NIKHIL RAJESH NIGAM

+1 (608)-960-6420 · niknigam96@gmail.com · linkedin.com/in/nikhil-nigam-31131b12a · github.com/nnigam96

Experienced ML Engineer delivering productionized AI/ML systems. Proven track record of leading projects from concept to deployment, with expertise in LLMs/RAG, recommendation systems, Agentic AI and MLOps.

TECHNICAL SKILLS

ML Frameworks, Libraries & Tools: PyTorch, Lightning, LangChain, Milvus, MLflow, Weights & Biases, TensorFlow, Ollama

Languages: Python, SQL | **Deployment & Cloud:** Flask, Docker, AWS (boto3, EC2, S3, EFS, Cognito), Linux, Git

WORK EXPERIENCE

modlee Inc, *Machine Learning Engineer*   Sept 2023 - Nov 2025
- **Engineered and deployed a Milvus vector search system** for sub-second model recommendation, analyzing **240+ meta-features** via a two-stage retrieval algorithm to drive a **5x experiment catalog expansion**.
- **Accelerated a compliance workflow** with a privacy-first hybrid OCR pipeline using **EasyOCR** + local **Ollama**, layering a **LangChain AI agent** for grounded, natural-language coverage reasoning.
- **Packaged and shipped a reusable Agentic Framework** with a formal **Tool Registry**, enabling partners to deploy schema-driven **text-to-SQL/JSON agents** and drastically cutting enterprise onboarding time.
- **Standardized MLOps velocity** with integrated **MLflow** and built a **"Retcon" pipeline** that backfilled **100s** of machine learning experiments from **WandB**, regenerating artifacts and meta-features for the vector system.

InterDigital Inc, Emerging Technologies Lab - *Deep Learning Intern*   May 2022 - Aug 2022
- Achieved a **40% bandwidth reduction** over the SOTA baseline by developing a **novel CNN autoencoder**, demonstrating feasibility for more efficient data transmission over constrained wireless networks.
- Improved signal reconstruction quality by integrating **Convolutional Block Attention Modules** into a Fully Convolutional Autoencoder architecture, validated against standard image quality benchmarks.

Vodafone Intelligent Services, Vodafone PLC - *Software Engineer*   Jul 2018 - Jul 2021
- Designated **Subject Matter Expert** of FinX, an in-house tool that automated all expense management and cost benefit analysis in Vodafone Group PLC.
- Led the end-to-end development of a mission-critical financial API within a CI/CD pipeline, **automating the expense approval** lifecycle for a 5,000+ employee organization.
- Cut average expense approval time by 50% by designing and implementing a secure, 'lazy approval' UX integrated with the company's primary email framework.

EDUCATION

University of Wisconsin - Madison (WI, USA), *M.S in ECE, (Specialization: Machine Learning),* **GPA:** 3.6/4.0   May 2023
Maharashtra Institute of Technology - Pune (MH, India), *B.Engg. in Electronics & Telecommunication,* **Grade:** A   May 2018

RESEARCH PROJECTS

Distributed LLM Systems: Federated Learning & Split Inference
- **Architected a fault-tolerant distributed inference engine** using raw TCP and torch.rpc, enabling **federated LoRA** training for LLMs across heterogeneous consumer devices with high network latency

Privacy-Preserving Edge Agent: Automated Context Synthesis
- **Developed an agentic workflow** using LangGraph and quantized Llama 3 to **orchestrate a full RAG pipeline**, combining semantic retrieval on 1000s of data samples with context-aware generation of email drafts via strictly local inference

Doc-Assist: Local RAG & Evaluation Pipeline
- **Engineered a privacy-preserving retrieval system** for document-level Q&A, implementing a custom evaluation framework to measure factual consistency i.e Hallucination Rate and Answer Relevance

Denoising and Intensity Inhomogeneity Correction of MRIs
- **Optimized N4 Bias Correction pipelines** by integrating a MONAI-based UNet with **Stable Diffusion schedulers**, reducing image clarity error rate metrics by 68% for coil-constrained MRI scanners