

DARPAN SARDA

AI / GenAI Engineer | RAG | Agentic AI | Python

Ahmedabad, India | 9081249454 | darpansarda7@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

AI/GenAI Engineer with 1.5+ years of experience building production-grade agentic AI systems — from NL2SQL platforms to multi-agent SaaS workflows and multimodal RAG pipelines. Skilled across the full stack: LLM orchestration (LangGraph, CrewAI), retrieval architecture (Qdrant, hybrid search, reranking), and backend/frontend development (FastAPI, React, Next.js). Focused on reliable AI systems with grounding, evaluation, and observability integrated into production workflows.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, SQL

Generative AI: LLMs, RAG, Agentic AI, Prompt Engineering

NL2SQL: Vanna AI

AI & Agent Frameworks: LangGraph, CrewAI, LlamaIndex

Retrieval & Vector Search: Qdrant, BM25, Hybrid Search, RRF, BGE Embeddings, BGE Reranking

Backend & APIs: FastAPI, REST APIs, WebSockets

Databases & Storage: PostgreSQL, MongoDB, Redis, MinIO

Frontend: React, Next.js, Tailwind CSS

LLM Inference & Infrastructure: vLLM, Docker

Observability & Evaluation: Arize Phoenix, RAGAS

PROFESSIONAL EXPERIENCE

Silvertouch Technologies, Ahmedabad

AI Full Stack Developer | Jun 2025 – Present AI/ML Intern | Jan 2025 – Jun 2025 (converted to full-time)

TenderFlow AI — Multi-Tenant SaaS Platform for End-to-End Tender Management

- Contributed to a multi-tenant SaaS platform for end-to-end tender drafting, review, and approval across a 3-stage workflow (Draft → Review → Approve) covering 6 tender types (Open, Limited, RFP, RFQ, EOI, Single Tender).
- Built AI-assisted section-by-section tender generation driven by configured PQ/TQ criteria and scope of work, with human-in-the-loop accept/reject/refine controls.
- Developed a query management module to ingest GeM/CPPP clarification queries, classify them, and auto-generate RAG-based answers grounded in tender documents, feeding into corrigendum management.
- Built a bid evaluation engine matching bidder-uploaded documents against PQ/TQ criteria with confidence scoring, supporting L1, LCS, and QCBS evaluation methods, on a platform with multi-LLM support (OpenAI, Anthropic Claude, Azure OpenAI), LLM usage tracing, and SLA dashboards.

Talk-to-DB (Ask DB) — NL2SQL Chatbot

- Built a natural language-to-SQL chatbot enabling business users to query the NCD database conversationally, using Vanna AI on an on-prem Qdrant vector store with a self-hosted GPT-OSS 20B model.
- Engineered the retrieval knowledge base from DDLs with column-level semantic annotations, business rules, and curated question-SQL pairs to ground SQL generation.
- Designed intent classification to route queries across business modules (federations, societies, banks) before retrieval, implemented a SELECT-only security layer with a 3-attempt retry mechanism, and evaluated system quality using RAGAS.

MyBotGenie — SaaS RAG Chatbot Platform

- Built a SaaS product that delivers an embeddable, script-based RAG chatbot trained on customer-provided documents and web-scraped content.
- Owned the retrieval pipeline end-to-end: implemented BGE small, base, and large embedding models alongside sparse embeddings for hybrid retrieval, combined with a BGE reranker to improve retrieval accuracy.

SAP Copilot — Agentic SaaS Platform for SAP HANA Operations

- Built a human-in-the-loop UI card system to pause agentic workflows for structured input, and real-time WebSocket-based agent status updates between frontend and backend.
- Enhanced Talk-to-HANA-DB logic (via Vanna Cloud) for SAP sales/purchase querying, plus analytics and auto-generated summary features.

EDUCATION

Bachelor of Engineering, Computer Engineering

Vishwakarma Government Engineering College, Ahmedabad | CGPA: 8.56/10

Sep 2022 – Jul 2025