Y Harsha Vardhana Reddy

8971871833 harshavyr@iisc.ac.in
ai.harsha@outlook.com Harsha

Education

M.Tech in Computational and Data Sciences — **Aug. 2024 – current**
Indian Institute of Science, Bangalore — *CGPA: 7.9/10*

B.Tech in Aeronautical Engineering — **2013 – 2017**
Nitte Meenakshi Institute of Technology — *CGPA: 8.54/10*

Projects

Large Language models for Rare Disease Concept Normalization (M.Tech)

* Built a phenotype normalization model using GPT embeddings to map HPO terms into a vector database
* It normalizes phenotypes by applying NER to extract phenotypes from clinical text, then uses a sentence transformer and FAISS for similarity search, enabling accurate retrieval of the closest HPO IDs for the identified phenotypic terms.

BIOMed_NER: Named Entity Recognition for Biomedical Entities

* Developed a **Named Entity Recognition (NER)** model using **DeBERTaV3 base** to tag 41 biomedical entities.
* Utilized the **MACCROBAT2020** dataset with 400 clinical notes, split 80% for training and 20% for validation, to accurately tag biomedical terms in clinical text.Published the model on Hugging Face as **BIOMed_NER**, making it available for the research and healthcare community.

Deep Convolutional and Wasserstein GANs for High-Quality Animal Image Synthesis

* Implemented both DCGAN and WGAN from scratch using PyTorch on an Animal Image Dataset containing 90 classes.
* Trained the models to generate realistic animal images, with WGAN providing enhanced stability and quality in generated samples through improved loss optimization.

Efficient GPT-2 Implementation for Text Auto-Completion Tasks

* Implemented and trained a GPT-2 model from strach using pytorch on the FineWeb-Edu dataset of 10 billion tokens.Validated the model using the HellaSwag dataset, achieving an accuracy of 31.33%.
* Utilized Distributed Data-Parallel (DDP) to efficiently leverage 2 RTX A6000 GPUs (48GB each) for training.

Machine Translation using Transformers

* Developed a Transformer model from scratch using PyTorch for machine translation.
* Processed 300,000 English-Kannada translation samples and evaluated model performance using BLEU score

Mini-Projects

* **Classification Models:** Implemented logistic regression and K-Nearest Neighbors (KNN) for binary and multi-class classification tasks using Python
* **Algorithms from Scratch:** Implemented Decision Tree, PCA, and K-Means clustering algorithms using Python.
* **Neural Networks:** Implemented Multi-Layer Perceptron (MLP) and Convolutional Neural Networks (CNN) from scratch using Python and tested them on MNIST and CIFAR-10 dataset.

Relevant Coursework

* Machine Learning for Data Science
* Advanced deep representation learning
* Deep Learning for Computer Vision

Technical Skills

Languages | **Tools**: : Python, C++, PyTorch, NumPy, Matplotlib, Pandas, SQL.
Expertise: NLP, Computer Vision, Generative Modeling, Deep Learning, Machine Learning