

Anish Kumar Maurya



+91-7307800543



anishkumarmaurya12@gmail.com



LinkedIn



GitHub

PROFESSIONAL SUMMARY

ML Engineer with 1.6+ years of industry experience and M.Tech in Cyber-Physical Systems (IIT Jodhpur), specializing in **LLM fine-tuning, quantization, and CPU-optimized inference**. Proven track record of deploying **production-grade GenAI systems** using FastAPI, Celery, Docker, and OpenVINO. Strong expertise in **Agentic AI, RAG pipelines, NLP, Document AI**, and scalable ML infrastructure.

EDUCATION

M.Tech in Cyber-Physical Systems

Indian Institute of Technology (IIT), Jodhpur

Jun 2022 – Jun 2024

CGPA: 7.72

B.Tech in Electrical Engineering

Institute of Engineering and Rural Technology

Aug 2013 – Jul 2017

CGPA: 7.5

EXPERIENCE

•Datafoundry Private Ltd

June 2024 – November 2025

ML Engineer

Bengaluru, On-site

- Designed and deployed a Dockerized, event-driven LLM pipeline using FastAPI and Celery to perform NER on 40K+ Supreme Court judgments; asynchronously extracted entities, indexed in Elasticsearch, and auto-synced to Dgraph via optimized knowledge graphs with <2s query latency.
- Conducted EDA on medical NER datasets (entity distribution, class imbalance, token length analysis) and fine-tuned Gemma3 and Qwen3 using LoRA with 4-bit quantization; optimized and exported models to OpenVINO, ONNX, and GGUF, achieving up to 33% faster CPU-only inference.
- Fine-tuned a YOLO11 object detection model on large-scale annotated medical documents, achieving ~99% checkbox detection accuracy; enabled region-aware OCR.
- Fine-tuned BERT for sentiment analysis on Prime Minister's speeches, achieving 95.8% precision and 96.5% recall, outperforming Random Forest and Logistic Regression.

•Indian Institute of TechnologyJodhpur

June 2022 – May 2024

Teaching Assistant

Jodhpur, India

- Teaching Assistant: Data Science , ML Lab , Embedded Systems , Computer Vision Lab

PROJECTS


•Automatic Image Classification for Surveillance-based CPS Applications

Computer Vision

M.Tech Project

- Developed a real-time image detection model using Python, OpenCV,YOLO, and the VisDrone dataset, applying deep learning techniques for efficient training and validation; deployed the model on Raspberry Pi with Google Coral TPU, achieving 28 FPS and demonstrating edge-based intelligent surveillance.

•Agentic AI Chatbot

Agentic AI Systems 

Personal Projects

- Designed and deployed a sentiment-aware, human-in-the-loop bug-analysis and intelligent email-automation system, along with an agentic AI chatbot powered by LangGraph,LangChain,LangSmith, SQLite-based conversational memory, and multi-tool orchestration (RAG, web search, weather, finance, utilities); implemented an automated DevOps CI/CD delivery pipeline on AWS EC2.

TECHNICAL SKILLS

- Programming:** Python , C/C++,HTML/CSS, JavaScript, SQL
- Tools & OS:** Git, Linux, Windows, Visual Studio Code,OCR
- Frameworks:** LangChain, LangGraph, LangSmith, Streamlit, TensorFlow, Keras, PyTorch,VectorD
- Libraries:** Pandas, NumPy, Scikit-learn,Ollama, FAISS, Chroma, Seaborn, Matplotlib
- Computer Vision:** YOLO,Object Detection,image segmentation ,Image Annotation
- Cloud & DevOps:**Microsoft Azure, AWS (LLM Deployment, EC2, S3), Docker,CI/CD

ACHIEVEMENTS

- GATE 2022**Qualified (EE) with 97.48 percentile (AIR 1756) and (IN) with 97 percentile (AIR 601)