Pranav Rebala

U.S Citizen - Jersey City, New Jersey
+1 (617) 987 1268 | navrebala@gmail.com | Github | Linkedin | Portfolio

PROFESSIONAL SUMMARY

Gen AI/ML Developer with 3+ years building production-ready software and model pipelines. Expert in Python, React, and deploying generative models (vision text) end-to-end. Strong at shipping robust APIs, reproducible CI/CD workflows, containerized services, and developer-facing tooling that enable data-driven products and creative applications

TECHNICAL SKILLS

Languages: Python, JavaScript, SQL, R
Generative AI/ML: PyTorch, Hugging Face, Diffusers, Transformers, DECA, PIFuHD
Frameworks: React, Node.js, Express, FastAPI, Flask, Tailwind CSS
Backend / APIs: Flask, FastAPI, REST, Sockets, Kore.AI
Cloud/DevOps: Azure (Functions, SQL), AWS EC2, Docker, Git, Bitbucket
Tools: Postman, Jupyter, Pandas, NumPy, Tableau, Jira, VS Code
Development & Tooling: Unit Testing, Dataset Versioning, Linux, Git

WORK EXPERIENCE

Forfend Cybernatics | **Software Developer** *Nov 2023 – Jul 2025*

- Designed and deployed a **LightFM-based** hybrid recommendation model within a **FastAPI** microservice (**Dockerized on AWS with RDS backend**), predicting vehicle preferences from historical sales and behavioral data; handled **20,000** monthly kiosk and sales requests with **60%** higher engagement and **22%** higher conversion rates
- Implemented a **Generative AI**-powered conversational chatbot with **intent classification** and **LangChain**-based **RAG** retrievers for appointment booking, inventory queries, and expert assistance in automobile dealerships; reducing customer support operational expenses by **50%**
- Developed an **Agentic AI** system for shop-floor management, integrating **time-series forecasting** (**Prophet/XGBoost**) and **RL-based** bay allocation to monitor live queue congestion across multiple service centers. Orchestrated decisions via a **LangGraph** agent and delivered insights through a **live analytics dashboard** with predictive workload optimization, improving **operational efficiency** by **60%**

Wipro Ltd | **Full Stack Intern** *Jul 2023 – Oct 2023*

- Led a **3-person** team to design and ship real-time dashboards accessing **Azure**-hosted financial data for a **230,000+** employee vertical, cutting reporting lag from **2 days** to **2 hours**
- Implemented **role-based access controls** across 6 groups (1,200 users), reducing access-request tickets **40%** and provisioning time from **3 days** to **8 hours**
- Automated **ETL/ingestion workflows** and stabilized pipelines (**Python**), cutting manual reporting time **40%** and accelerating decision speed **30%**

Flutura Business Solutions | **Data Science Intern** *Jun 2022 – Aug 2022*

- Analyzed **drone telemetry** and optimized battery usage, yielding simulation models with **25%** better efficiency versus baseline
- Built **predictive maintenance models** from sensor data with **85%** precision, lowering pilot maintenance costs **20%**
- Evaluated and selected the best-fit **AR/VR** diagnostic platform (**Vuforia**) from **10+** tools, integrating with **Unity** to enhance engineering troubleshooting capabilities by **35%**

PROJECTS

Football Base | **Python, OpenAI, LangChain, FAISS, FastAPI, PyTorch** *Oct 2025*

- **Architected** and deployed a **production ready LLM driven conversational analytics platform** using **Next.js, FastAPI, LangChain** and **OpenAI** that supports **real time streaming queries** and **contextual follow ups**
- Implemented **RAG** and **FAISS** based **semantic search** over **8,900+ player** season records with **OpenAI embeddings** to enable **fuzzy name resolution**, **intent parsing** and **high accuracy retrieval**
- Built **scalable data** and **visualization pipelines** using **pandas, NumPy** and **matplotlib** to generate **time series, comparison** and **radar charts** as **base64 images** for fast **multi player analysis**
- Engineered **MLOps** and **caching** with **batch embedding workflows**, an **SQLite agent cache** and **environment driven config** to ensure **low latency, repeatable evaluation** and simple **cloud deployment**
- Built a **React/Next.js** frontend with **TypeScript** and **Tailwind, WebSocket** real-time chat + image viewer, deployed on **Netlify**

