

# NIKHIL RAJESH NIGAM

+1 (608)-960-6420 • [niknigam96@gmail.com](mailto:niknigam96@gmail.com) • [linkedin.com/in/nikhil-nigam-31131b12a](https://linkedin.com/in/nikhil-nigam-31131b12a) • [github.com/nnigam96](https://github.com/nnigam96)

*ML Engineer with 5+ years of experience delivering production 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
- **Deployment & Cloud:** Flask, Docker, AWS (boto3, EC2, S3, EFS, Cognito), Linux, Git
- **Languages:** Python, SQL

## WORK EXPERIENCE

modlee Inc, *Machine Learning Research 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 (**EasyOCR** + local **Ollama**), layering a **LangChain/OpenAI agent** for grounded, natural-language coverage reasoning.
- Packaged and shipped a **reusable Agentic Framework** (MCP Server) 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 external runs (W&B, MLflow), regenerating artifacts and meta-features for the vector system.

InterDigital Inc, *Emerging Technologies Lab - Deep Learning Research 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

Doc-Assist: LLM Powered Document Retrieval and Q&A System

- Built a **privacy-preserving RAG system using FAISS and LLaMA 2**, validated on an internal Q&A dataset to ensure high factual consistency and relevance in generated answers for document-level search.

Denoising and Intensity Inhomogeneity Correction of MRIs

- Adapted N4 Correction with MONAI based **UNet via stable diffusion with modified noise scheduler** to streamline usage of coils needed for high quality MRI denoising

## CERTIFICATIONS

- **Specialization in Natural Language Processing** (*Advanced Specialization offered by Deep Learning.ai*)
- **Advanced Computer Vision** (*Advanced Specialization offered by Deep Learning.ai*)

Jul 2023

Jul 2023