Vanshika Sharma

+1 862-433-9846 | vanshika.sh3143@gmail.com | LinkedIN | GitHub | TowardsDataScience

Summary

Data Scientist with an extensive experience of 5 years in leveraging and deploying Machine learning models focused in Natural Language Processing and Generative AI, using frameworks like Large Language Models and multi-modal learning. Specialized in Software Development and Data Science/ML Algorithms having a Master's degree in Computer Science, focused in Data Science.

Technical Skills

Programming: Python, PyTorch, TensorFlow, Keras, PySpark, SQL
Technologies: Amazon Sagemaker, Azure Databricks, Azure Data Factory, Natural Language Processing, Database Systems, Large Language/Transformer Models

Work Experience

Machine Learning Engineer **Jan 2023-Present**
Nagarro Inc. - *Natural Language Processing and Database Systems* *Atlanta,GA,USA*

- Working on devising and executing LLM pipelines using RAG and Prompt Engineering to extract valuable insights from client-specific data sources with over 90 percent tested credibility
- Contributed extensively to a 6-member team to perform Data Migration between 2 database systems, for a major automotive client in the US using Azure Data Factory pipelines and Databricks PySpark scripts

Research Assistant (Machine Learning) **May 2022-Sept 2023**
SUNY Stony Brook University - *Context Aware NLP Multi-modal Applications Research* *New York, USA*

- Generated Lung CT Images based on conditions over diverse organ sizes using multi-modal Probabilistic Diffusion Models, integrated with GPT Models
- Integrated Visual and Sensor Tracking to perform real-time indoor object tracking using Dead Reckoning Principles with 90.7 percent accuracy using YOLOV3 model and Transformers

Machine Learning Engineer **Jun 2019-Aug 2021**
TMotions Global Ltd. - *Natural Language Model Development and Cloud Deployment* *New Delhi, India*

- Leveraged BERT, for Named Entity Recognition (NER), efficiently automating the extraction of structured information from Fine Tickets for Portuguese Government's traffic management system
- Performed sentiment analysis for classifying product reviews for an e-commerce website by into 'Highly Satisfied', 'Satisfied' and 'Disappointed' categories using LSTMs
- Orchestrated and managed end-to-end Machine Learning Dev pipelines for deployment of the ML models on AWS

Education

Stony Brook University, State University of New York **Aug 2021-May 2023**
Masters in Computer Science, specializing in Data Science/ML Domain

Birla Institute of Technology and Science Pilani **Aug 2015-Jun 2019**
Bachelor's in Engineering, Computer Science

Research/Academic Projects

Lung CT Image Generation | *Pytorch/GPU* **Jan 2022 - May 2022**

- Generated accurate Lung CT Images integrating Denoising Probabilistic Diffusion Models and GPT models, enhanced by 90 percent
- Incorporated a conditioned element by fine-tuning the model based on organ size into the image generation process, achieving an 80 percent accuracy rate in generating condition-specific images
- Link to project and results: | LINK |

IMU and Visual Sensor Pedestrian Tracking | *Tensorflow/OpenCV* **Jan 2022 - May 2022**

- Real-time Indoor Pedestrian Tracking system using IMU data and CCTV feed, achieving 90 percent accuracy.
- Integrated Dead Reckoning using transformers with YOLOV3 for seamless tracking in dynamic environments.
- Link to project and results: | LINK |