Anuj Patel

347-401-7880 | amp10162@nyu.edu | www.panuj.com | linkedin.com/in/panuj

Education

New York University **Sep 2023 – May 2025**

Master of Science in Electrical Engineering (GPA: 3.9/4.0) *New York, NY*

- Coursework: Machine Learning, Deep Learning, Big Data, Probability & Stochastic Processes, High Performance Machine Learning, Computer Vision, Wireless Communications, Digital Signal Processing, Internet Architecture & Protocols
- Graduate Teaching Assistant: Fundamentals of Communication Theory

Vellore Institute of Technology **Jul 2019 – May 2023**

Bachelor of Technology in Electronics and Communication Engineering (GPA: 9.16/10) *Vellore, India*

Professional Experience

NYU WIRELESS **Jan 2025 – May 2025**

Graduate Researcher *New York, NY*

- Designed and implemented an **mmWave channel sounder at 57.51 GHz** using AMD Xilinx RFSoC4x2 and Sivers EVK06002 with BFM06010 RF Module, **enabling high-fidelity** wireless channel measurements.
- Integrated **automated measurement capabilities** using an optimized robotic system and a customized rotating turntable, **streamlining data collection** for large-scale mmWave propagation studies.

Johnson & Johnson **Jun 2024 – Aug 2024**

Data Science Intern *New Brunswick, NJ*

- Developed and deployed **Random Forest** and **XGBoost** models on **AWS SageMaker**, achieving an R^2 **value of 0.92** and a **17% improvement** in model accuracy over baseline methods.
- Built an **end-to-end MLOps pipeline** leveraging **Kubernetes, Docker**, and **CI/CD automation** for scalable model training, deployment, and monitoring in a production-ready environment.
- Optimized model performance through **hyperparameter tuning** and **ensemble learning**, **boosting predictive accuracy by 40%** on a multimodal dataset with electrical, experimental, and thermographic surgery parameters.

Indian Space Research Organization (SAC) **Dec 2022 – May 2023**

Machine Learning Researcher – Quantum Communications *Ahmedabad, India*

- Engineered ML models to **automate polarization component characterization** for a QKD testbed, **reducing manual intervention by 85%**, leveraging Python and C++ for data processing and model implementation.
- Led and deployed **signal processing algorithms** for a free-space BB84 QKD system using NI USRP, achieving a **bit error rate (BER) of 10E-10** and a **correlation of 0.8** for secure quantum communication.

Indian Oil Corporation Ltd. **May 2022 – Jul 2022**

Network Intern *Vadodara, India*

- **Reduced routing speeds by 20%** for MPLS and SD-WAN networks **by identifying bottlenecks** via a smart data center survey, **conducting A/B testing** on routing strategies, and exploring alternate paths.

Technical Skills

Programming: Python, R, SQL, Pandas, NumPy, MATLAB, C, C++, CUDA, Java, Bash, Verilog, VHDL
AI/ML: TensorFlow, PyTorch, Scikit-Learn, MLFlow, NLP, **LLM**(Hugging Face, LangChain, LangGraph, Transformers)
Data & Analytics: Tableau, PowerBI, D3.js, matplotlib, seaborn, plotly, ggplot, **Big Data**(Hadoop, PySpark, Hive)
Cloud: **AWS**(Sagemaker, EC2, Lambda, S3, ELB, Redshift), **GCP**(Vertex AI, BigQuery, AutoML)
Tools and DevOps: Jupyter Notebook, Apache Airflow, Apache Kafka, Apache Spark, NoSQL, PostgreSQL, CosmosDB, MongoDB, MLOps, Databricks, Git/Gitlab, CI/CD, Kubernetes, Docker, ETL, MS Office

Projects

Efficient Federated Learning using Gradient Pruning and Adaptive Methods | *PyTorch* **Sep 2024 - Dec 2024**

- Pioneered an efficient FL framework with gradient-based pruning and adaptive federated optimization to mitigate I/O overheads, **cutting training time by 22%** and **boosting model generalization** in distributed systems.
- **Increased bandwidth efficiency by 143%** using gradient compression techniques and mix-precision training, ensuring accuracy on ResNet models with PyTorch DDP, DeepSpeed and Hugging Face Accelerate.

Medical Chatbot with RAG Architecture | *Llama-3, Hugging Face, LangChain* **Sep 2024 - Dec 2024**

- **Developed a Retrieval-Augmented Generation (RAG) pipeline** for a Medical Chatbot by integrating finetuned Llama-3, Llama-2, Gemma 1.1, Mistral-7B, and DistilGPT2 with **LangChain**.
- Leveraged **Vector DBs** such as **Weaviate** to index and retrieve medical knowledge from the **PubMedQA** dataset, ensuring accurate and relevant context.
- Achieved a **precision of 88.3%** and a **BERT score of 0.87** while maintaining memory efficiency at **4.1GB**, optimized for **low-cost compute environments**.

Transformer-Based Multi-Modal Emotion Recognition System | *PyTorch, OpenCV, HPC* **Sep 2024 - Dec 2024**

- Enhanced a transformer based framework for emotion recognition, **achieving 33.96% top-1 and 98.13% top-5 precision** on RAVDESS data integrating both facial and vocal cues.
- Applied **advanced modality fusion techniques** with feature extraction and preprocessing pipelines, **processing 4,000+ video and audio signals to improve robustness** in noisy/incomplete datasets.

Movie Recommendation System with NCF | *Python, PyTorch, PySpark, SQL* **Sep 2024 - Dec 2024**

- Built and deployed a scalable movie recommendation system using NCF, **achieving 50% Hit Ratio** on MovieLens 1M with **distributed Spark processing and SQL integration** for data warehousing.
- Streamlined a 1M-record data pipeline with preprocessing, negative sampling, SQL-driven backend, and a Streamlit frontend, **reducing recommendation retrieval time by 33%.**