CharGen: Synthesizing 3D Faces and Bodies from Text | **Flask, React, MongoDB, Python** *May 2024*

- Engineered a prompt-to-3D pipeline integrating **DECA** and **PIFuHD** to produce configurable 3D faces and bodies from attribute inputs; built **REST APIs** and a **React** frontend for user flows
- Added **dataset/versioning**, **unit-tested preprocessing**, and **CI** pipelines to ensure reproducible experiments and traceable results; published methodology in **IEEE**

Image Caption Generator | **Python, CNN, LSTM, Flask** ***GitHub*** *Apr 2023*

- Implemented a deployable image→text pipeline combining **CNN** feature extraction with sequence decoders, optimized for **inference efficiency** and checkpointed model releases
- Served batch and real-time inference via lightweight **Flask endpoints** and production-ready inference scripts

Hadoop MapReduce Simulator| **Python, Socket Programming** | | ***Github*** *Mar 2023*

- Simulated Hadoop-style distributed system in Python with client-server architecture; enabled parallel MapReduce operations over partitioned datasets across multiple nodes.
- Demonstrated 2.5x faster execution time compared to baseline script when tested on large text data; validated system logic for education/demo purposes.

LightweightMMM Attribution Pipeline | **LightweightMMM, JAX, Python, SQL, Tableau** | ***GitHub*** *Aug 2025*

- Built a **LightweightMMM** pipeline leveraging **Meta's Robyn dataset** plus Google Ads and Analytics exports to estimate **channel elasticities**, **adstock lags**, and seasonality and produce posterior predictive estimates and **marginal ROI** curves
- Exported parameterized scenario grids and hero mixes for interactive Tableau decisioning, validated with **rolling origin backtests** and **posterior predictive checks**, and delivered a pilot reallocation with an estimated **9% incremental revenue uplift**

Socket Chatbot | **Python, Socket Programming, Regex** | ***GitHub*** *Mar 2023*

- Built a **socket-based client to server chatbot** prototype with **real-time agent handoff** and server orchestration across customer, agent, and server terminals
- Implemented **regex-driven intent matching** and resilient **handoff logic** to demonstrate escalation flows for customer support

Hadoop MapReduce | **Python, Hadoop, Distributed Systems** | ***GitHub*** *Nov 2023*

- Implemented a **multi-node MapReduce orchestration** demonstrating partitioned writes, shuffling, and distributed reduce for scalable **word-count ETL** workflows
- Designed **client**, **master**, and **worker** components to illustrate **parallel execution** and **data partitioning** fundamentals

Optimal Team - Player Selection & Squad Optimization | **Python, Pandas, StatsModels** | ***GitHub*** *Mar 2023*

- Built a **player-rating pipeline** using ball-by-ball and match-level data to score **450+ players** and generate optimal playing elevens via **constrained optimization**; validated selections with season-level backtests
- Produced **visualization-ready outputs** and an **executive memo** for selection strategy; published methodology in **IEEE**

EDUCATION

Babson College | **M.S. in Business Analytics** *May 2025*

Relevant Coursework: Machine Learning in Business(R), Storytelling with Data, Managerial Economics, Entrepreneurship Analytics

PES University | **B.S. Computer Science, Machine Intelligence & Data Science** *May 2024*

Relevant Coursework: Database Management System, Data Structures and Its Applications, Engineering Mathematics I II, Object Oriented Analysis Design with Java, Cloud Computing, Blockchain, Compiler Design, Software Engineering, Linear Algebra, Python for Computational Problem Solving, Operating Systems, Machine Intelligence, Data Analytics, Database Technologies, Big Data, Design and Analysis of Algorithms, Computer Networks, Research Methodology, Software Testing, Design of IoT Solutions

PUBLICATIONS

Synthesizing 3D Faces and Bodies from Text — IEEE 9th I2CT *2024*

Optimal IPL Playing 11 Team Selection — IEEE 8th I2CT *2023*

CERTIFICATIONS

AWS Educate: Introduction to Cloud 101 *2023*

PCEP – Certified Entry-Level Python Programmer *2022*

PERSONAL DETAILS

Gender: Male | Ethnicity: South Asian | US Citizen | Do not Require Sponsorship