JAYANT PATEL

Korba, Chhattisgarh 495452 | +91 9358957326 | jayant.07124@gmail.com | linkedin.com/in/jayantpatel71 | github.com/Jayant71

PROFESSIONAL SUMMARY

AI Developer specializing in Large Language Models (LLMs), Agentic AI systems, and production-grade MLOps pipelines. Experienced in building scalable AI applications with PyTorch, LangChain, and FastAPI, with proven outcomes in improving retrieval relevance, model efficiency, and end-to-end deployment reliability across real-world systems.

EDUCATION

University Teaching Department (CSVTU) | CPI: 8.34/10 **December 2021 - July 2025**
B.Tech (Honours) in Computer Science & Engineering (Artificial Intelligence) *Bhilai, Chhattisgarh*

SKILLS

Programming Languages: Python, SQL, JavaScript/TypeScript;

Technologies: Machine Learning, Deep Learning, AI Agents, Agentic AI, Computer Vision, Object Detection, Large Language Models (LLMs), NLP, Retrieval Augmented Generation (RAG), Artificial Intelligence (AI), MLOps, Vector DBs, OpenAI API, LlamaIndex;

Frameworks / APIs: LangChain, PyTorch, FastAPI, React, REST APIs, OpenAPI, n8n, Firebase, Supabase, Streamlit, Gradio, Hugging Face, Transformers;

Relevant Coursework: Data Structures & Algorithms, Predictive Modeling, Data Science, Database Management, OOPs, Cloud Computing, Data Visualization.

INTERNSHIP

Clonefutura **August 2025 – Present**
AI Agent Developer Intern *Remote*
- Developed automations for educator tracking, attendance, payments, and timetables, improving operational workflow efficiency across 50+ active users.

Iron Willed Tech Ltd. **January 2025 – June 2025**
Software Developer Intern *Remote*
- Engineered outfit-creation features for an Android app, enabling drag-and-drop layering, image import, and consistent rendering via modular code architecture.
- Integrated MediaPipe face-mesh to prototype AI-driven makeup effects with real-time responsiveness
- Developed a virtual try-on pipeline that transmitted clothing and user images to a GenAI backend, enabling high-quality outfit generation and automated blending.

IIT Jammu **February 2024 – July 2024**
Research Intern *Jammu, Jammu & Kashmir - Hybrid*
- Benchmarked and optimized YOLO models (v3–v8) on the CHV PPE dataset; achieved 80.95% model size reduction and 83.75% parameter reduction with minimal accuracy loss.
- Conducted hyperparameter tuning, pruning, and image augmentation to improve inference efficiency on edge-deployed systems.

PROJECTS

Retrieval Augmented Generation (RAG) | *Python, FastAPI, Streamlit, LlamaParse, OpenaiAPI* | *Link* **November 2025**
- Developed production-grade, multi-tenant RAG platform (FastAPI, Qdrant, PostgreSQL, MinIO, LlamaIndex, LlamaParse, GPT-4o) providing 10 retrieved chunks/query, top-3 reranking and 30–40% improved retrieval relevance across 5+ document formats.
- Built an infrastructure-agnostic system with swappable DB/storage backends (SQLite/PostgreSQL, Local/S3-MinIO) via the Strategy pattern, reducing setup effort 50%. Delivered 8 REST CRUD APIs and a Streamlit UI for 2–3× faster testing and easy external integration.

AI-Powered Book Reading Attention Monitor | *Python, PyTorch, Object Detection* | *Link* **January 2025 - May 2025**
- Designed a real-time attention monitoring system integrating L2CS gaze estimation and a custom YOLOv12s model for book detection.
- Achieved an mAP of 0.87 on the validation set for book presence and state using a curated dataset. Using this model implemented an attention analysis module correlating 3D gaze vectors with book locations, processing at ~25 FPS on GPU.

PUBLICATIONS

AI Powered Attention Monitoring System For Book Reading – High Technology Letters, Vol 31, Issue 7, July 2025.
Development of Mobile Application for Sickle Cell Anemia Treatment Support System – YMER, Vol 23, Issue 06, June 2024.

CERTIFICATIONS

- **Foundation of Generative AI Nanodegree** – Udacity, February 2025 - Certificate
- **Google Cloud Fundamentals: Core Infrastructure** – Google Cloud & Coursera, November 2023 - Certificate