

DIXIT JAIN

AI Engineer | Generative AI Engineer | Agentic AI & LLM Systems
dixitj13@gmail.com | India
[linkedin.com/in/thedixitjain](https://www.linkedin.com/in/thedixitjain) | github.com/thedixitjain | thedixitjain.com

SUMMARY

Generative AI Engineer with 2+ years building and shipping production AI systems across enterprise, government, and international client environments. Specializes in **Agentic AI, LLM applications, RAG, MCP, multi-agent orchestration, and AI automation**, with hands-on ownership across architecture, deployment, evaluation, observability, security, and reliability. Delivered **25+ production systems**, achieving **85%+ retrieval accuracy** and reducing inference cost by **~70%** through open-weight model deployment.

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript, TypeScript, Java

Generative AI & Agents: LangGraph, LangChain, Model Context Protocol (MCP), ReAct Agents, Multi-Agent Orchestration, Agentic AI, AI Agents, Retrieval-Augmented Generation (RAG), Hybrid Search, Function Calling, Tool Use, Text-to-SQL

LLM Platforms & Models: OpenAI API, Groq, Ollama, vLLM, Qwen, DeepSeek, Gemma, Hugging Face, Whisper, ElevenLabs

Evaluation & Observability: RAGAS, LLM-as-Judge Evaluation, Langfuse, OpenTelemetry, OTel Collector, Grafana, Loki, Tempo, Prometheus, Audit Logging

Security & Reliability: Prompt-Injection Defense, PII Redaction, Scope Allowlisting, Pydantic Validation, Keycloak, OIDC, RBAC, Kong API Gateway

Backend & Data: FastAPI, REST APIs, Server-Sent Events, WebSockets, RabbitMQ, asyncio, PostgreSQL, MongoDB, Redis, Qdrant, Pinecone, Next.js

Cloud & DevOps: Docker, Kubernetes, Azure AKS, Azure DevOps Pipelines, AWS EC2, Git, CI/CD, pytest

Automation & Enterprise: n8n, ERPNext, CRM/ERP Integrations, API Integrations, Workflow Automation

Machine Learning: PyTorch, Scikit-learn, YOLOv8, OCR, Pandas, NumPy, Computer Vision, NLP

EXPERIENCE

Generative AI Engineer

Jun 2026 - Present

In Time Tec

India

Engagement: Confidential Government Client - sovereign, on-premise executive AI assistant

Stack: *Python, FastAPI, LangGraph, MCP, Qdrant, MongoDB, PostgreSQL, Redis, RabbitMQ, Kubernetes, Keycloak, Langfuse, OpenTelemetry*

- Engineered the agent layer of a sovereign, air-gapped executive AI assistant, mapping **100%** of generated assertions to stored source references with **zero** external data egress.
- Deployed a bilingual Arabic and English agent using LangGraph, SSE token streaming, **5-way** intent classification, ReAct tool loops over MCP connectors, Qdrant retrieval, and multi-turn memory in Redis.
- Orchestrated asynchronous RabbitMQ insight pipelines scoring portfolio records against **4** PostgreSQL business rubrics and publishing **5** schema-validated contracts with idempotency and audit snapshots.
- Implemented **3** guardrail profiles for prompt-injection scanning, scope allowlists, PII redaction, and Pydantic validation, alongside a RAGAS harness covering **6** quality dimensions.
- Instrumented OpenTelemetry signals across Tempo, Loki, and Grafana with Prometheus alerting, Langfuse tracing, and request-level audit middleware, achieving a **90%** observability coverage gate on GPU-enabled Kubernetes.

Engagement: Government Policy Institution - conversational analytics over national statistical datasets

Stack: *Python, SQL, LLMs, RAG, Vector Search*

- Architected a multilingual text-to-SQL analytics system with query validation and source-cited answers, achieving **85%+ retrieval accuracy** and reducing analyst requests from **hours to minutes**.

AI Engineer

Jan 2025 - Jun 2026

UB Webs

India

- Delivered **25+** production AI systems for international clients across multi-agent support automation, document intelligence, conversational AI, and outbound workflows handling **1,000+ conversations per month**.
- Designed LangChain and LangGraph multi-agent workflows for support and internal operations, resolving the majority of tier-1 tickets with **no human handoff**.
- Integrated ERPNext, PostgreSQL, and **2** vector stores into hybrid retrieval pipelines, sustaining **85%+ top-k accuracy** on enterprise document search.
- Launched a real-time voice AI agent using Whisper and ElevenLabs, keeping end-to-end latency below **1.5 seconds** through streaming transcription and speculative synthesis.
- Migrated inference from hosted APIs to self-hosted open-weight models, reducing model spend by **~70%** with **no measurable quality drop** on held-out evaluation sets.

- Automated candidate screening with Salesforce Flow and created recruitment dashboards adopted by talent acquisition, eliminating manual handoffs from every requisition.

SELECTED PROJECTS

AgentFlow | *LangGraph, FastAPI, Groq, PostgreSQL, Qdrant, Next.js* | agentflow.thedixitjain.com

- Architected a multi-agent document intelligence platform converting **4** file formats into cited executive briefs, reducing review time by **85%** in user testing.
- Orchestrated retrieval, verification, and summarization agents in LangGraph with source-span validation that removes unsupported claims.

Industrial Manufacturing AI Assistant | *Generative AI, RAG, Agentic Workflows, Enterprise Integrations*

- Created an AI assistant for an international industrial manufacturing environment, enabling conversational access to domain knowledge and enterprise workflows.
- Connected enterprise data, retrieval, and workflow automation to support operational use cases while maintaining client confidentiality.

MarqHire AI | *LLMs, FastAPI, PostgreSQL, Pinecone, Voice AI* | marqhireai.techmarqx.com

- Launched an AI hiring platform unifying **4** screening stages: resume parsing, semantic job matching, candidate ranking, and autonomous voice interviews.
- Integrated email ingestion with vector search over an indexed resume corpus, converting inbound applications into ranked profiles with **zero manual sorting**.

Wellnix | *Python, YOLOv8, OCR, FastAPI, RAG* | wellnix.thedixitjain.com

- Combined YOLOv8 posture scoring, OCR label extraction, and a RAG nutrition knowledge base into one AI intelligence pipeline; selected for **Startup Mahakumbh 2025**.

EDUCATION

Bachelor of Technology in Computer Science

Aug 2021 - May 2025

JK Lakshmipat University

Jaipur, India

CERTIFICATIONS & ACHIEVEMENTS

Deep Learning Specialization, DeepLearning.AI | **Google Data Analytics Professional Certificate**, Google**Amazon ML Challenge 2024** - Finished in the top **15%** of the national leaderboard**Startup Mahakumbh 2025** - Selected for Wellnix, an AI fitness and nutrition platform