

Shreya Bhat

Data Scientist

www.linkedin.com/in/shreyabhat11 | game-portfolio-five-tau.vercel.app | github.com/Shreyabhat11 | candidates.interviews.club/u/shreya-bhat

SUMMARY

Data Scientist with experience developing GenAI, RAG, and anomaly-detection systems in manufacturing environments. Holds a B.E. degree and specialized post-graduate training, with expertise in processing time-series sensor data and document retrieval pipelines.

EXPERIENCE

Trainee – Manufacturing & Formulation

CIPLA Ltd

Jul 2024 – Jun 2025

- Developed GenAI, RAG, and sensor anomaly-detection systems to support manufacturing plant operations.
- Built RAG document processing pipelines over 1,000+ logs and SOPs.
- Processed time-series data from 1,400+ sensors across 15 SCADA and 47 BMS systems.
- Executed LLM evaluations using RAGAS to improve retrieval model performance.

Trainee – Manufacturing & Formulation

CIPLA Ltd

Jul 2024 – Jun 2025

- Developed GenAI, RAG, and sensor anomaly-detection systems to support manufacturing plant operations.
- Built RAG document processing pipelines over 1,000+ logs and SOPs.
- Processed time-series data from 1,400+ sensors across 15 SCADA and 47 BMS systems.
- Executed LLM evaluations using RAGAS to improve retrieval model performance.

EDUCATION

Bangalore Institute of Technology · B.E. · 2020 – 2024

Great Learning · PGP · 2025 – 2026

SKILLS

Languages and Frameworks: C, Java, JavaScript, Jupyter Notebook, Python, SQL, TypeScript, CrewAI, CTGAN, FastAPI, Flask, Gradient Boosting, Hugging Face Transformers, Keras, LangChain, LangGraph, LightGBM, Matplotlib, Next.js, NumPy, OpenCV, Pandas, PyTorch, Random Forest, Scikit-learn, SDV, Sentence Transformers, SMOTE, spaCy, Streamlit, TensorFlow, TF-IDF, Whisper, XGBoost

Databases: FAISS, MySQL, pgvector, PostgreSQL, Qdrant, SQLite, Vector Databases

Tools: Agentic AI, CNN, Deepgram, Docker, DVC, Evidently AI, Git, GitHub, Google Gemini API, Google STT, Grafana, JiWER, JWT, LLaVA, LLM systems, MLflow, Power BI, Prefect, Qwen-VL, RAG, RAG pipelines, RAGAS, ReAct, REST APIs, Transformer, WebSockets, Zerodha Kite Connect

PROJECTS

Multimodal Financial Document Understanding System

Built a multimodal document-understanding pipeline using open-source vision-language models to extract text, tables, charts, signatures, and layout from multi-page financial documents. Extracted structural layout, text, tables, and signatures using LLaVA and Qwen-VL.

Orchestrated complex extraction workflows utilizing LangGraph, CrewAI, and ReAct frameworks.

Multimodal LLMs, LLaVA, Qwen-VL, LangGraph, CrewAI, ReAct

AI Hallucination Detection System

Built an LLM validation pipeline performing claim extraction, web evidence retrieval, and semantic similarity scoring to detect hallucinations. Implemented claim extraction and semantic similarity scoring pipelines to detect model hallucinations. Integrated Google Gemini API, FAISS vector index, FastAPI, and Streamlit for live validation.

LLM Evaluation, RAG, FAISS, FastAPI, Google Gemini API, Streamlit

AskQL - Natural Language to SQL Query System

Built an AI-powered natural language-to-SQL system converting plain English queries into structured SQL commands using rule-based and NLP-driven pipelines. Converted natural language queries into structured SQL using classical NLP and LLM techniques. Deployed backend query execution engines using FastAPI and React.

NLP, SQL, LLM, FastAPI, React

Additional projects: Multimodal Financial Document Intelligence, AI Hallucination Detector, Industrial RAG + Anomaly Intelligence, GPT from Scratch - PyTorch, System Design RAG Assistant, AI MOM - Meeting Minutes Generator, AskQL - Natural Language to SQL, ASR Shootout - Indian Conversational Speech, Fraud Detection MLOps Platform, Fake News Trend Drift Detector, Synthetic Data Generation Platform, AI-Powered Investment Portfolio Platform, AI Symptom Checker, Health Insurance Fraud Detection System, Structural Defect Analyzer, Credit Card Fraud Detection System, AI Resume Skill Matcher, Data Quality & Pipeline Monitoring Dashboard, Synthetic Data Generation Platform, AI-Powered Investment Portfolio Platform, AI Symptom Checker, Health Insurance Fraud Detection System, Multimodal Financial Document Intelligence, AI Hallucination Detector, Industrial RAG + Anomaly Intelligence, GPT from Scratch - PyTorch, System Design RAG Assistant, AI MOM - Meeting Minutes Generator, AskQL - Natural Language to SQL, ASR Shootout - Indian Conversational Speech, Fraud Detection MLOps Platform, Fake News Trend Drift Detector, Structural Defect Analyzer, Credit Card Fraud Detection System, AI Resume Skill Matcher, Data Quality & Pipeline Monitoring Dashboard

AWARDS

Performer of the Month