

ANUJ PATEL

New York, NY | 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

- **Co-authored** a textbook on “**Fundamentals of Communication Theory**” with **Dr. Unnikrishna Pillai**.
- Coursework: ML, Deep Learning, CV, High Performance Machine Learning, Probability, Big Data
- Research: Developed a mmWave channel sounder at 57.51 GHz under **Dr. Sundeep Rangan** for wireless channel measurements.

Vellore Institute of Technology

Jul 2019 – May 2023

Bachelor of Technology in Electronics and Communication Engineering (GPA: 9.2/10, Rank: 4)

Vellore, India

- Coursework: Applied Linear Algebra, Statistics, Cryptography and Network Security, Object Oriented Programming

WORK EXPERIENCE

New York University

Jan 2024 – Dec 2024

Artificial Intelligence Engineer

New York, NY

- Adapted and deployed a **RAG-based GenAI assistant** using **LangChain**, **OpenAI APIs**, and **Pinecone**, enabling natural language search over 100+ academic and policy documents.
- Engineered modular retrieval pipelines with **text-embedding-ada-002**, **Docker**, and **GitHub Actions**, achieving **sub-300ms median latency** and readiness for seamless internal rollout.
- Collaborated with product teams to develop a course recommendation engine using **TF-IDF** and **user embeddings**, deployed via **FastAPI** + **PostgreSQL** for 5K+ students.

Johnson & Johnson

Jun 2024 – Aug 2024

Data Science Intern

New Brunswick, NJ

- Led development of an **LLM-powered clinical assistant** using **LLaMA 2-7B** with **LoRA fine-tuning** over 1M+ anonymized patient records—reducing physician query time by **46%** across **450+ daily clinical queries** during pilot phase.
- Built and productionized a **real-time, multimodal ML pipeline** (text, imaging, vitals) using **PyTorch**, **HuggingFace** and **AWS**—achieved **28% lift in outcome prediction accuracy** and scaled to serve 20M+ patient records.

Indian Space Research Organization

Dec 2022 – May 2023

Artificial Intelligence Researcher

Ahmedabad, India

- Trained and deployed a **GAN-based super-resolution model** on RISAT-1A SAR data, boosting spatial resolution **2×** while retaining speckle-aware texture priors for terrain analysis.
- Integrated and optimized **quantized CNNs** in an Edge AI framework for real-time cloud detection in Microsat’s onboard inference pipeline, reducing transmission data by **67%** under compute and power limits.

PROJECTS

Multi-Agent RAG Chatbot for E-commerce Customer Support

Aug 2025 - Sep 2025

- Designed an **agentic RAG chatbot** with **7 specialized tools** and **Voyage-3.5 embeddings**, orchestrating **dual-database queries** (SQLite FTS5 + ChromaDB) to process **20,000+ parts** with **2 LLM calls** for cost optimization.
- Architected **multi-agent workflow** with **RAGOrchestrator**, implementing **intent classification** and **semantic retrieval** via **DeepSeek API** for real-time appliance troubleshooting and parts compatibility.

Efficient Federated Learning using Gradient Pruning and Adaptive Methods

Sep 2024 - Dec 2024

- Pioneered an efficient FL framework with gradient pruning and adaptive federated optimization, **reducing training time by 22%** and boosting model generalization.
- Increased **bandwidth efficiency by 143%** via gradient compression and mix-precision training, validating ResNet accuracy with **PyTorch DDP**, **DeepSpeed** and **Hugging Face Accelerate**.

Transformer-Based Multi-Modal Emotion Recognition System

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.

SKILLS

Languages

Python, TypeScript, C/C++, Golang, CUDA, MATLAB, SQL, Bash

Frameworks

PyTorch, HuggingFace, LangChain, LangGraph, GraphQL, Numpy, Pandas, Wandb

Cloud

AWS (SageMaker, EC2, ELB, S3, Redshift), GCP (Vertex AI, BigQuery, AutoML)

DevOps

Kubeflow, Airflow, Spark, Kafka, Kubernetes, Docker, CI/CD, Git, Slurm

Databases

PostgreSQL, MongoDB, SQLite, ChromaDB, Weaviate, Pinecone