Shaik Abdulla

aadilshaik.dtov@gmail.com 95021 76287 aadil-shaik Aadil-Shaik Portfolio

Professional Summary

AI Engineer (3 years) who builds and ships production-grade **end-to-end LLM and RAG systems** with measurable business impact. Experienced across the full stack — from model fine-tuning (LoRA, Transformers, YOLO11) and agentic workflows to high-performance APIs (FastAPI) and responsive interfaces (React/Next.js). I take full ownership of complex AI initiatives from problem discovery to production deployment and iteration.

Skills

AI/ML & LLM Engineering: LLMs, RAG Architecture, Agentic AI, LangGraph, Multi-Agent Systems, LoRA Fine-tuning (SFT + ORPO), Prompt Engineering, Evaluation Frameworks, GraphRAG

Computer Vision & Document Intelligence: YOLO11, OpenCV, Object Detection, Document Intelligence Pipelines

MLOps & Deployment: FastAPI, Docker, Kubernetes (OpenShift), CI/CD, Model Serving, Latency Optimization, Production Deployment

Backend & Data: Python, FastAPI, Oracle SQL, PostgreSQL, MongoDB, ETL Pipelines, Vector Databases (Qdrant, ChromaDB, FAISS)

Frontend: React, Next.js, TypeScript, Bootstrap

Cloud & Tools: Azure OpenAI, GCP Vertex AI, Git/GitHub, n8n

Experience

Software Engineer (AI/ML) — **Federal Bank – FedServ**, Bengaluru *Sept 2025 – Present*

Tech Stack: LangChain, LangGraph, LoRA (SFT + ORPO), Llama 3.1, Qdrant, YOLO11, FastAPI, n8n, OpenShift

- Led development of a production **agentic AI workflow** using n8n and a **LoRA fine-tuned LLM** (SFT + ORPO) that enabled natural language discovery of relevant Oracle DB tables across thousands of schemas. Designed a multi-vector RAG pipeline combining table name, column metadata, and LLM-generated business description embeddings — achieving **85% success rate within 2–3 conversational turns** and reducing manual discovery time from weeks to minutes.
- Engineered an end-to-end **document intelligence pipeline** using Llama 3.1, Qdrant, and cross-encoders, achieving **95% accuracy** in extracting 21+ fields from diverse insurance documents.
- Custom-trained and productionized **YOLO11s** computer vision models for signature and stamp detection with parallel OpenCV inference pipelines for high-throughput verification.
- Architected a **hybrid RAG pipeline** combining dense vector search, BM25 retrieval, and cross-encoder reranking to significantly improve precision on domain-specific banking terminology.
- Designed a **multi-heuristic confidence scoring engine** with fuzzy entity resolution to automate maker-checker workflows across six weighted risk dimensions.
- Built and containerized a high-performance async **FastAPI** backend with mTLS security on Red Hat OpenShift, delivering enterprise-wide AI capabilities.

Associate Data Scientist — **Dzitalinfo**, Hyderabad *May 2024 – Aug 2025*

Tech Stack: GPT-4o, Gemini 1.5 Pro, LangChain, Whisper, React/Next.js, FastAPI, MongoDB

- Designed and deployed multiple production **GenAI chatbots** for client websites using **GPT-4o** and **Gemini 1.5 Pro**, integrated with custom RAG pipelines and LangChain, reducing customer support tickets by **over 60%**.
- Built an end-to-end **meeting transcription and intelligence system** using **Whisper medium** for transcription + diarization, followed by automated chat note generation with **Gemini 1.5 Pro**.
- Architected and scaled full-stack AI-powered applications (Inventory Management, Restaurant Billing, Gokarting Park Booking) using React/Next.js, FastAPI/PHP, and MongoDB/SQL.

Data Science Intern — *Dzitalinfo, Vizianagaram* *Nov 2023 – Apr 2024*

- Developed a full-stack **Attendance Management System** with real-time reporting and React/SQL backend while starting work on production GenAI chatbots using advanced prompt engineering and Chain-of-Thought techniques.

Web Developer (Intern) — *ION Technology Solutions, Vijayawada* *Jul 2023 – Oct 2023*

- Built the complete corporate website siddachem.com from scratch using **Bootstrap**, HTML5, CSS3, and JavaScript for the CEO's oilfield chemicals export company.

Projects

American Society for Reproductive Medicine (NGO) *Remote*

MAC Tool: Built a large-scale educational platform using Next.js (100+ pages, 1700+ components) and integrated a production GenAI chatbot (GPT-4o + RAG) for context-aware personalized learning assistance for medical professionals.

Edge Tool: Developed an interactive AI-powered grading and real-time feedback system with intelligent performance analytics and automated guidance.

Education

GMR Institute of Technology (GMRIT) CGPA: 8.42 / 10 *Sept 2020 – Apr 2024*

B.Tech in Information Technology — AI & ML Specialization