

Rajarshi Das

rajarshi Axel@gmail.com

+91-7002136866

Guwahati, Assam

[LinkedIn](#)

[GitHub](#)

[Portfolio](#)

PROFESSIONAL SUMMARY

AI/ML Engineer specializing in the **local open-weights LLM ecosystem**, agentic AI workflows, and production-grade machine learning systems. Experienced in running, benchmarking, and optimizing open-source models using **Ollama, llama.cpp, and LM Studio**; proficient in **model quantization, RAG pipelines, and Model Context Protocol (MCP)**. Built AI-powered CRM systems, deployed custom **YOLOv11** food-detection models to TensorFlow Lite, and fine-tuned **ResNet-50** achieving 79% classification accuracy on 3D MRI datasets. Skilled in **LangChain, LangGraph, PyTorch, TensorFlow, FastAPI, and React**, delivering end-to-end AI solutions from model training to scalable API deployment.

EXPERIENCE

Full-Stack & AI Engineer

2023 – Present

Auctus Digitech

- Engineered production-ready CRM systems and AI chatbots leveraging local open-weights LLMs (Gemma/Qwen via Ollama) and agentic workflows for 5 clients, reducing API operational costs by up to 60% while improving lead conversion rates by ~30%.
- Developed personalized email generation systems integrating LLMs and RAG with internal CRM databases to automate tailored communication workflows, boosting engagement rates by 40%.
- Designed and integrated secure, scalable REST APIs handling 10K+ daily requests, connecting ML workflows with client-facing web interfaces.
- Optimized and deployed 4+ production web applications using React, Vite, and TypeScript, slashing page load times by ~40% through code splitting and lazy loading optimizations.

Tech: Python, LLMs, LangChain, LangGraph, FastAPI, React, TypeScript, PostgreSQL, Docker

PROJECT

CalCal – AI-Powered Fitness & Nutrition App [Play Store](#)

2025 – Present

React Native, TensorFlow Lite, Python, YOLOv11, FastAPI, Node.js

- Fine-tuned a custom PyTorch YOLOv11 segmentation model on 40,000+ food images (~87% mAP), porting it to ONNX and then to TensorFlow Lite for edge deployment with 300ms on-device inference.
- Integrated a YOLOv8 object detection model to identify known reference objects, enabling real-time scale calibration for accurate food volume and quantity estimation.
- Architected and published the full-stack mobile application to the Google Play Store, spanning 3 microservices (FastAPI, Node.js, TFLite inference) and integrating the CalorieNinjas API.

Tech: YOLOv11, YOLOv8, ONNX, TensorFlow Lite, Python, FastAPI, React Native, Node.js, Computer Vision

EDUCATION

Bachelor of Technology – Information Technology

2019 – 2023

Sikkim Manipal Institute of Technology

KEY SKILLS

Languages

Python • TypeScript • JavaScript • Kotlin • Java • SQL

Frameworks & Libraries

PyTorch • TensorFlow (Lite) • Keras • OpenCV • HuggingFace • NumPy • Pandas • FastAPI • React • React Native • Node.js

Databases & Vector Stores

PostgreSQL • MySQL • Qdrant • ChromaDB • DynamoDB

AI / ML & Deep Learning

LLMs • NLP • Generative AI • Computer Vision • Model Fine-Tuning • Model Quantization • Object Detection & Segmentation • RAG • Agentic Workflows • LangChain • LangGraph • MCP • Ragas

LLM Tools & Platforms

Ollama • llama.cpp • LM Studio • Continue • Cline • Arize Phoenix • LangSmith • OpenTelemetry

Tools & Infrastructure

Git • Docker • Kubernetes • Linux • AWS (EC2) • Microservices • Vite • ComfyUI • Unity (AR/VR)