Karan Mudaliar

Boston, MA 02130 | (617) 314-1536 | karan.mudaliar1999@gmail.com | LinkedIn | GitHub

EDUCATION

Northeastern University, Boston MA — Expected May 2025
Master of Science in Data Science — **GPA: 3.89/4.0**
Related Courses: Supervised & Unsupervised Machine Learning, Algorithms, Data Processing & Management, Deep Learning

Vellore Institute of Technology, India — May 2021
Bachelor of Technology in Computer Science and Engineering
Related Courses: Object Oriented Programming, Data Visualization, Statistics, Linear Algebra, Data Structures

WORK EXPERIENCE

Liberty Mutual Insurance, Applied Science Co-op — August 2024 - Present

- **Cut costs** and dependence on GPT-4/OpenAI and **reduced latency** by 20% by **fine-tuning** Small Language Models (**SLMs**) like **Phi-3, Qwen2.5** for **function calling** internal tools for the Director's and Officers' Copilot project
- Developed custom evaluation pipelines, curated high quality **synthetic** and real world data sets and performed **MLflow** experiments to fine-tune and monitor model performance and evaluated benchmarks for tool use.
- **Tuned and evaluated hyperparameters for 15 models (135M to 7B parameters)**, with the best model achieving a **92.5%** fuzzy match score (GPT-4o scored **98%)** despite being a 7B model on a single 24GB GPU.
- Researched best practices for fine tuning Small Language models, and contributed to developing Liberty Mutual's internal Language Model fine tuning library. (**Features added – LoRA, Quantization, gradient checkpointing and accumulation**)

CVS Health, Machine Learning Engineering Intern — May 2024 - August 2024

- Saved 10 engineering hours/week by by implementing a **Text-to-SQL** agent using LangChain and GPT-4o that automated dashboard generation through natural language queries.
- Conducted **literature reviews** to identify the current state-of-the-art text-to-SQL solutions and evaluated multiple Large Language Models (LLMs) to determine their effectiveness for SQL query generation
- Devised a caching framework to **optimize token usage** for generating queries, reducing the required input tokens by approximately 20%

Conde Nast, Data Engineer — February 2021 - July 2022

- Developed and automated end-to-end data-science **ETL** pipelines for enterprise data vendors such as Google Analytics and Spotify, processing up to 0.3 million rows during peak volume
- Led cross-team development between Data Engineering, Platform and QA teams to develop data validation tools, **reducing** validation **time by 80%**
- Executed **backfills** for terabytes of data spanning a period of 3 years to meet the evolving requirements from upstream ML/BI teams

Capgemini, Data Analyst Intern — April 2019 – June 2019

- Produced performance reports analyzing portfolios and identified 3 key underperforming accounts, providing actionable insights to optimize portfolio performance
- Crafted descriptive **reports** using Tableau, analyzing data over 24 months, explaining causes for the drop in performance

RESEARCH EXPERIENCE

Northeastern University, Graduate Research Assistant — October 2023 - Present

- Applied Deep Learning models such as **Crystal Graph Convolutional Neural Networks** to predict surface properties of materials, reducing computational complexity by a factor of 5
- Using ML algorithms to discover new materials for renewable energy applications, contributing to ongoing research efforts

TECHNICAL SKILLS

Programming & ML : R**, Python, SQL,** Pandas, TensorFlow**, Spark, PyTorch,** NumPy, Scikit-learn, Flask, Langchain
Tools & Technologies : Git, MySQL, **Databricks**, **Airflow**, OpenSearch, Neo4j, **Tableau,** AWS, **ETL**, CI/CD, Docker, vLLM
Models & Domains : Hypothesis Testing, CNNs, Logistic and Linear Regression, Decision Trees, OpenAI, GNNs

ACADEMIC PROJECTS

Generative Image and Language Modelling

- Built and trained an **LLM** from scratch for Multi30k German to English dataset with 88% translation accuracy
- Designed and trained a custom Diffusion model to generate MNIST digit images