DILIP FRANCIES

+1-812-803-4479 | dfranci@iu.edu | LinkedIn | GitHub | Portfolio

EDUCATION

Indiana University, Bloomington — Bloomington, IN

Master of Science in Data Science | ***GPA: 3.84/4.0*** — *Aug 2023 – May 2025*

- *Relevant Coursework: Deep Learning Systems, Applied Machine Learning, Big Data Applications, Cloud Computing, Applied Algorithms, Computer Vision, Elements of AI, Scientific Visualization*

EXPERIENCE

Senior Data Analyst - Machine Learning and Computer Vision — Nov 2021 – Jul 2023

Hexaware Technologies — *Mumbai, India*

Enterprise ML Authentication & Monitoring Platform

- **Architected and deployed** end-to-end ML system serving 2,000+ employees for secure workspace monitoring, integrating facial recognition (99% accuracy) and real-time object detection (92% accuracy) for automated policy enforcement
- **Production Deployment at Scale**: Packaged optimized models as Windows executable distributed across enterprise; processed 10K+ daily authentications with <200ms latency, saving estimated **$100K+ annually** in security overhead
- **Model Optimization**: Reduced model size **93%** (467MB → 30MB) using TensorRT and TensorFlow Lite quantization (INT8/FP16), enabling edge deployment across 2,000 devices
- **Cloud Infrastructure**: Deployed parallel containerized version using Docker + TensorFlow Serving on AWS ECS; conducted A/B testing proving edge deployment **reduced latency 3× and infrastructure costs 80%** vs cloud
- **Full-Stack Engineering**: Built 30+ FastAPI endpoints for a React dashboard enabling real-time monitoring; designed a MongoDB Atlas schema supporting 2,000+ concurrent users; implemented a GitHub→Heroku CI/CD pipeline.
- **Database Architecture**: Designed MongoDB Atlas schema supporting authentication, session tracking, violation logs, and analytics for 2,000+ concurrent users
- *Recognition: **Hexaware YUVA Award** for exceptional technical contribution*

Audio Analytics Intelligence Platform

- **Engineered serverless** ML pipeline processing **100k+** call center recordings using AWS Lambda, S3, Transcribe, and AssemblyAI for automated sentiment analysis at scale
- **Infrastructure Management**: Designed FastAPI microservices and MongoDB schema; managed AWS IAM, cost optimization, and service architecture for production workloads

Research and Data Analyst Intern - Machine Learning — Jan 2024 – May 2025

Indiana University - Institutional Analytics — *Bloomington, IN*

Natural Language to SQL System with LLM Fine-Tuning

- Built production NL2SQL system integrating FAISS vector databases with PEFT-tuned models (SQLCoder-7B, LLaMA-3B) using RAG architecture, enabling natural language queries across university data warehouse (1,000+ tables)
- **LLM Optimization**: Implemented LoRA fine-tuning **reducing training compute 80%** while maintaining query accuracy

Explainable AI for Student Outcomes

- **Predictive Modeling**: Developed tree-based models with SHAP analysis identifying key drivers of student success; delivered insights to VP of Student Affairs affecting 40,000+ students
- **Published Research**: *Co-authored conference paper* on interpretable ML for educational outcomes

Fairness & Bias Mitigation

- **Algorithmic Fairness**: Reduced demographic bias **75% in equalized odds** using IBM AIF360, ensuring fair predictions across protected groups

Scalable ML on HPC

- **Distributed Training**: Built PyTorch experimentation framework on Slurm HPC clusters, **accelerating training 3×**
- **Large-Scale Clustering**: Processed 20M+ student records using Dask and variational autoencoders, clustering 16,000 courses with **7% improved silhouette score**

Data Visualization

- *Winner representing IU at **Big Ten Data Visualization Championship***

Computer Vision ML Engineer Apr 2025 – Present

AI & Economics Research Collaboration (Luddy School, Kelley School, Wharton, Yale) *Bloomington, IN*

Large-Scale Facial Analysis Pipeline

- Engineered CV pipeline processing **2M+ images** using Vision Transformers, ArcFace, and StyleGAN2 for Big Five personality research across 4 institutions
- **Distributed Computing**: Implemented multi-GPU training (PyTorch DDP, Ray) on HPC infrastructure processing 500K+ images with batched inference
- **MLOps**: Established WandB logging, Ray Tune hyperparameter optimization, and model versioning for reproducible multi-institutional research
- **Multimodal Modeling**: Integrated facial embeddings with aesthetic metrics achieving **1.2× improvement** in salary/credit prediction accuracy
- **Research Impact**: Contributing to forthcoming publication on personality-appearance correlations | *Paper*

Multimodal ML Engineer May 2024 – Present

Indiana University - Department of History *Bloomington, IN*

AI-Powered Historical Artifact Retrieval

- Built multimodal search system using DINOv2 and CLIP for **67K+ ivory artifacts**, outperforming commercial AI tools in semantic accuracy with sub-second FAISS-based retrieval | *Archives of Ivory*

Technical Projects

Emergency Response Multi-Agent System LangGraph + GPT-4o

- Built AI-powered 911 disaster response system using LangGraph/LangChain coordinating Police/Fire/Medical agents
- Integrated GPT-4o, MongoDB geospatial indexing, and React dashboard for real-time incident management

GPT-2 Inference Optimization Quantization + Flash Attention

- Achieved **3.6× speedup and 75% memory reduction** through quantization, Flash Attention, & speculative decoding
- Converted Conv1D to Linear layers for Tensor Core optimization on RTX 4070

Technical Skills

- **ML/AI Frameworks**: PyTorch, TensorFlow, Scikit-learn, XGBoost, Hugging Face Transformers, PyTorch Lightning
- **Computer Vision**: ResNet, Vision Transformers (ViT), YOLO, CLIP, DINOv2, Facial Recognition, Object Detection, Image Segmentation, Grad-CAM
- **LLMs & NLP**: LangChain, LangGraph, vLLM, Text Generation Inference, Fine-tuning (LoRA, PEFT), Prompt Engineering, RAG, Vector Search (FAISS)
- **MLOps & Deployment**: Docker, TensorFlow Serving, TFLite, AWS (ECS, Lambda, S3, EC2, Transcribe), Model Optimization (TensorRT, Quantization), CI/CD, Monitoring (Prometheus, Grafana)
- **Distributed Systems**: PyTorch DDP, Ray, Dask, Multi-GPU Training, Slurm HPC, Model/Data Parallelism
- **Data Engineering**: MongoDB, PostgreSQL, MySQL, ETL Pipelines, FastAPI, REST APIs
- **Cloud & Infrastructure**: AWS (extensive), Azure, Heroku, IAM Management, Cost Optimization
- **ML Experimentation**: MLflow, WandB, Neptune.ai, Ray Tune, Optuna, A/B Testing
- **Programming**: Python, R, SQL, Bash
- **Data Science**: NumPy, Pandas, Dask, CuDF, SciPy, OpenCV, Scikit-image, Matplotlib, Seaborn, Plotly, Tableau

Certifications

- Deep Learning Specialization (Coursera - Andrew Ng)
- Convolutional Neural Networks Specialization (Coursera)
- Image Segmentation (Coursera)

Awards & Recognition

- **Hexaware YUVA Award** - Outstanding technical contribution beyond role expectations (2023)
- **Big Ten Data Visualization Championship** - Selected to represent Indiana University (2025)