Anuj Patel

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

EDUCATION

New York University		Sep 2023 – May 2025 New York, NY
Coursework: Ml	L. DL, High Performance Machine Learning, Probability & Stochastic Processes	, CV, Big Data
• Graduate Teach	ing Assistant: Fundamentals of Communication Theory	, , , ,
Vellore Institute of Technology Bachelor of Technology in Electronics and Communication Engineering (GPA: 9.16/10)		Jul 2019 – May 2023 Vellore, India
WORK EXPE	RIENCE	
Data Science Intern Johnson & Johnson		Jun 2024 – Aug 2024 New Brunswick, NJ
• Built and deploy engineering a N	red Random Forest and XGBoost models on AWS SageMaker (achieving ILOps pipeline with Kubernetes, Docker, and CI/CD automation.	an R^2 of 0.92 and 17% lift), while
• Enhanced mode learning .	l accuracy by 40% on a multimodal surgical dataset via strategic hyperparam	eter tuning and ensemble
Machine Learning Researcher – Quantum Communications Indian Space Research Organization (SAC)		Dec 2022 – May 2023 Ahmedabad, India
• Engineered ML intervention b	models in Python and C++ to optimize polarization characterization for a QKI y 85% .	D testbed, cutting manual
• Deployed signal correlation .	processing algorithms for a free-space BB84 QKD system via NI USRP, ${\bf achiev}$	ring a BER of 10E-10 and 0.8
Network Intern Indian Oil Corporation Ltd.		May 2022 – Jul 2022 Vadodara, India
• Reduced rout exploring altern	ing speeds by 20% for MPLS and SD-WAN networks by identifying bottlened ate paths.	ks, conducting A/B testing, and
SKILLS		
Programming AI/ML Data Analytics Cloud DevOps Databases	Python, R, SQL, Pandas, NumPy, MATLAB, C/C++, CUDA, Java, Bash, Verilog, VHDL TensorFlow, PyTorch, Scikit-Learn, NLP, LLM (Hugging Face, LangChain, Transformers) Tableau, PowerBI, D3.js, matplotlib, seaborn, plotly, ggplot, Hadoop, PySpark, Hive AWS (Sagemaker, EC2, ELB, S3, Redshift), GCP (Vertex AI, BigQuery, AutoML) Kubeflow, Airflow, Kafka, Spark, Git, CI/CD, Kubernetes, Docker, ETL PostgreSQL, CosmosDB, MongoDB, ChromaDB	
PROJECTS		
Efficient Federated • Pioneered an eff 22 % and boosti	Learning using Gradient Pruning and Adaptive Methods PyTorch icient FL framework with gradient pruning and adaptive federated optimization ng model generalization.	Sep 2024 - Dec 2024 a, reducing training time by

• Increased **bandwidth efficiency by 143%** via gradient compression and mix-precision training, validating ResNet accuracy with PyTorch DDP and DeepSpeed and Hugging Face Accelerate.

Sep 2024 - Dec 2024

Sep 2024 - Dec 2024

• Developed a RAG pipeline integrating finetuned LLMs (Llama-3, Llama-2, Gemma 1.1, Mistral-7B, DistilGPT2) with LangChain and leveraging Weaviate to index/retrieve data from PubMedQA.

• Achieved a precision of 88.3% and BERT score of 0.87, with memory efficiency at 4.1GB for low-cost compute.

Movie Recommendation System with NCF | Python, PyTorch, PySpark, SQL

Medical Chatbot with RAG Architecture | Llama-3, Hugging Face, LangChain, HPC

- Built a scalable movie recommender using NCF, achieving 50% Hit Ratio on MovieLens 1M with distributed Spark processing and SQL data warehousing.
- Optimized a 1M-record pipeline with preprocessing, negative sampling, SQL backend, and a Streamlit frontend, reducing retrieval time by 33%.