

Pranav Raj

[Contact](#) | [Linkedin](#) | [Github](#) | [Personal Website](#) | [Google Scholar](#) |

EDUCATION

Georgia Institute of Technology

Atlanta, GA

Bachelor of Science and Master of Science in Computer Science

4.0 GPA — Aug. 2021 – May 2025

- Courses: Computer Organization (Assembly + C Programming), Systems + Networks, Algorithms, Probability + Statistics, Robotics + Perception, High-Performance Computing, Automata and Complexity, Intro to AI, Computer Vision, Deep Learning, Natural Language Processing, **Teaching Assistant for Machine Learning.**

EXPERIENCE

GPU Applied Architecture (Deep Learning Workloads) Intern

May 2024 – Dec 2024

NVIDIA

Santa Clara, CA

- Performance Captures and Proprietary Analysis of LLM Inference for 10+ Llama, Phi, and Mistral models.
- Corresponding interactive spreadsheet models to understand how/why performance differs based on resource constraints and input parameters (input len, output len, weight precision, flash attention, etc.)
 - * Operation-level, specific breakdown of the Context and Generation phases.

Machine Learning Engineer Intern

May 2023 – Aug 2023

Skytap

Remote

- Developed models (Poly Regression, Exponential Smoothing, Autoregressive Integrated Moving- Average) to forecast resource (CPU, RAM, and storage) utilization using historical utilization/sales data.
- Worked with Stationary and Non-Stationary data; Transformed Non-Stationary data.
- Explored AWS SageMaker DeepAR and Holt's Winter Seasonal Exp Smoothing forecasting algorithms.

Software Engineer Intern

May. 2022 – Aug. 2022

Intuitive Surgical

Sunnyvale, CA

- Created a customized front-end form where end-users input essential information about their API
- Upon form submission, the project generates and runs a Tavern test file, displaying detailed test results.
- Handles a variety of scenarios: certificate checks, enrolling a certificate without a CSR, comparing against secure databases, OKTA Authorization, lifetime bearer token generation, text file checks, etc.
- Integrated within Digital Organization's Jenkins CI/CD pipeline to enable organization-wide use.

Machine Learning Research Lead

Sep. 2021 – Present

Stadium-IoPT Project Team

Atlanta, GA

- Led dev of white paper (accepted in to ISACE 2024); Ran the model on PACE supercomputer Phoenix cluster.
- Improved CNN model that predicts play type from play clip with data cleaning + hyperparameter tuning.
- Created binary classification model and Markov Chain model based on prior play information.
- Created a final neural network that fuses the Markov Model and LSTM Neural Network results.

Socioeconomic Poverty Prediction (SPP) and Parkinson's Lead

Sep. 2021 – Present

Big Data Big Impact

Atlanta, GA

- Improved the socioeconomic indicator to account for new land classifications using Sentinel-2 dataset.
- SPP: Led Platform (collect clean satellite imagery), Analysis (classify imagery and compute final score), and Data Visualization (create Heat Map) teams.
- Parkinson's: Facilitated creation of distinct models to predict Parkinson's (trained SVM classifier/XGBoost algorithms on voice data, Convolutional Neural Networks on handwriting data, etc.)

Data Science Intern

May 2021 – Aug. 2021

Innception

Remote

- Performed massive web scrape (100000+ records) of a freelance services marketplace using Selenium.
- Analyzed data, retrieved from pSQL, to provide valuable insight for the company's major release.

TECHNICAL SKILLS

Python, Java, C++, HTML, Git, PostgreSQL, AWS, Docker, CUDA, REST API, CI/CD (Jenkins), Pyodbc, Databases Pandas, Numpy, Pytorch, Tensorflow, Keras, Jupyter, Selenium, Scikit Learn, NLP, LLMs, DL, CV, Time Series Analysis