

# ANUJ PATEL

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

## SUMMARY

Curious how YouTube “knew me”, I dove into it’s recommendation systems — and that curiosity grew into a career (highly passion driven, indeed!). With 3+ Y.O.E, I’ve built ML systems powering GenAI tools, RecSys engines, and MLOps pipelines — driving decisions across EdTech, space-tech, and healthcare.

## WORK EXPERIENCE

### NYU WIRELESS

Jan 2025 – May 2025

Applied Researcher (Advisor: Prof. Sundeeep Rangan)

New York, NY

- Led the development of a **mmWave channel sounder** at **57.51 GHz** using AMD Xilinx RFSoC4x2 and Sivers RF Modules for high-fidelity wireless channel measurements.
- Automated large-scale data acquisition via a custom **rotating turntable**, accelerating 5G propagation studies.

### New York University

Jan 2024 – Dec 2024

Machine Learning Engineer

New York, NY

- Built 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 low-latency responses and readiness for seamless internal rollout.
- Developed and productionized a personalized course recommendation engine using **TF-IDF**, **cosine similarity**, and **user embeddings**; served via **FastAPI** + **PostgreSQL** for 5K+ students.

### Johnson & Johnson

Jun 2024 – Aug 2024

Data Scientist

New Brunswick, NJ

- Led development of an **LLM-powered clinical assistant** fine-tuned on 10M+ patient records and medical dialogues; reduced physician query time by **45%** and influenced \$12M+ in operational savings.
- Built and productionized a **real-time, multimodal ML pipeline** (text, imaging, vitals) using **PyTorch**, **HuggingFace**, and **AWS**—achieved **40% lift in outcome prediction accuracy** and scaled to serve 20M+ patient records.

### Indian Space Research Organization

Dec 2022 – May 2023

Machine Learning 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.
- Built and optimized **quantized CNNs** in an Edge AI framework for real-time cloud detection in Microsat’s onboard inference pipeline, reducing transmission data by **50%** under compute and power limits.

### VIT University

Aug 2020 – May 2022

Software Engineer

Vellore, India

- Maintained and scaled the VITian app (100K+ downloads, 25K+ MAUs) using **React Native**, **FastAPI**, **PostgreSQL**, and **GCP**; implemented user analytics and backend data pipelines to support ML-driven personalization.
- Led system design and CI/CD setup using **Docker**, **GitHub Actions**, and **Firestore**, improving API latency by **60%** and achieving **75% crash-free sessions**; collaborated cross-functionally to ship high-impact features at scale.

## 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 Prof. Unnikrishna Pillai
- Coursework: ML, Deep Learning, CV, High Performance Machine Learning, Probability, Big Data

### Vellore Institute of Technology

Jul 2019 – May 2023

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

Vellore, India

## SKILLS

|            |   |
|------------|---|
| Languages  | Python, TypeScript, C/C++, Golang, CUDA, MATLAB, SQL, Bash                  |
| Frameworks | PyTorch, HuggingFace, LangChain, LangGraph, GraphQL, Numpy, Pandas, W&B     |
| Cloud      | AWS (SageMaker, EC2, ELB, S3, Redshift), GCP (Vertex AI, BigQuery, AutoML)  |
| DevOps     | Kubeflow, Airflow, Spark, Kafka, Kubernetes, Docker, CI/CD, Git, Slurm, ETL |
| Databases  | PostgreSQL, MongoDB, CosmosDB, ChromaDB, Weaviate, Pinecone                 |

## PROJECTS

### Efficient Federated Learning using Gradient Pruning and Adaptive Methods | *PyTorch*

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 and DeepSpeed and Hugging Face Accelerate.

### 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.