# Vanshika Sharma

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

## Summary

A Machine Learning Engineer with an extensive experience of 5 years in leveraging Generative AI models, frameworks like LLMs, Dall-e, LlamaIndex, Langchain, Retrieval Augmented Generation (RAG); creating end-to-end data modeling pipelines through Azure and AWS. Specialized in Software Development and ML Algorithms with a Master's degree in Computer Science, with a curriculum focused in Data Science.

## Technical Skills

**Programming**: Python, PyTorch, TensorFlow, Keras, PySpark, Natural Language Processing, Statistics, SQL
**Technologies**: Azure Cloud, Git, AWS Sagemaker, Generative AI, LLMs, Azure Datbricks, Azure Data Factory, MLFlow
**Packages**: Scikit-Learn, OpenCV, NLTK, SpaCy, Transformers, Docker, Kubernetes, MLflow, PyTorch Lightning

## Work Experience

**Machine Learning Engineer** **Jan 2023-Present**
*Nagarro Inc.* *Atlanta,GA,USA*

- Leveraged LLMs to extract precise database insights through chatbots developed for a major Automotive Client's Warranty data domain, fine-tuned models via PEFT, and deployed with Azure ML
- Worked on devising and executing end-to-end LLM pipelines using RAG and Prompt Engineering to extract valuable insights from client-specific data sources with over 90 percent tested credibility
- Contributed to a 6-member team in design and construction of an Azure Data Factory pipeline integrated with Databricks PySpark scripts to perform seamless Data Migration between two different database platforms

**Research Assistant (Machine Learning)** **May 2022-Sept 2023**
*SUNY Stony Brook University* *New York, USA*

- Performed accurate conditional Lung CT Image Generation using Denoising Probabilistic Diffusion Models to build high resolution augmented, conditioned datasets used for diagnosis model training and validation
- 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
- Integrated Visual (CCTV) and Sensor (IMU) Tracking to perform real-time indoor pedestrian tracking with over 90 percent accuracy using YOLOV3 models and Transformers

**Machine Learning Engineer** **Jun 2019-Aug 2021**
*TMotions Global Ltd.* *New Delhi, India*

- Orchestrated and managed end-to-end Machine Learning Dev pipelines from development to deployment on AWS for Portuguese Government's traffic management system
- Utilized Amazon SageMaker for model training and hosting, Docker for containerization, and AWS Lambda for serverless computing, used CloudWatch for real-time model performance monitoring
- Developed a Few-Shot Learning Model for an E-commerce website using Siamese Neural Networks to classify over 5000 jewellery designs, reducing manual effort by over 2 hours, streamlining cataloging processes
- Leveraged BERT, a state-of-the-art transformer model, for document parsing and Named Entity Recognition (NER), efficiently extracting structured information from unstructured data sources for Portuguese Government's traffic management system
- Gathered requirements directly from clients, proposed and developed end-to-end solutions, ensuring alignment with client needs and satisfaction

## 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 using Denoising Probabilistic Diffusion 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/Pose Estimator* **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 |