JAYANT PATEL

+91 9358957326 | jayant.07124@gmail.com | linkedin.com/in/jayantpatel71 | github.com/Jayant71

PROFESSIONAL SUMMARY

AI Developer with a track record of building scalable **Agentic systems** and **Advanced RAG pipelines**. Leveraged **LangChain and LangGraph** to deliver **measurable improvements** in retrieval accuracy and model efficiency, ensuring high-fidelity outputs in complex, real-world deployments.

SKILLS

Languages: Python (Advanced), SQL, TypeScript

AI & LLM Stack: Agentic AI (LangGraph), Advanced RAG (Hybrid Search, Reranking), OpenAI/Anthropic APIs, Vector Databases (ChromaDB, Pinecone), Semantic Chunking, Ragas (Evaluation)

Frameworks & Tools: LangChain, PyTorch, FastAPI, n8n (Automation), Streamlit, Hugging Face, Git, Docker

Database & Cloud: Supabase, PostgreSQL, RESTful APIs, Firebase

Relevant Coursework: Data Structures & Algorithms, LLM Application Development, Neural Networks, DBMS

EXPERIENCE

Ushree - AI Agent Developer Intern   **August 2025 – December 2025**
- Engineered a mobile-first internal management platform using **Supabase (PostgreSQL)** for **real-time data synchronization**, supporting educator activity tracking for 50+ users.
- Developed **automated data pipelines** for triggered notifications and dynamic report generation, streamlining manual coordination and improving **operational efficiency by 50%**.
- Architected scalable database schemas and **automated logic modules** for timetable and payroll management, ensuring system-wide data integrity and **100% calculation accuracy**.

Iron Willed Tech Ltd. - Software Developer Intern   **January 2025 – June 2025**
- Developed a **GenAI-powered virtual try-on engine**, streamlining the backend pipeline for **high-fidelity outfit generation** and real-time responsiveness.
- Integrated **MediaPipe face-mesh** and computer vision models to prototype **real-time AI-driven makeup effects** on mobile devices.
- Built a modular outfit-layering system for Android using a **decoupled architecture** to support scalable feature updates.

IIT Jammu - Research Intern   **February 2024 – July 2024**
- Conducted a comprehensive **comparative analysis of YOLO architectures**, benchmarking trade-offs between **inference latency, FLOPs, and mAP** across diverse datasets.
- Engineered a **lightweight model optimization** for YOLOv8n, achieving an **80.95% reduction in model size** and an **83.75% parameter compression** for edge deployment.

PROJECTS

Retrieval Augmented Generation (RAG) | *Python, LangChain, OpenAI API*   **November 2025 - January 2026**
- Engineered a **high-precision pipeline** using **Hybrid Search (BM25 + Vector)**, **Flashrank Reranking**, and **Multi-Query expansion** to maximize retrieval accuracy.
- Developed an automated processing engine supporting PDF/DOCX formats with **Semantic Chunking** to improve context alignment and **minimize hallucinations**.
- Integrated **Ragas and LangSmith** for **metric-driven optimization**, ensuring **production-grade performance** across faithfulness and answer relevancy.

AI-Powered Book Reading Attention Monitor | *Python, PyTorch, CV*   **January 2025 - May 2025**
- Architected a **real-time attention monitoring system** for book reading integrating **L2CS gaze estimation** and **YOLOv12s** model.
- Achieved an **mAP of 0.87** for book presence and implemented an attention analysis module processing at $\sim$25 **FPS**.

PUBLICATIONS AND CERTIFICATIONS

AI-Powered Attention Monitoring System For Book Reading – High Technology Letters, Vol 31, July 2025
Development of Mobile Application for Sickle Cell Anemia Support – YMER, Vol 23, June 2024
Advanced RAG with Vector Databases and Retrievers – Coursera, Nov 2025