

Gaurav Raul

Mumbai | Maharashtra

Email: raulgaurav96@gmail.com

Mob: +91 9890471562

CAREER OBJECTIVE

Lead Data Scientist & AI Engineer with 7.5+ years of experience deploying production-grade, scalable AI systems across FinTech, Telecom, and Enterprise Automation. Specialized in end-to-end GenAI and Agentic architectures, advanced RAG, contextual retrieval, and LLM evaluation systems using LangChain, LangGraph, and Vector Databases. Expert in transformer models, fine-tuning, and cloud-native AI infrastructure, with a proven track record of engineering low-latency inference pipelines and production APIs. Integrates robust MLOps, CI/CD, and observability practices to deliver reliable, cost-optimized intelligent automation that drives measurable enterprise business impact.

Professional Experience:

Organization: Fractal Analytics (Permanent Role)

Sep 2024 – Present

Role: Senior Data Scientist

- **Enterprise AI Framework Engineering:** Architected **Assist Framework** and **FlowAssist** to standardize multi-agent system development; engineered a **Text-to-DAG compiler** within FlowAssist that transforms natural language prompts and code into executable visual workflows with editable node-level orchestration.
- **Multi-Agent Platforms & Generative UI:** Designed **Sales Assist** (UHNI wealth management) and **Ops Assist** (back-office reconciliation) enterprise accelerators to drive client acquisition. Integrated autonomous planner, simulation, profiling, and portfolio agents utilizing a **Generative UI framework** to dynamically adapt frontend components based on conversational intent.
- **Financial Operations & Automated RCA:** Built autonomous reconciliation engines leveraging specialized **Root Cause Analysis (RCA)** and **Next Best Action (NBA)** agents to identify transaction anomalies, automate complex diagnostics, and generate actionable remediation pathways.
- **Hybrid Infrastructure & SLM Optimization:** Engineered dual-layer architectures combining distributed batch-processing pipelines with low-latency, real-time inference services. Fine-tuned **Phi-3** and optimized **Gemini 2.0 Flash** pipelines using context compression, elevating billing accuracy past **92%** and answerability to **96%** while saving millions in token overhead.

Organization: Fractal Analytics (via Meraki Talentworks Pvt. Ltd.)

Feb 2024 – Sep 2024 Role:

Senior Data Scientist

- **Retail GenAI Copilot:** Developed a production-grade retail AI copilot leveraging Retrieval-Augmented Generation (RAG) and Gemini 1.5 to enable real-time conversational summarization, intent detection, and contextual product recommendation workflows for retail support agents.

Organization: Cognext.ai

May 2021 – Feb 2024

Role: Data Scientist

- **Financial AI & Domain NLP:** Architected end-to-end sentiment intelligence platforms using Python, PyTorch, and Flask; developed a proprietary Financial Language Model (CFSAM) leveraging fine-tuned transformer architectures.
- **GenAI & Advanced Retrieval:** Designed NL-to-SQL systems and multi-document intelligence workflows utilizing LangChain, ChromaDB, and open-source/proprietary LLMs for contextual financial retrieval.
- **Document AI & Multilingual OCR:** Engineered high-throughput digitization pipelines using Detectron2, image processing, and fine-tuned Tesseract OCR to extract structured data from complex multilingual Arabic/financial documents.

Organization: Ipredictt Data Labs Pvt. Ltd.

January 2020 – April 2021

Role: Manager – Data & Technology

- **Predictive NLP Modeling:** Architected NLP lead-prediction and recommendation systems that tripled lead volume and improved quality by 50%; engineered Flask APIs deployed on AWS (EC2, Lambda, S3, RDS, IAM).
- **Full-Stack Architecture:** Built and launched Careerletics Genius and JobCoach from scratch using Django, Python, and SQL, owning end-to-end cloud infrastructure, deployment, and third-party integrations.
- **Data Operations & Automation:** Led technical product strategy for client campaigns and engineered analytics

dashboards and automation systems to improve operational workflows, reporting automation, and business visibility.

Organization: Margo Networks Pvt. Ltd.

April 2018- July 2019

Role: Software Developer

- Developed full-stack microservices, REST APIs, and enterprise dashboards using Java, Spring Boot, and Angular 7, including an early recommendation engine proof-of-concept using collaborative filtering.

Skill Sets

GenAI & Agentic Architectures	LangChain, LangGraph, LlamaIndex, Model Context Protocol (MCP), Agentic Workflows, Retrieval-Augmented Generation (RAG), Contextual Retrieval, LLM Fine-Tuning (PEFT, LoRA/QLoRA), LLM Evaluation Systems, OpenAI API, Gemini Enterprise, Llama 3
AI Infrastructure & Backend	Python, SQL, FastAPI, Flask, Django, Scalable Production APIs, Low-Latency Inference Pipelines, Model Serving, Async Processing, Distributed Systems
ML & Language Intelligence	PyTorch, Hugging Face (Model Training), Transformers, BERT, PySpark, Scikit-Learn, Recommendation Systems, NLP & Language Intelligence (NER, Text Summarization, Topic Modeling)
Cloud & MLOps Infrastructure	AWS (SageMaker, EC2, S3, RDS, Lambda, VPC, IAM), GCP, DigitalOcean, Docker, GitHub Actions (CI/CD), MLflow, Model Observability, CI/CD Automation
Vector Databases & Storage	FAISS, ChromaDB, PostgreSQL, Redis, Hybrid Semantic Retrieval

EDUCATION

- Atharva College of Engineering (Mumbai University)
Bachelor of Engineering in Computer Science & Engineering

Mumbai, Maharashtra
2013 - 2017

Web Links

- **LinkedIn:** <https://www.linkedin.com/in/gaurav-raul-09388810a/>
- **Github:** <https://github.com/gauravraul/>

