

Rushikesh Jadhav

AI/ML Engineer | Applied AI & LLM Systems | Pune, India

jrushikesh44@gmail.com · github.com/Rushi543 · linkedin.com/in/rushi-jadhav32 · +91 8446030311

SUMMARY

Machine Learning Engineer at TCS Research with experience developing, evaluating, and deploying ML and LLM systems. Built production-oriented pipelines for performance prediction, vector search, agentic AI, and multimodal information extraction using Python, PyTorch, LangGraph and Qdrant.

TECHNICAL SKILLS

ML & Deep Learning: PyTorch, scikit-learn, Transformers, Computer Vision, NLP, OCR

LLM & Applied AI: LLMs, LangGraph, RAG, OpenAI, Hugging Face, Embeddings, Qdrant, FAISS

ML Infrastructure: MLflow, KServe, Kubernetes, Docker, AWS, MinIO

Languages & Backend: Python, SQL, FastAPI, Git, Linux

EXPERIENCE

Tata Consultancy Services — TCS Research (TRDDC)

Machine Learning Engineer - Pune | May 2025–Present

- Developed a multi-target Random Forest regression pipeline using 1,000+ Qdrant benchmark runs across hardware, workloads, concurrency, and HNSW parameters; predicted average, p95, p99 latency, and load duration with $R^2 = 0.9$, with automated evaluation.
- Built a production multi-turn LLM agent using LangGraph StateGraph with an 8-node workflow for session hydration, relevance checking, structured extraction, recommendation generation, and response formatting; maintained typed agent state and execution traces across conversational workflows.
- Engineered parallel LLM extraction using 3 concurrent inference calls, reducing extraction latency from 15s to 5s; implemented timeout handling, regex fallbacks, relevance gating, and a 9-layer input security pipeline for resilient LLM processing.
- Developed multimodal architecture-diagram extraction pipelines using Faster R-CNN (ResNet50-FPN), OpenCV, and EasyOCR; trained and evaluated object detection models with confidence thresholding, non-maximum suppression, and mAP-based evaluation to convert visual diagrams into structured data.

Odoo

Software Developer Intern - Gandhinagar | Jan 2024–July 2024

- Wrote SQL queries to repair corrupt database records, improving data integrity by 20%.
- Fixed critical bugs and improved the RD editor and JavaScript framework, contributing to a 15% improvement in system stability and performance; enhanced keyboard navigation and screen-reader accessibility.

PROJECTS

Vibe Finder — Qdrant, FastAPI, Embeddings, UMAP, OpenAI | [GitHub](#) | [Live Demo](#)

- Built a cross-domain semantic recommendation system by embedding user preferences from GitHub, Steam, Spotify, and AniList and retrieving similar items through Qdrant vector search.
- Deployed a Python/FastAPI backend on a self-managed Oracle Cloud VM with a React frontend, supporting multi-source data ingestion and personalized recommendations.
- Integrated an LLM pipeline to generate concise explanations for recommendations and applied UMAP to visualize the learned embedding space through an interactive dashboard.

Conversational AI Chatbot — PyTorch, Hugging Face Transformers, DialoGPT

- Fine-tuned DialoGPT-medium on 4,000+ dialogue pairs, building preprocessing, tokenization, training, and evaluation pipelines for conversational response generation.
- Deployed the trained transformer model as a Discord chatbot and published the model to Hugging Face for reproducible inference.

EDUCATION

Indian Institute of Information Technology, Vadodara: B.Tech, Computer Science, 2020–2024