

# REET NANDY

reet.nandy@nyu.edu | reetnandy.com | github.com/techpertz | linkedin.com/in/reetnandy | +1 (518) 930-6116

## EDUCATION

- New York University** – Masters in Computer Science (MS) May 2025
- *Relevant Coursework:* Data Structures and Algorithms, Cloud Computing, Machine Learning, Big Data Analytics, Database Systems, Software Engineering, Object-oriented Design, Probability and Statistics
  - *Graduate Teaching Assistant:* CS GY 6233 - Operating System, CSCI UA 0310 - Algorithms
- Manipal University Jaipur** – Bachelors in Computer Science (B. Tech) May 2023

## SKILLS

**Frontend + Backend:** Python (FastAPI, Django), TypeScript (Next.js), Tailwind CSS, Node.js (Express.js), Java (Spring Boot)  
**AI / LLM:** Transformer (GPT, BERT, LLaMA), Graph + RAG, Fine-tuning, Embeddings, Quantization, Multi-Agent  
**Cloud + DevOps:** AWS (EC2, Lambda, S3, SageMaker, Prometheus), GCP, Kubernetes, Docker, CI/CD, ELK, Microservices  
**Data + Pipeline:** SQL (PostgreSQL), NoSQL (MongoDB), Vector DBs, Redis, Kafka, Airflow, ETL / ELT, REST, gRPC, Agile

## EXPERIENCE

- Backend Development Intern (ML)** June 2024 – December 2024  
Mobility Intelligence, New York City, USA *React, FastAPI, PyTorch, PostgreSQL, Airflow*
- Engineered regression + time series model with Kalman filtering, reducing forecast errors to 5% for 90-day horizons.
  - Delivered FastAPI-ReactJS app enabling real-time ML model integration, delivering predictions with sub-1s latency.
  - Constructed ETL pipelines for batch processing 15M+ daily records saving 3 man hours per dataset.
- Software Engineering Intern (R&D)** January 2023 – June 2023  
Defence Research & Development Organisation, India *Python, PyQt, Redis, PostgreSQL, Airflow, Docker, SciPy*
- Designed PyQt desktop app with concurrent LiDAR data processing for real-time 3D mapping (97% accuracy).
  - Developed streaming ETL pipeline with zero-copy buffers and SciPy transforms, reducing calibration errors by 70%.
  - Built Redis-PostgreSQL hybrid retrieval using geospatial indexing to reduce GPS API latency from 1sec to 150ms.
- Software Engineering Intern** April 2022 – December 2022  
Solar Industries India Ltd, India *Python, AWS, Django, Kafka, Redis, SQL, Docker*
- Directed automation initiative with 5 interns for 25+ plants, deploying 5 Django projects totaling 75K+ lines of code.
  - Pioneered Kafka pipeline for high-throughput data processing, achieving 2.5M rows/sec with 100K+ daily requests.
  - Optimized workflows across 5 cross-functional teams by implementing distributed caching, cutting API latency by 20%

## PROJECTS

- [AI Agent] GraphRAG - LLM Document Compliance (Github)** *Next.js, FastAPI, GraphRAG, Anthropic, FAISS, Neo4J, SQL*
- Launched an Agentic SaaS with real-time document edit, approval and audit reports via Graph based RAG and LLM.
  - Implemented a GraphRAG and PDF parser from scratch for unstructured PDFs using PDFMiner & Tesseract (OCR).
  - Executed semantic chunking + NLP NER + BART-CNN summarization, achieving 70% relation extraction accuracy.
  - Synthesized hybrid retrieval (Vector + Graph + metadata) boosting compliance accuracy to 90%.
- [AWS / Fullstack] AI-Fitness Analytics Dashboard (Github)** *AWS, ReactJS, Django, PostgreSQL, DynamoDB, KNN*
- Deployed Django API on Elastic Beanstalk with Google Fit integration, configured auto-scaling groups + health probes.
  - Orchestrated pipelines with Lambda-SageMaker, deploying KNN models for recommendations at 92% accuracy.
  - Coordinated event-driven metrics processing via SNS/SQS, achieving 800ms p95 latency for real-time health data.
- [Python / Core] Hierarchical Vector Database from Scratch (Github)** *FastAPI, Pydantic, Cohere, AsyncIO, Kubernetes*
- Built embedding database (library - document - chunk) w/ async collection mutexes; 12K ops/sec at <0.1% conflicts.
  - Added 3 indexing algorithms (LinearScan/KD-Tree/LSH) for vector search on 10M vectors in 18ms.
  - Led Kubernetes Helm deployment along with custom made CLI toolkit, reducing onboarding complexity by 100%.
- [AI Workflow] Multi-Platform LLM Task Orchestrator (Github)** *Flask, LLM, RAG, NoSQL, Vector DB, GOOGLE API*
- Created Flask API with Gmail/Slack/Trello via OAuth2, automating task extraction with OpenAI priority classification.
  - Configured MongoDB RAG system with vector indexing for context memory management via chat history tracking.
  - Automated cross-platform workflow triggers using LLM decision trees, reducing manual task management by 60%