

Prasanth Muntha

New York, NY | pmuntha@mail.yu.edu | (551) 225-4184 | LinkedIn | Portfolio

Summary

Graduate student in Artificial Intelligence with a strong builder mindset and experience developing end-to-end AI systems, from data ingestion to deployment. Comfortable working across ML, backend, and experimentation in fast-paced environments.

Education

Yeshiva University

Master of Science in Artificial Intelligence

New York, NY

Exp. Dec 2026

JNTU, India

Bachelor of Science in Computer Science

Hyderabad, India

May 2023

Technical Skills

Languages: Python, SQL, JavaScript, Bash

Machine Learning & AI: Machine Learning, Deep Learning, NLP, Computer Vision, BERT, LLMs, RAG, LangChain

ML & CV Frameworks: PyVista, VTK, MoveNet, ONNX Runtime, OpenCV, YOLOv8, Scikit-learn, Pandas

Data Engineering: Apache Airflow, AWS Glue, Athena, S3, ETL Pipelines

Backend & APIs: FastAPI, Flask, REST APIs, GraphQL, Streamlit

Cloud & DevOps: AWS, Docker, Git

Work Experience

Aloha Technologies

Software Developer

India

2022 – 2025

- **3D Scoliosis Detection System:** Built a production-grade 3D AI inference pipeline integrating geometric computer vision with deep learning to automate scoliosis diagnosis from raw 3D body scans.
- Designed a mesh rectification algorithm using Oriented Bounding Boxes and K-Means clustering to correct scan misalignment and eliminate manual preprocessing.
- Implemented semantic orientation detection with ONNX Runtime + MoveNet, achieving **99% accuracy** in anatomical view classification.
- Developed ROI segmentation using statistical regression to suppress surface noise and improve Cobb Angle measurement accuracy.
- **AWS Airflow Data Migration:** Migrated legacy Pentaho workflows to Amazon MWAA, refactoring monolithic ETL jobs into modular Python-based DAGs.
- Improved pipeline reliability and fault tolerance, reducing runtimes by **40%**.
- **High-Throughput Data Ingestion:** Built ingestion pipelines processing millions of XML and CSV records daily with **sub-200ms latency**.
- Developed a real-time GraphQL monitoring dashboard, reducing engineering support overhead.

Projects

Glacix — AI Chatbot

Nuxt 4, Vue 3, TypeScript, Vercel AI SDK, PostgreSQL, Drizzle, NVIDIA NIM

- Built a full-stack AI chatbot with streaming chat, multi-model support (NVIDIA NIM), and RAG over uploaded documents (PDF, CSV, Excel, Word).
- Implemented live web search with Tavily API and citation links; integrated per-user token usage tracking in PostgreSQL with a REST API for analytics.
- Shipped with GitHub OAuth, anonymous sessions, S3/Vercel Blob file uploads, and deployment on Vercel with Neon PostgreSQL.

GitHub Codebase Decoder

Python, LLMs, Static Analysis, RAG, AWS

- Build a system to automatically analyze, summarize, and explain large codebases using AST parsing, embeddings, and retrieval-augmented generation.
- Design the system to help researchers and students understand unfamiliar repositories and software architecture quickly.