embeddings (all-MiniLM-L6-v2), cutting diagnostic manual lookup time by 90% with 100% offline capability.
- Architected an AI diagnostic pipeline using **LangChain** to bridge edge-deployed predictive maintenance models (XGBoost/Random Forest) with a generative AI reporting system.
- Integrated LLMs (**Gemini/Ollama**) to translate complex mechanical fault codes into actionable, plain-English Situation Reports (SITREPs) for tank operators.

EDUCATION

Guru Gobind Singh Indraprastha University

Delhi, India

Bachelor of Technology in Computer Science (CGPA: 9.0 / 10.0)

Aug. 2024 – Present

PROJECTS

FrameIQ: Multimodal Semantic Video Search Engine | *FastAPI, LanceDB, Groq LPU, Gemini Flash, FFmpeg* Sep. 2026

- Engineered a cross-modal Video-RAG search engine using **FastAPI** and **LanceDB** for moment-level semantic retrieval of short-form videos without captions or tags.
- Architected a dual-stream ingestion pipeline using **Groq Whisper API** for audio tokenization and **Gemini Flash** for visual scene decomposition, enforcing strict **Pydantic** schemas to prevent JSON hallucinations.
- Implemented zero-copy columnar indexing with **LanceDB** and Gemini embeddings for sub-millisecond cosine similarity searches, returning precise timestamps and on-the-fly **FFmpeg** thumbnails.

ScaleGuard: Distributed Auth & Rate-Limiting API | *FastAPI, Redis, PostgreSQL, Nginx, Docker* Feb. 2026

- Developed a distributed backend system using **FastAPI** and **Nginx**, scaling across multiple replicas to handle high-concurrency traffic via load balancing.
- Engineered an atomic **Token Bucket Rate Limiter** using **Redis Lua scripting**, ensuring thread-safe operations and preventing race conditions across distributed nodes.
- Implemented stateless **JWT** authentication with **Alembic** migrations; orchestrated a 4-service stack via **Docker Compose** and validated rate limits using **Locust**.

ACHIEVEMENTS & CERTIFICATIONS

1st Runner-Up – AzinHack 2026 | *USAR, GGSIPU* | [GitHub](#) | [Live Demo](#)

Oct. 2026

- Secured **2nd place** out of **250+ teams** in a 24-hour AI hackathon by building **Tinyfah**, an autonomous web vulnerability analysis agent using **LangGraph** and the **TinyFish API** to audit web apps against **345 OWASP ASVS v5.0.0** checks.

Google Cloud Skills Boost Profile | *Diamond League (3337+ points)*

- Completed hands-on cloud infrastructure badges including **Monitoring in Google Cloud**, **Cloud Speech API**, **App Engine**, and **Cloud Run Functions**.

COMPETITIVE PROGRAMMING & DSA

Algorithmic Problem Solving | *LeetCode, Codeforces, AtCoder*

- **LeetCode**: Solved **486** DSA problems (308 Medium, 45 Hard) with a Max Rating of **1815**; competed in **56** total contests across platforms (Max Rating **927** on AtCoder).

TECHNICAL SKILLS

Languages: Python, SQL, C/C++, Java, JavaScript, Lua, R

Backend Frameworks: FastAPI, Flask, Pydantic, SQLAlchemy, Alembic

Data Science/ML: LangChain, LangGraph, HuggingFace, NumPy, Pandas, Matplotlib

Database & Infrastructure: PostgreSQL, Redis, Docker, Nginx, Git

Developer Tools: Neovim, Linux, Google Cloud Platform, Postman