RAHUL MANIKANDAN MURUGAN

+1 734 548 1717 | New York, USA | rahul.m@columbia.edu | http://www.linkedin.com/in/rahulmurugan/

EDUCATION

Columbia University **New York, USA**
M.S. in Electrical Engineering **(Specializing in Data-Driven Analysis and Computation)** Aug 2024 - Dec 2025
Coursework: Generative AI and Modern Deep Learning, Applied Machine Learning, Neural Networks and Deep Learning, Analysis of Algorithms, Data-Driven methods in Finance

SSN College of Engineering **Chennai, India**
Bachelor of Engineering in Electrical Engineering **(CGPA : 3.63/4)** Nov 2020 - May 2024

TECHNICAL SKILLS

- Programming Languages: Python, C++, R, SQL, MATLAB, HTML/CSS
- Libraries & Frameworks: TensorFlow, Keras, SciPy, Pandas, Scikit-Learn, Flask, PyTorch, Langchain, Matplotlib, Yolo, Streamlit
- Software: Visual Studio, Jupyter, CUDA, Google Cloud Platform, AWS

EXPERIENCE

Columbia University Department of Computer Science **New York, USA**
Graduate Research Assistant – Dr. Vishal Misra Sep 2024 - Present

- Currently serving as **Project Lead** for the **CS Advising Assistant project**, architecting an advanced **chatbot** system to handle student inquiries. Optimizing the Retrieval-Augmented Generation (**RAG**) pipeline for enhanced accuracy and efficiency
- Utilized **OpenAI vector embeddings** with **FAISS** search and **FAQ filtering** for student-specific retrieval, added **out-of-scope detection** with unsupervised learning, and optimized caching to reduce API costs and improve scalability

Indian Institute of Technology (IIT) Madras **Chennai, India**
AI Research Intern Oct 2023 - Jan 2024

- Created a Retrieval-Augmented Generation (**RAG**) pipeline for **Pfizer**, streamlined data extraction from company's **25,000** patent text files with efficient **vector embeddings** and text-splitting tools under the guidance of Dr. Raghunathan Rengaswamy
- Established an effective **question-answering system** within pipeline and increased accuracy up to **98%** by testing and evaluating latest large language models (**Mistral 7b**, **Zephyr 7b**) with use of **GPU** capabilities

Robert Bosch centre of Data Science and Artificial Intelligence – IIT Madras **Chennai, India**
Summer Research Intern May 2023 - Aug 2023

- Automated the process of **data mining** and **extraction** from sustainability reports with **1500+ pages** to validate a **Climate Action tool** by leveraging advanced **NLP** techniques and extracted information from **texts**, **tables**, and **graphs**
- Designed a **cost-effective** methodology and reduced team's validation time of the tool by **60%** by integrating **GPT-4 API** and **Langchain** framework, through efficient testing and verification of model's accuracy using challenging prompts
- Built a robust **Machine Learning model** to analyze India's energy consumption and generation patterns over last decade, achieving a **MAE** of **0.25** and facilitating precise predictions that enhanced decision-making process

RESEARCH & PROJECTS

EQUATOR: A Deterministic Framework for Evaluating LLM Reasoning with Open-Ended Questions Oct 2024 - Jan 2025

- Published on **arXiv**, **EQUATOR** introduces a deterministic evaluation framework for assessing Large Language Model (LLM) reasoning with **open-ended questions**. Automated evaluation pipelines with smaller LLMs like **LLaMA 3.2B**
- Engineered EQUATOR to rigorously evaluate over **1,005+** logic and reasoning questions, enabling testing of **hundreds** of models, including **Ollama**, locally. Initiated plans for public accessibility through a collaborative website to share evaluations and insights

Fine Tuning and Evaluating Large Language Models for HTML Code Generation Jan 2024

- Fine-tuned the **Falcon 7b** language model to generate **HTML** code from natural language prompts. Carried out **BitsAndBytes** quantization with **bnb_4bit** configuration to optimize memory usage and computational efficiency
- Executed **LoRA** (Low-Rank Adaptation) for model **fine-tuning**, accomplishing a training loss reduction to **0.3374** with gradient checkpointing and mixed-precision (**fp16**) training. Assessed the model by exact match comparison

Sentiment Analysis using VADER – Price returns vs Sentiment Aug 2023

- Built an end-to-end **data pipeline** by deploying web scraping methods with **requests** and **BeautifulSoup** to analyze over **100** daily stock news articles from **Finviz**, integrated **VADER** tool to visualize **sentiment shifts** across five trading days
- Integrated the **yFinance API** to **automate** historical price returns extraction, reducing manual data collection by **70%** and with results demonstrating a **90%** correlation between sentiment shifts and stock price movements

LEADERSHIP & ACHIEVEMENTS

- Selected for the highly competitive summer internship at Robert Bosch Centre for Data Science and Artificial Intelligence (RBCDSAI)- **IIT Madras**, among only **20** students chosen from a pool of over **2500+** applicants **nationwide**
- Founded a **Non-Profit Organization (GG cares)**. Raised over **INR 50,000** to set up **2 Covid Care facilities** with over 50 hospital beds during the pandemic. Commemorated by the Minister of Publicity and Information