

# Vraj Patel

vjpatel4423@gmail.com — +1-631-780-2456 — Stony Brook, NY —  LinkedIn —  GitHub —  Portfolio —  DevPost

## Education

**Stony Brook University** – *M.S. in Data Science and Minor in Quantitative Finance*

May 2027

**Pandit Deendayal Energy University** – *B.Tech in Computer Engineering (CGPA: 8.72/10)*

May 2025

## Technical Skills

**Languages:** Python, SQL, R, C++, C, Java, JavaScript/TypeScript, Go, HTML/CSS, Bash

**Cloud/DevOps:** AWS (EC2, S3, Lambda), GCP (Cloud Run), Docker, Git, CI/CD, Cloudflare

**Reporting & Visualization:** Data Storytelling, Business Process Mapping, Process Flow Diagramming

**Frameworks:** FastAPI, React, Flask, Node.js, Streamlit, REST APIs, WebSockets

**Big Data & Databases:** Spark, Hadoop, Kafka, Airflow, PostgreSQL, MySQL, MongoDB, Supabase, Redis

**Data/ML:** PyTorch, TensorFlow, Pandas, NumPy, Scikit-learn, SpaCy, OpenCV, Tableau, Excel


## Featured Projects

**The Lot Lab** — **1st Place, NVIDIA Spark Hack NYC**

Apr 2026

*Python, RAPIDS, NVIDIA cuOpt, FastAPI, React, MapLibre GL, NemoTron, Flux, OpenClaw*

- Built an on-device urban planning simulator ingesting 15+ NYC open datasets using RAPIDS cuDF/cuGraph/cuML for GPU-accelerated preprocessing, lot scoring, and graph-based accessibility analysis.
- Designed district-level planning with NVIDIA cuOpt, producing population-weighted multi-lot recommendations across 12 human & environmental facility types for NYC vacant parcels.
- Developed a local multi-stage pipeline with zoning RAG, cuML impact modeling, NemoTron-driven recommendations, and Flux-based concept rendering, packaged as an OpenClaw skill for civic planning workflows.

**SBU VibeCheck – Real-Time Event Aggregation Platform** —  **Website**

Nov 2025

*React, TypeScript, FastAPI, PostgreSQL, ChromaDB, OpenAI (gpt-4o), WebSockets, Auth0, GCP*

- Winner of SBU Hackathon: Architected a full-stack event platform tracking live sentiment and active RSVPs, deploying the FastAPI backend on auto-scaling GCP Cloud Run.
- Precision RAG Chatbot: Engineered a time-aware assistant using OpenAI gpt-4o and function calling to dynamically translate natural language into SQL filters and semantic searches over a ChromaDB vector store.
- Automated AI Pipelines: Developed async Python scrapers to parse nested RSS/HTML feeds into PostgreSQL. Leveraged batched LLM inference to autonomously generate contextual polling questions, reducing API costs by ~90%.
- Real-Time Sync & Auth: Integrated WebSockets for sub-200ms live poll synchronization and designed a secure OAuth 2.0 flow via Auth0 for strict .edu credential verification.

**CampusPool (WolfiePool)** —  **GitHub** —  **Website**

Feb 2026

*React, TypeScript, FastAPI, Supabase, Gemini API, ElevenLabs, Mapbox, Uber API*

- Architected a geospatial clustering engine matching users by directional bearing, proximity, and time windows to dynamically calculate distance-weighted Uber fare splits.
- Integrated Google Gemini 2.5 Flash API to power a context-aware chatbot, predictive ride-matching probabilities, and personalized weekly insights tracking CO2 reductions.
- Developed a full-stack platform using React, Mapbox, and FastAPI with Supabase Realtime for live chat, incorporating ElevenLabs TTS for voice-based ride alerts.

## Experience

**FinTech Data Scientist Intern, Flits (India)**

Feb 2025 – May 2025

- Built a real-time analytics engine processing NSE/BSE tick data for 200+ contracts via an async Websocket client with auto-reconnection, reducing manual monitoring by 40%; engineered ETL pipelines to parse a 170K-row instrument file into a normalized PostgreSQL schema with 30+ indexed categories for downstream BI.
- Designed process flow diagrams for end-to-end data ingestion, communicating functional requirements via technical documentation to align data workflows across engineering teams; deployed Streamlit & Plotly dashboards with 30s polling, request throttling, and connection pooling ensuring 100% uptime against exchange API rate limits.

**Data Analyst Intern, YHONK – Noise Pollution Mitigation**

May 2024 – Jun 2024

- Executed geospatial analysis on 2.6M GPS-tagged events using PostGIS and QGIS R-tree spatial indexing to compute violation hotspots with  $O(\log n)$  lookup efficiency; automated violation detection via ST\_Within spatial joins for law enforcement reporting.
- Built a distributed scraping pipeline via a 4-node Selenium Grid with rotating proxies and exponential backoff, extracting 50K+ school records for noise zone mapping; developed Tableau dashboards to visually narrate complex geospatial findings for stakeholder presentations.

**International Research Intern, AIT Brain Lab (Thailand)** —  **Website**

Jul 2023 – Aug 2023

- Fine-tuned T5-base & FactorSum transformers on 350K scientific paper pairs, achieving a 0.42 ROUGE-L score for abstractive summarization via transfer learning on V100 GPUs; built a Flask REST API with Celery task queue and request batching, reducing P95 latency from 4.2s to 1.8s.
- Integrated SpaCy NER and NLTK for preprocessing with PyPDF2 PDF extraction (OCR fallback) and MathJax rendering.

## Certifications

**Azure Administrator Associate, AWS Certified Cloud Practitioner, Alteryx Designer Core**