

Swaminathan Sankaran

Buffalo, NY | +17165332867 | swaminathansankaran21@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

MS Data Science student at University at Buffalo with experience in developing and deploying NLP and computer vision systems on AWS. Built multimodal models, MLOps pipelines, and semantic search systems across academic and production environments.

EDUCATION

University at Buffalo, The State University of New York

Masters of Science, Engineering Science (Data Science)

- GPA: 3.67

Aug 2025 - Dec 2026

Buffalo, NY

Vellore Institute of Technology

Bachelor of Technology, Computer Science and Engineering with Specialization In AI and ML

Aug 2019 - Jul 2023

Chennai, IN

WORK EXPERIENCE

Zolvit (formerly Vakilsearch)

Machine Learning Engineer Intern

Feb 2024 - Aug 2024

Chennai, IN

- Eliminated ~23 hours of weekly manual data entry by engineering a fully automated ingestion pipeline for 350+ complex documents, combining Google Vision OCR with a fine-tuned T5-large model to replace a fragile legacy process.
- Drove 99.9% service uptime and eliminated recurring outages by deploying Dockerized ML microservices on AWS EC2 with FastAPI, restart policies, and real-time monitoring via Grafana.
- Automated routing of customer-uploaded documents across categories by achieving 92% classification accuracy, benchmarking Logistic Regression, SVM, and XGBoost on TF-IDF and Doc2Vec features extracted via Google Vision OCR.
- Improved legal precedent retrieval precision by ~28% for lawyers by building a hybrid search engine combining Elasticsearch keyword search with Pinecone vector embeddings across 10,000+ documents.
- Maintained reliability of production ML systems by handling incident response, debugging failures, and improving robustness of legacy pipelines.

PROJECTS

DRIFT-AWARE MLOps PIPELINE | [GitHub](#)

Feb 2026

- Designed a drift-aware MLOps pipeline to detect silent model degradation and trigger automated retraining, improving model reliability under changing data distributions.
- Built and containerized the system using FastAPI, Airflow, and MLflow with Evidently AI for drift detection and Docker for deployment.
- Monitored system health in real time using Prometheus and Grafana dashboards.
- Validated end-to-end performance by simulating concept drift, triggering automated retraining that recovered +0.1178 ROC-AUC on shifted data.

MULTI-MODAL MOLECULAR SIMILARITY REGRESSION | [GitHub](#)

Jan 2026

- Modeled molecular similarity for drug discovery to improve candidate selection and ranking of chemical compounds.
- Designed a multi-modal architecture fusing 2D image features (ResNet-18), 3D geometry (SchNet), and fingerprint embeddings, trained with contrastive learning across 291,742 molecules.
- Achieved ~0.92 Pearson correlation on 200 expert-annotated pairs, outperforming the Tanimoto similarity baseline on unseen molecular pairs.

PATCH-LEVEL CT TAMPER CLASSIFICATION | [GitHub](#)

Dec 2025

- Built to detect tampering at the patch level within CT scans, where localized manipulation is harder to catch than whole-scan forgeries.
- Jointly trained a 3D convolutional compressor with an ImageNet-pretrained ResNet-18 across 169 volumetric lung CT scans to convert 16-slice 3D patches into compact 2D feature maps.
- Achieved 0.95 validation AUC, outperforming 2.5D, 3D, and projection-based baselines at lower computational cost.

SKILLS

- **Programming Languages:** Python, C/C++, SQL, R
- **Machine Learning & AI:** PyTorch (torchvision, PyTorch Geometric), TensorFlow, Scikit-learn, XGBoost, Hugging Face, embeddings, RAG, Representation Learning (self-supervised & contrastive learning), LangChain, LangGraph, TensorBoard
- **Data Processing & Analysis:** Pandas, NumPy, Medical Imaging (CT, DICOM, pydicom, volumetric data), RDKit
- **Data Visualization & BI:** Matplotlib, Seaborn, Plotly, Streamlit, Tableau, PowerBI
- **MLOps & Cloud:** Docker, Kubernetes, AWS, Apache Airflow, MLflow, Evidently AI, Prometheus, Grafana
- **Developer Tools & Systems:** Linux, Git/GitHub, Pytest, REST APIs, CUDA/GPU training, Coding Agents (Cursor, Claude Code, Codex)
- **Databases & Data Engineering:** PostgreSQL, MySQL, Elasticsearch, Pinecone, Spark, Snowflake

AWARDS & CERTIFICATIONS

- Felix Infausto Scholarship - University at Buffalo, 2025
- Coursera - Probability & Statistics for Machine Learning & Data Science, Machine Learning Specialization, Deep Learning Specialization