

# NAGHARJUN MATHI MARIAPPAN

917-688-8843 • [nagharjun2000@gmail.com](mailto:nagharjun2000@gmail.com) • [Portfolio](#) • [LinkedIn](#) • [Github](#) • New York City, NY

## TECHNICAL SKILLS

**Programming:** Python, C++, SQL

**Libraries/ML:** PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy

**MLOps & Data:** MLflow, Airflow, Spark, Docker, Kubernetes, Git, Tableau, Linux

**Cloud:** AWS - EC2, S3, Lambda, Bedrock, Redshift, ECS; Azure - VMs, Functions, Cognitive Services

**Machine Learning:** Supervised ML, Clustering, PCA, CNNs, LSTMs, Attention, LLMs, Hypothesis Testing

## PROFESSIONAL EXPERIENCE

### Data Scientist - [SCAN](#)

July 2024 - Present

*Mount Sinai Health System, New York, NY*

- **Generated** airflow signal for home sleep-apnea studies lacking nasal cannulas by training a Dilated Residual CNN + Attention on plethysmography, reaching **84% recall** and **75% precision**.
- **Restored** a failed MLflow service within **24h** and **100% data integrity** via custom backup/restore pods and automated MinIO data snapshots using Kubernetes CronJobs.
- **Flagged** and removed sensitive entities on clinical notes with a lightweight zero-shot GLiNER NER, blocking **98%** of protected health information (**PHI**) **leaks** before LLM inference via AWS Bedrock.
- **Reduced** token usage by **40%** by clustering redundant clinical note chunks in **RAG** using cosine similarity.

### Data Science Intern

January 2024 - May 2024

*LOCOMeX, New York, NY*

- **Lifted** contract engagement **20%** using a two-tower recommender system combining BERT text embeddings and company/contract metadata.
- **Slashed** document-parsing time **90%** by deploying a GPT-3.5 Lambda API for PDF/DOCX/PPT extraction.
- **Saved** 8 h/week with Apache Airflow pipelines feeding AWS Redshift and S3.

## ACADEMIC PROJECTS

### AdEvade: Outsmarting ML Ad Blockers - [Link](#)

December 2023 - March 2024

- Caused a **15% detection drop** in a YOLOv5 ad-blocker (90% mIoU) via adversarial tweaks, then retrained to recover accuracy.

### CitiBikeFlow: NYC's Cycling Visuals - [Link](#)

December 2023 - February 2024

- Live-streamed Citi Bike feeds through AWS S3 → Snowpipe → Snowflake; Airflow on EC2 powered Tableau usage dashboards.

### NYC StationSense: Predicting Crowds - [Link](#)

February 2023 - May 2023

- Forecasted subway station crowding using SparkML regression ( $R^2 = 0.62$ ) on 11M records after parallel ETL in Apache Spark.

## EDUCATION

### New York University, New York, NY

September 2022 - May 2024

Master of Science in Computer Engineering

GPA: **3.96/4**

Relevant coursework: Big Data, Machine Learning, Deep Learning, DS and Algorithms, Image Processing, Databases

## OPEN SOURCE CONTRIBUTIONS

### Minimizing KL Divergence Equals Minimizing Cross-Entropy - [Link](#)

May 2025

- Published a Medium article showing the theoretical and empirical link between KL Divergence and Cross Entropy.

### Exporting MLflow Experiments from Restricted HPC Systems - [Link](#)

April 2025

- Published a Towards Data Science article on secure MLflow experiment migration from restricted HPC clusters.