

Anupam Verma

AI Engineer



📍 Mumbai, Maharashtra, India

☎ 9445864375

✉ anupam215769@gmail.com

🔗 LeetCode

🔗 Github

🔗 LinkedIn

🔗 Portfolio Website

🔗 Tableau Public

SKILLS

Generative AI

RAG, Fine-tuning, LangChain, LangGraph, Transformers, LLMs

Machine Learning

TensorFlow, PyTorch, Scikit-learn

NLP & Computer Vision

Hugging Face, NLTK, spaCy, Gensim, OpenCV

Data Analysis & Visualization

Tableau, Power BI

Programming

Python, C++, C

Frontend & Backend

Chainlit, Streamlit, FastAPI, Flask

Databases / Graph Databases

MySQL, PostgreSQL, MongoDB, Neo4j, ArangoDB

VectorDBs / Vectorstores

ChromaDB, PGVector, Qdrant, Milvus, Weaviate

Cloud

AWS, Azure, Docker, Git

EDUCATION

Post Graduate Diploma in Data Science, Symbiosis Centre for Distance Learning

Jul 2023 – Jun 2025

Passed with 83.2%

B.Tech in CSE, Vel Tech University

Jul 2019 – Jun 2023

Passed with 9.08 CGPA

EXPERIENCE

Financial Software and Systems (FSS), AI Engineer

Sep 2025 – Present

- Built a **RAG pipeline** for **FSS BLAZE™ Cosmos** on ACS, Recon and Chargeback source code. This helped **developers and QA** with test generation and bug detection, cutting debugging time by 50%.
- Created a **ReAct agent** using **PGVector** and **ArangoDB**. It handles user queries with reasoning and action, boosting query **accuracy by 60%**.
- Used **metadata filtering, reranking, and similarity search** to retrieve function definitions. This improved **retrieval precision by 70%**.
- Fetched forward and backward **function dependencies** with dynamic AQL queries and generated **dependency graphs** using **Cytoscape.js** of the **methods and classes** for the same.
- Served and optimized **Qwen3-Coder-30B-A3B-Instruct** via **vLLM**, integrating **LiteLLM** for monitoring, fallback handling, and guardrails (PII + **shield-gemma-2b** for safety), achieving **24x throughput** and **~65% lower latency**.
- Implemented **per-user logging and auditability** using **Merlin Forge** (VS Code extension) and added **Azure AD authentication**, enhancing security, observability, and developer workflow efficiency.

Comcast, Engineer 1 - Software Development & Engineering

Jan 2023 – Aug 2025

Generative AI

- Engineered an **Agentic Retrieval-Augmented Generation (RAG)** system with **LangChain** and **LangGraph**, querying millions of code lines and cutting manual reviews by **40%** through automated suggestions, bug detection, test generation, and more efficient documentation of new RDK contributions using **Azure OpenAI gpt-4.1** and **o3 models**.
- Developed an interactive **chatbot** with **Streamlit** as the front-end and **FastAPI** as the **RESTful API** back-end, reducing latency by **60%** while creating **custom vector embeddings** through **microservices**.
- Integrated multiple agents with a **ChromaDB** vector store, combining **Similarity and Full-Text Search** to improve retrieval accuracy by **95%** and enable more precise queries of RDK documentation and codebase.
- Architected persistent **user session management** and chat history storage using **MongoDB**, enabling seamless retrieval of **past interactions** and supporting **context-aware** responses, which enhanced user satisfaction scores by **35%** and increased user engagement by **50%**.
- Integrated a **Neo4j Knowledge Graph** to map and analyze RDK **component-level dependencies**, accelerating dependency mapping by **40%** and enabling precise identification of **inter-component relationships** and creation of architecture diagrams.
- Leveraged **Selenium-based web scraping** to automate extraction of key contribution details from **Gerrit patches**, achieving **90% accuracy**, and integrated the results into **Jira via API**, increasing cross-team visibility and tracking efficiency by **70%** for over **500 code changes**.
- Deployed the end-to-end application on an **AWS EC2** instance using **Docker** containers, automating **CI/CD workflows** on **GitHub Actions** and achieving a **65% reduction** in deployment time while ensuring scalable, reliable operation with **99.9% uptime**, which facilitated smoother production.

Tableau

- Developed **50+** interactive **RDK dashboards** in **Tableau**, including clone, code, and contribution **metrics portals**, boosting insights by **75%** and significantly improving data accessibility and efficiency for **stakeholders**.
- Leveraged advanced Tableau techniques to create visually appealing **charts, graphs, and maps**, representing **complex data** more effectively and enhancing **stakeholder** understanding by **50%**.
- Implemented innovative **data integration** with **Tableau Prep Builder** for **ETL** process like cleansing and **transformation**, **cutting data cleaning time by 85%** and improving loading efficiency into **MySQL** while maintaining a streamlined workflow.

CERTIFICATES

- Fundamentals of Deep Learning by NVIDIA
- AWS Academy Machine Learning Foundations