Pratyush Muthukumar

pratyush.muthukumar@gmail.com

EXPERIENCE

NVIDIA | Generative AI Data Scientist
October 2023 – Present | Santa Clara, CA

- Developed GPU-parallelized **NVIDIA DALI** data pipeline to preprocess long context video pre-training data with **7X** increase in throughput
- Implemented training & inference scripts for video multimodal vision-language model (VLM) leveraging **Megatron-LM** and **TensorRT-LLM** and trained to convergence exhibiting **4X** speedup across H100 & A100 GPU hardware and **48%** decrease in error vs image-only VLM in NeMo
- Developed **multi-turn synthetic question answer data** generation algorithms to curate training data for training RAG models in the NeMo LLM service at scale
- Led design of CLI-based evaluation tool to monitor & visualize embedding space drift with **t-SNE** & **PCA** and uniformly subsample by embeddings via **Voronoi tesselations**

TESLA | Data Engineer Intern
June 2023 – September 2023 | Palo Alto, CA

- Developed dashboard tools to execute **Spark / Trino** queries and visualize vehicle fleet data with **32 TB daily ingestion rate** hosted on AWS
- Produced time-series reliability predictive model to query vehicle metrics via **SQL** for harmful trend detection in vehicle firmware updates and configured ETL data pipelines via **Airflow**.

LAWRENCE BERKELEY NATIONAL LABORATORY | Research Intern
February 2021 – June 2023 | Berkeley, CA

- Developed robust monitoring tools through the SDNs for End-to-End Networked Science at the Exascale (SENSE) project
- Deployed AI-powered network monitoring system on the SLAC and NERSC labs using **Prometheus** and **Grafana** frameworks to collect and query metrics on end-to-end dataflows of over **2 petabytes**.

NASA JET PROPULSION LABORATORY | Research Associate
August 2019 – December 2022 | Los Angeles, CA

- Used a sequential encoder-decoder **ConvLSTM** model to predict atmospheric remote-sensing imagery of PM2.5 spatially continuously over LA with results showing **over 30% reduction in error** and **24x temporal resolution** over existing approaches.
- Applied Generative Adversarial Network to predict aerospace stock trends through financial text and numerical data showing **over 65% decrease in error** over first-day predictions and **over 70% decrease in error** over 30-day predictions.

PROJECTS

EDVENTURE | Stanford + Meta Corporate Project
Jan 2023 – June 2023 | Stanford, CA

- Designed VR web app companion utilizing **React**, **HTML/CSS/JS**, and **PlayCanvas** to enable virtual exploration of historical sites with the Meta GHA platform
- Prototyped VR knowledge sharing tools with feedback from the Meta **New Product Experimentation** Team and presented at Meta HQ

SATELLITE IMAGE WELL PAD MAPPING | Stanford AI Laboratory (SAIL)
Jan 2023 – June 2023 | Stanford, CA

- Developed multi-task object detection model for mapping methane-emmision sources from global imagery through the Stanford ML Group advised by Professor **Andrew Ng.**
- Achieved all-time high mAP scores in **8 of 10** test regions while preserving scalability

EDUCATION

STANFORD UNIVERSITY
M.S. in Computer Science
Specialization in AI
Graduated June 2024
Supported by NSF GRFP
GPA: 3.9 / 4.0

UNIVERSITY OF CALIFORNIA, IRVINE
B.S. in Computer Science
with minor in statistics
Graduated June 2022
GPA: 3.9 / 4.0

SKILLS

PROGRAMMING
3+ years:
Python • SQL • HTML/CSS
2+ years:
Java • C++ • MATLAB

TECHNOLOGY
AI • AWS • Linux • Git
Spark • Web Dev • Tensorflow
Docker • Trino • Networking
Kubernetes • Keras • PyTorch

COURSEWORK

UNDERGRADUATE
Artificial Intelligence
(Teaching Assistant)
Data Structures
Object-Oriented Programming
Calculus I-III
Differential Equations
Linear Algebra
Probability Theory

GRADUATE
NLP with Deep Learning
Principles of Robot Autonomy
Experimental Robotics
Machine Programming
Financial Trading Strategies

LINKS

Github://
PannuMuthu
LinkedIn://
pratyush-muthukumar
Portfolio://
pratyushmuthukumar.com