Ganesh Rajput

grt7045@gmail.com | +91 9023584367 | LinkedIn | GitHub

Surat, Gujarat, India
Open to Relocation / Remote roles

Professional Summary

Python engineer who ships both sides of an AI feature: the LLM layer (RAG, agents, structured document extraction, safety validation) and the backend that keeps it alive in production (FastAPI, Django REST Framework, PostgreSQL). Currently maintain three live client systems and have built LangChain, FAISS, and DSPy-based AI applications with source citations, Pydantic-validated outputs, and Pytest coverage. Comfortable tracing a bad model answer or a slow query back to its cause through logs, API responses, and data. B.E. in Computer Engineering, 9.46/10 CGPA.

Technical Skills

Languages	:	Python, SQL, Java, TypeScript
Generative AI	:	LLM APIs, Prompt Engineering, Prompt Chaining, Structured Outputs (JSON), Output Parsing, Guardrails & Output Validation
LLM Platforms	:	Google Gemini, Groq, OpenRouter, Hugging Face, Ollama (local LLMs)
RAG & Retrieval	:	LangChain, FAISS, ChromaDB, Embeddings, Semantic Search, Vector Search, Document Ingestion, Chunking, OCR
Agentic AI	:	AI Agents, Tool Use, Conversational Memory, Multi-step Orchestration, Recursive LLM Workflows, DSPy
Backend & APIs	:	FastAPI, Django REST Framework, Django, REST APIs, Pydantic, JWT, RBAC, Request Validation, OpenAPI/Swagger
Data & ML	:	Pandas, NumPy, Feature Engineering, Model Evaluation (Accuracy, Precision, Recall), PostgreSQL, MySQL, SQLite, Redis
Tools & Delivery	:	Pytest, Docker, Git, GitHub Actions, Linux, Bash, Postman

Experience

Python Developer – AI & Backend **Jul 2025 – Present**
BitCoding Solutions Pvt. Ltd. Surat, Gujarat

- Design and build **RAG pipelines** (ingestion, chunking, embeddings, FAISS/ChromaDB retrieval) and serve them through **FastAPI** REST APIs.
- Build **LangChain agents** for document extraction, summarization, and Q&A using prompt chaining, memory, and **structured JSON outputs validated with Pydantic**.
- Add **guardrails and tests**: emergency and off-topic filtering, post-generation validation, and Pytest suites for prompt and safety workflows.
- Work across Gemini, Groq, OpenRouter, and Hugging Face models; investigate wrong or unstable outputs and fix them at the prompt, retrieval, or validation layer.
- Build and maintain backend services for three live client applications (property auctions, warehouse management, e-commerce) using **Django REST Framework and FastAPI**; sole backend owner of the auction platform.
- Manage PostgreSQL schema changes, migrations, and ORM queries; trace defects across API responses, logs, and database state; apply JWT authentication and RBAC to production APIs.
- Write SQL utilities and database-backup scripts that support daily warehouse operations.

AI/ML Intern **Feb 2025 – Mar 2025**
Sparks To Ideas Ahmedabad, Gujarat

- Prepared datasets and engineered features with Pandas and NumPy for supervised learning tasks.
- Trained and compared models on accuracy, precision, and recall to judge which one actually performed better.

Projects

KnowledgeHub AI – Private Document Data Extraction *Author, Full-stack AI*
Python, FastAPI, PostgreSQL + pgvector, Ollama, RapidOCR, Next.js, TypeScript, Docker

- Built a multi-tenant platform that turns invoices into **reviewed, structured spreadsheet data** using **local LLMs (Ollama)**, so sensitive files never reach a cloud service; **hybrid extraction** combines rule-based readers, OCR for scans, and an LLM that fills the gaps field by field.
- Added deterministic checks (GST tax-ID check digits, totals, tax-rate sanity, duplicates) and a Next.js review screen with approval and Excel export, backed by a Postgres job queue, tenant isolation, and an audit log.

Healthcare AI Assistant *Co-developer, LLM & RAG*

Python, LangChain, Gemini 2.5 Flash, Groq, Hugging Face Embeddings, FAISS, Pytest

- Built a **RAG assistant** over a curated healthcare knowledge base with dynamic PDF ingestion, session-scoped indexing, streaming responses, and source citations; supports Gemini and Groq-hosted models behind one interface.
- Added **safety guardrails** for emergency detection, diagnosis and prescription prevention, and off-topic filtering, plus post-generation output validation; covered prompt and safety workflows with Pytest.

RAG-Based Document Intelligence System *Co-developer, Retrieval Pipeline*

Python, LangChain, FAISS, ChromaDB, FastAPI

- Built the **document question-answering pipeline** end to end: ingestion, chunking, embeddings, and semantic retrieval over custom document collections.
- Exposed retrieval through a **FastAPI REST API** so other applications can query it in real time.

AI-Powered Data Extraction and Summarization Agent *Co-developer, Agent & Extraction*

Python, LangChain, LLM APIs

- Built a **LangChain agent** that extracts, summarizes, and answers questions over structured and unstructured sources.
- Used prompt chaining, memory buffers, and output parsers to hold conversational context and return **structured JSON**.

RLM Agent Framework *Author, Agent Orchestration*

Python, DSPy, OpenRouter, Pyodide, Deno

- Built a **recursive LLM framework** that splits a complex task into smaller steps, runs them through reusable pipelines, and aggregates the results.
- Generates and executes model-written code inside **Pyodide/Deno sandboxes** so unsafe output never touches the host.

InvenIO – AI-Powered Patent Marketplace *Backend & Search Engineer*

FastAPI, Python, PostgreSQL, Pydantic, Vector Embeddings

- Built FastAPI services for patent submission, lifecycle management, and marketplace workflows with **Pydantic-enforced input contracts and RBAC**.
- Implemented **embedding-based semantic similarity search** to flag likely duplicate patents before submissions are accepted.

ApiWall – API Security & Monitoring Platform *Author, Backend*

FastAPI, PostgreSQL, Redis, Pydantic, Docker

- Built reusable FastAPI middleware for **JWT authentication, RBAC, brute-force protection, and request validation**, with rate limiting, IP reputation scoring, and an activity dashboard.
- Kept distributed rate-limit state in Redis, modeled data in PostgreSQL, and documented every contract in OpenAPI/Swagger.

Live Property-Auction Platform – Client Backend *Sole Backend Developer*

Python, Django REST Framework, PostgreSQL, Linux

- Took over a legacy production backend and run it independently: auction lifecycle APIs, migrations, query tuning, deployments, and incident investigation for live bidding traffic.

Education

B.E. in Computer Engineering **2022 – 2026**
Pacific School of Engineering, Gujarat Technological University (GTU) CGPA: 9.46/10

Achievements & Certifications

- Semi-Finalist, SAP Code Unnati Innovation Marathon; SAP Code Unnati Advanced Technology Program in Software Engineering, AI/ML, Deep Learning, and IoT.
- Python Programming Certification, GeeksforGeeks.