

Rajan Jasani

Generative AI Engineer | Backend Engineer

rajan.developer09@gmail.com | +91 70165 66752 | linkedin.com/in/rajanjasani | github.com/Rajan-Jasani9

Summary

Software Engineer with 2+ years of experience building production backend and Generative AI systems using Python, FastAPI, Django, PostgreSQL, Redis, React, and AWS. Delivered citation-grounded retrieval, agentic workflows, multimodal ingestion, structured LLM outputs, and asynchronous systems supporting products with approximately 3,500 weekly users and 1,200 students.

Experience

Software Engineer L2

May 2024 – Present

Crest Infosystems Pvt. Ltd.

- Developed robust full-stack applications and REST APIs using Django, Django REST Framework, React, PostgreSQL, and AWS, with a focus on scalable backend architecture, performance, and production reliability.
- Built production Generative AI workflows spanning retrieval, tool calling, structured outputs, guardrails, multimodal ingestion, asynchronous processing, and LLM evaluation across user-facing products.
- Designed ingestion pipelines processing 200+ articles per day from 35+ RSS feeds using Scrapy, Playwright, background workers, semantic indexing, and event-driven processing patterns.

Backend Developer Intern

January 2024 – May 2024

Crest Infosystems Pvt. Ltd.

- Developed REST APIs, PostgreSQL schemas, validation, serialization, and business logic using Python, Django, and Django REST Framework.

Projects

Ask Hip Hop – Citation-First AI Knowledge and Creator Platform

Python, FastAPI, React, TypeScript, PostgreSQL, Redis, AWS

- Engineered citation-first retrieval using multilingual E5 embeddings and Amazon S3 Vectors, combining knowledge-base filters, trust-tier prioritisation, recency scoring, conversation-aware retrieval, and source-linked responses for approximately 3,500 weekly users.
- Orchestrated a multi-project, multi-knowledge-base platform using FastAPI, SQLAlchemy, PostgreSQL, and Redis workers, enabling 40+ APIs, JWT authentication, email verification, and asynchronous ingestion while cutting request latency by 150 ms.
- Built multimodal ingestion for PDFs, websites, YouTube, Instagram, TikTok, RSS, and CSV, including Whisper Large V3 transcription and automated feeds from 35+ article sources and 10+ YouTube channels processing 200+ articles daily.
- Implemented agentic workflows with tool calling, structured outputs, evaluation-oriented batch testing, guardrails, and event-driven background processing for research, chat, ingestion, and knowledge-base workflows.

JudyAI – Clinical Documentation Platform

React, FastAPI, PostgreSQL, AWS Bedrock, Amazon Transcribe Medical

- Built a Human-in-the-Loop clinical AI platform converting recorded and live doctor-patient consultations into clinician-reviewable SOAP notes, structured medical data, and speaker-labelled transcripts.
- Designed S3, Lambda, Step Functions, Transcribe Medical, Bedrock, and WebSocket workflows with Cognito authentication, role-based access, audit records, retries, guardrails, and automated tests.

BinaAI – Curriculum-Aware TCM Learning Platform

Django, FastAPI, LangChain, PostgreSQL, Redis, AWS

- Designed curriculum-aware retrieval that routes student queries to relevant course material and generates source-grounded responses from 60+ ingested textbooks and 500+ hours of video content.
- Scaled the platform to 1,200 students with a FastAPI embedding service, AI quiz engine, document/video ingestion, streaming responses, structured outputs, and asynchronous LLM evaluation workflows.

Autonomous AI Video Production Agent – Side Project

TypeScript, React, Remotion, Node.js, SVG

- Built an agentic system that converts narration and transcripts into lesson plans, synchronised captions, animated architecture diagrams, thumbnails, and production-ready 1080p videos using tool-driven AI scene planning and deterministic Remotion rendering.

Midland Purchase Optimization

Python, FastAPI, PostgreSQL, React

- Developed procurement services for 80+ suppliers, including business workflows, reporting, data processing, and Generative AI contextual insights.

Technical Skills

Languages: Python, TypeScript, JavaScript, SQL
Backend and Distributed Systems: FastAPI, Django, Django REST Framework, Flask, SQLAlchemy, Pydantic, REST APIs, WebSockets, microservices, event-driven architecture, Apache Kafka, Amazon SQS, Redis workers
Generative AI and Retrieval: RAG, GraphRAG, knowledge graphs, entity-centric retrieval, agentic AI, LLM orchestration, LLM evaluation, tool calling, structured outputs, guardrails, MCP (Model Context Protocol), LangGraph, LangChain, hybrid search, reranking, embeddings, semantic search, OpenAI API, AWS Bedrock, Hugging Face, Whisper
Data: PostgreSQL, MySQL, Redis, pgvector, Amazon S3 Vectors, Pinecone
Cloud and DevOps: AWS ECS, EC2, S3, Lambda, Step Functions, RDS, Cognito, IAM, KMS, CloudWatch, Docker, CloudFormation, CI/CD, Linux
Frontend and Testing: React, Vite, Tailwind CSS, Git, pytest, Playwright, Scrapy

Certifications

AWS Certified AI Practitioner – Amazon Web Services
Certified in AI practices and methodologies
Verification Code: d86819f14bd14addb6b92861a915890c

Verify Credential

Education

B.E. in Computer Engineering
Gujarat Technological University

2020 – 2024
CGPA: 8.74/10

Community

Technical Speaker – Django India Meetup, Surat: Delivered a technical talk on mastering the Django ORM.
Django Day India Contributor: Led development of the Django India community event website.