

ANSHUMAN SINGH

anshu.2005.12@gmail.com — +91-9125506309 — GitHub — LinkedIn

Summary

AI/ML engineer with hands-on experience building **LLM-based, GraphRAG, and real-time ML systems**. Developed production-grade pipelines processing **100+ documents, 10K+ records**, and supporting low-latency inference. Strong focus on retrieval optimization, hallucination reduction, and scalable deployment.

Education

PDPM IITDM Jabalpur, India

Bachelor of Technology – Electronics & Communication Engineering

2024 – 2028

Boys' High School, Prayagraj

Senior Secondary (Class XII) – 89%

Secondary (Class X) – 90%

Experience

Kalp Intelligence

May 5, 2026 – July 5, 2026

SDE-I Intern

- Contributed to the development of **Gram**, a food ordering platform, across the mobile app and website
- Worked on real-world product features and collaborated with the team to deliver quality work within deadlines
- Gained hands-on experience in software development, problem solving, and product collaboration

Skills Summary

- Languages:** Python, JavaScript, Rust
- Frameworks:** FastAPI, Django, React, LangChain
- AI / ML:** LLMs, RAG, Multimodal RAG, FAISS, Embeddings, XGBoost, Isolation Forest
- Domains:** NLP, Anomaly Detection, Information Retrieval, Generative AI
- Tools:** Git, Postman, REST APIs

Projects

Nsure-Graph-AI – Graph-based RAG Insurance Intelligence Platform

Live — GitHub

- Built a GraphRAG system processing **100+ insurance policy documents** and enabling multi-hop reasoning across **1K+ knowledge graph nodes**
- Reduced LLM hallucinations by **30–40%** through graph-based contextual retrieval
- Improved answer relevance by **1.5×** compared to vanilla RAG using embedding + graph traversal
- Deployed a full-stack application with **sub-2s inference latency** on cloud infrastructure

Indie-VOICE – Multilingual Text-to-Speech System

GitHub

- Developed a multilingual TTS system supporting **22+ Indian languages** for audiobook and podcast use cases
- Improved speech consistency by **30%** using chunked inference and deterministic decoding
- Reduced audio generation latency by **25%** through optimized WAV synthesis

Fraud_trnx_Flagger – Real-Time Fraud Detection

GitHub

- Built a fraud detection pipeline analyzing **10K+ synthetic transactions** using Isolation Forest and XGBoost
- Improved anomaly detection precision by **20%** via feature engineering and model tuning
- Deployed real-time inference API achieving **150ms response time**

Achievements

CYHI Hackathon – Winner

- Built an AI-based team recommendation engine used by **50+ participants**, reducing manual team formation time by **60%**
- Automated skill-based team formation, reducing manual effort

Certifications

- JavaScript (Basic) – HackerRank
- Postman API Fundamentals
- Master RAG: Retrieval-Augmented Generation (Advanced)