Neeraj Solanki

+1-312-375-3980 neerajsolanki2000@gmail.com LinkedIn Profile

Summary

Machine Learning Engineer specializing in **federated learning, LLMs, and cloud-native systems**. Experienced in **software architecture, anomaly detection, and scalable ML pipelines** with deployments on AWS and Kubernetes. Skilled in **Python, C++, PyTorch, TensorFlow, and MLOps**, with research contributions in **energy-efficient and privacy-preserving AI**.

Skills

Programming: Python, C++, Java, SQL
Machine Learning & AI: Deep Learning, Large Language Models (LLMs), Transformer Architectures, GANs, NLP, Model Optimization, Adversarial Robustness, Explainable AI
Frameworks & Libraries: PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face, Pandas, NumPy
Cloud & Deployment: AWS, MLOps, LLMOps, Docker, Kubernetes, CI/CD (Jenkins), Model Serving, Cloud-Native Inference
Data & Storage: MySQL, MongoDB, PostgreSQL

Experience

University of Illinois at Chicago Chicago, IL
Machine learning Researcher *09/2024 – Present*
- Developed **federated learning frameworks** with communication-efficient updates (MSB–LSB, adaptive gradient freezing), cutting overhead by up to **80%**.
- Built **early-exit neural architectures** for adaptive inference on edge devices, boosting energy efficiency without accuracy loss.
- Researched and deployed **Weightless Neural Networks (WNNs)** for lightweight AI in healthcare and signal processing.
- Designed **active learning modules** with uncertainty-based sampling, accelerating model convergence under limited labels.

Deloitte USI Mumbai, India
Machine Learning Engineer 01/2022 – 07/2024
- Designed and deployed **anomaly detection models** (Random Forest, Isolation Forest, Transformers) used in production, achieving **97% detection precision**.
- Built scalable **ML pipelines for real-time log and behavior analytics**, integrating 30+ data sources and reducing detection latency by **40%**.
- Applied **AI-driven threat modeling and root-cause analysis** aligned with MITRE ATT&CK, enhancing fraud detection and security insights.

Happiest Minds Mumbai, India
Machine Learning Intern *03/2021 – 06/2022*
- Developed and deployed **ML models for anomaly detection and predictive analytics**, improving reliability by **30%**.
- Built scalable **data pipelines and feature engineering workflows**, reducing model training time by **25%**.
- Integrated models into production using **REST APIs and Docker**, ensuring scalability and seamless deployment.

Education

University of Illinois at Chicago Chicago, IL
M.S. in Computer Science *08/2024 - 05/2026*

MIT - World Peace University Pune, India
B.Tech in Computer Engineering — CGPA: 8.86 *06/2018 - 06/2022*

Selected Projects

Converso: LLM-Powered Chatbot Platform
- Built a production-grade **customer support chatbot** using LangChain + OpenAI APIs, deployed with Docker and Kubernetes on AWS.
- Implemented retrieval-augmented generation (FAISS vector DB) for context-aware responses; scaled to **50K+ daily queries** with 99.9% uptime.

RecoTrack: Real-Time Recommendation Engine
- Designed a **personalized product recommendation system** using collaborative filtering and XGBoost, deployed via Flask REST APIs.
- Achieved an **18% lift in click-through rate**, integrated seamlessly with React-based front-end for e-commerce personalization.

FraudShield: Transaction Monitoring System
- Developed a streaming fraud detection pipeline using Kafka + PySpark + Random Forest for financial data streams.
- Detected anomalies with **95% accuracy** and ensured sub-200ms response time in a cloud-native AWS environment.

Publications and Leadership

- **ATM-Net:** Adaptive Termination and Multi-Precision Neural Networks for Energy-Harvested Edge Intelligence — *HPCA 2025*
- **SenGuard:** In-Sensor Privacy-Preserving Processing for Smart Imaging — *GLSVLSI 2025*
- **OrganoSense:** Biosignal Neural Processing via Organic Circuits — *MWSCAS 2025*
- Captain of UIC and Deloitte cricket teams with 5 tournament wins; 2-time National Rollball champion; State-level Boxer