

# Anurag Kumar

Associate Data Scientist

7970940623 | [anuragabcr@gmail.com](mailto:anuragabcr@gmail.com) | [linkedin.com/in/anuragabcr/](https://linkedin.com/in/anuragabcr/) | [github.com/anuragabcr](https://github.com/anuragabcr)

## SUMMARY

**Data Scientist (Generative AI & NLP)** with 2+ years of experience designing and implementing **LLM-powered AI solutions** for enterprise use cases. Strong hands-on expertise in **Generative AI, NLP, RAG architectures, vector databases, and MLOps**, with experience deploying scalable AI pipelines on **GCP and Azure**. Proven ability to translate business requirements into **production-ready AI systems**, ensuring **data quality, security, and responsible AI practices**.

## TECHNICAL SKILLS

**Generative AI & NLP:** LLMs (GPT, LLaMA), RAG Pipelines, LangChain, Hugging Face Transformers, Prompt Engineering, Embeddings, Semantic Search, FAISS, OpenAI API.

**MLOps & Cloud:** GCP (BigQuery, Dataproc), Azure (SQL), Docker, Airflow (Orchestration), CI/CD concepts, Git.

**Data Engineering:** PySpark, SQL, BigQuery Partitioning, Data Pipelines, Redis (Vector Caching), Data Cleansing & Curation.

**Core Machine Learning:** Python, PyTorch, Scikit-learn, Pandas, NumPy, Text Classification.

## EXPERIENCE

### Software Developer I

Jul 2024 – Sep 2025

*iBind Systems*

*Bangalore*

- Designed and implemented end-to-end data pipelines using **Airflow**, aligning with **MLOps practices** to support automated monitoring and model retraining.
- Built Python-based **RAG pipeline** workflows for unstructured data, ensuring high-quality **output** from Large Language Models (LLMs).
- Optimized **BigQuery** datasets using partitioning to support low-latency analytics, directly improving the performance of model inference workloads.
- Implemented guardrails for data validation and quality checks to ensure **reliable AI inputs**, improving accuracy and consistency across analytical systems.
- Collaborated with cross-functional teams to translate business requirements into **data-driven and AI-ready solutions**.

### Full Stack Developer

Oct 2023 – Jun 2024

*HikeCommerce Business Services*

*Noida*

- Collaborated with product teams to define data schemas aligned with business KPIs, curating datasets for machine learning experimentation.
- Developed backend APIs in **Python** to serve analytical data and support ML-driven dashboard requirements.
- Worked closely with product and analytics teams to define data schemas aligned with **business KPIs and AI use cases**.
- Performed extensive data cleaning and normalization using SQL and Python to ensure consistency across business logic and reporting layers.

## PROJECTS

### Enterprise Document Q&A using LLMs | *Python, LangChain, OpenAI API, FAISS, Redis*

- Developed a **Retrieval-Augmented Generation (RAG)** architecture to solve enterprise knowledge retrieval challenges.
- Implemented **document chunking strategies** and generated embeddings to populate a **FAISS vector store**, enabling semantic similarity search.
- Designed context-aware **prompt templates** to ground LLM responses in enterprise data, reducing hallucinations and ensuring relevance.
- Established evaluation metrics based on semantic similarity to assess output quality and coherence.

### SAP Query Chatbot | *Python, OpenAI, LangChain, LangGraph, SAP, FAISS, Postman*

- Developed an AgenticAI chatbot integrated with SAP systems

- Created tools to check user roles, validate SoD conflict, finding and raising Applications request on SAP system and checking request status.
- Created SQL agent to query SAP database for application roles and their respective T-codes.

#### **NLP-based Text Classification & Summarization** | *Python, Pandas, NumPy, Hugging Face Transformers*

- Fine-tuned pre-trained **Transformer models** (BERT-based) for domain-specific text classification tasks.
- Implemented preprocessing, tokenization, and evaluation workflows in Python.
- Deployed the inference pipeline on **GCP**, automating data ingestion and preprocessing using Python workflows.

#### EDUCATION

---

##### **CMR University, Bangalore**

*Master of Computer Application (MCA)*

Bangalore, India

*2021 – 2023*

##### **Sardar Patel University, Gujarat**

*Bachelor of Computer Application (BCA)*

Gujarat, India

*2018 – 2021*

#### CERTIFICATIONS

---

##### **Google Cloud Engineer Learning Path**

*Google Cloud Skills Boost*

##### **IBM Data Engineering Professional Certificate**

*Coursera*