

Education

New York University, Courant Institute of Mathematical Sciences

September 2023 – May 2025

Masters of Science in Computer Science **3.75/4 GPA**

New York, United States

Indian Institute of Technology Palakkad

September 2018 – May 2022

Bachelor of Technology in Electrical Engineering **8.06/10 GPA**

India

Skills

Languages/Tools: C++, Python, Java, C, HTML/CSS, JavaScript, SQL, CUDA, Kafka, Spark

Technologies/Frameworks: Pytorch, Tensorflow, Numpy, Pandas, React.js, Node.js, Kubernetes, AWS, MongoDB, Git

Technical Skills: Machine Learning, Computer Vision, Natural Language Processing, Reinforcement Learning, GPU

Experience

Princeton NLP Group

May 2024 – Present

Machine Learning Researcher

Princeton, NJ, USA

- Developing an advanced automated dashboard for web search queries using **LLM Agents**, incorporating personalized web agents with **chain of thought (CoT)** reasoning, long-term memory, and **RAG**. Additionally, building a communication framework, utilizing **HTTP** and **REST** protocols for interactions between LLM Agents and APIs.

Grossman School of Medicine, NYU

December 2023 – May 2024

Machine Learning Associate

New York, NY, USA

- Generated synthetic histopathology images using a **Multi-Modal Vision-Language Model (VLM)** based on a **Latent Diffusion Model** additionally guided by **Self-Supervised Learning (SSL)** embeddings from **ImageBERT**.
- Achieved **96.12%** accuracy in colorectal cancer type classification on real and synthetic images by training a **Vision Transformer**-based encoder with Self-Supervised techniques like SimCLR and StableRep.

Serre Lab, Brown University

July 2021 – July 2023

Machine Learning Researcher

Providence, RI, USA

- Enhanced scale invariant object recognition using biologically-inspired **Convolutional Neural Network**, achieving **30-40%** higher accuracy than **ConvNext** and **ResNet**. Secured **4th** place in the **NeurIPS** Sensorium Challenge.
- Generated synthetic images for color correction tasks, achieving state-of-the-art results on the Color Checker Dataset with a **Recurrent Convolutional Neural Network** model that is **10x** more efficient in size.

Publications

- Arjun et al., "**Introducing Attention Mechanism for EEG Signals: Emotion Recognition**", IEEE Engineering in Medicine Biology Society (EMBC), 2021 [Link]
- Arjun et al., "**Subject Independent Emotion Recognition using EEG Signals Employing Attention Driven Neural Networks**", Elsevier Biomedical Signal Processing and Control (BSPC) [Link]
- Arjun et al., "**HMAX Strikes Back: Self-supervised Learning of Human-Like Scale Invariant Representations**", Cognitive Computational Neuroscience, 2024 (Oral Talk) [Link]

Projects

Understanding the Effects Of RLHF and DPO on LLMs | *Huggingface, Python, HPC*

[GitHub]

- Analyzed the impact of Direct Preference Optimization (**DPO**) and Reinforcement Learning from Human Feedback (**RLHF**) on LLMs' output generalization and diversity by fine-tuning the **Mistral-7B-v0.1** model for summarization tasks using advanced techniques like **PEFT LoRA** adapters and **4-bit quantization**.

Advanced Music Generation using Language Models | *PyTorch, Python*

[GitHub]

- Developed a **Transformer-XL** based Language Model for generating novel multi-instrument music compositions. Integrated **Vector Quantised Variational Auto Encoder** for high-fidelity, diverse music generation, achieving near-lossless **6x** audio compression for efficient training and inference.

Text-to-SQL Context-Aware Query System | *Huggingface, LangChain, ChromaDB, Python*

[GitHub]

- Created a Text-to-SQL system utilizing LLMs augmented with Retrieval Augmented Generation (**RAG**) to generate context-aware SQL queries. Performed **Parameter Efficient Fine Tuning (PEFT)** of **Llama2-7b** using **LoRA** adapters on WikiSQL & Spider datasets and created a user-friendly interface for querying and interacting with IPEDS.

Content-Based Music Recommendation System | *Kafka, PySpark, Spark SQL, Mllib*

[GitHub]

- Implemented a content-based music recommendation system using **Apache Kafka** for streaming data, **PySpark** for data processing, and **Spark SQL** for querying structured data. Utilized **Mllib** for **KMeans** and **PCA** analysis on the Spotify Tracks Dataset, and visualized results with Streamlit, recommending similar songs based on liked playlists.

Achievements

- Reviewer for IEEE Transactions on Neural Networks and Learning Systems