ROHIT TIWARI

✉ knowrohit.work@gmail.com 🐦 Twitter Github

EXPERIENCE

TartanHQ   **July '22 – Present**
Data Scientist
- Successfully deployed and fine-tuned Document extraction OCR models using LayoutLM v3 and BERT on AWS EC2, replacing third party APIs with inbuilt solution
- Built and deployed models for Document Classification, Payslip Forgery, and Blur Detection using Tesseract, Torchvision, Detectron2, and Spacy NLP, achieving a 95 percent accuracy rate.
- Developed state-of-the-art NLP solutions for text summarization in Payroll information of employees using GPT-3 and XLNet, improving client satisfaction by 30 percent.
- Analyzed Data inflow, timestamps of Payslips, and their metadata using Apache Superset (MySQL); contributed to the development of a custom annotation tool for OCR.

Trill Marketplace   **September '21 – March '22**
MLOps Engineer
- Developed an RNN and ChatBot model for customer assistant using Transformers, GPT-NEO, resulting in a 30 percent improvement in creative output; incorporated CI/CD pipelines for seamless model updates and deployment.
- Worked on ML systems using PyTorch in production to scrape and learn from the clustered data of streetwear customers, used MLflow for tracking experiments, reproducibility, and model versioning.
- Collaborated with data engineering teams to establish robust data pipelines, using Apache Kafka, ensuring data quality and consistency for ML model training.

GooseAI   **April '21 – July '21**
Data Science - Intern
- Worked on Recurrent Neural Nets for NLP and experimental ML with physics such as Phononics, dimensionality reduction, and federated learning, achieving a 90 percent success rate in simulations.

PROJECTS

SCRIPTWRITING USING RNN | *TensorFlow, Keras, Numpy* | Github
- Developed a recurrent neural network for scriptwriting that detects user input texts and generates new movie scripts
- Improved text generation quality by 25 percent with additional training epochs

Bernard-The first sentient AI Chatbot | *Hugging Face, Transformers, GPT-2, PyTorch* | Github
- Bernard combines a sophisticated neural network machine learning model and scripted dialogue content.
- Utilized transformers and decoding methods to mimic the user's talking style, improving user engagement by 40 percent
- Bernard's heart lies a complex autoregressive language model GPT-2 that utilizes deep learning to produce human-like text.

Product Recommendation Engine | *VGG16, AWS, Docker, HuggingFace, BS4* | GitHub
- Developed a recommendation engine using a dataset of 44k images scraped from Amazon.
- Implemented feature extraction using Cosine similarities, VGG hidden layers, and Azure image tags.
- Achieved accurate predictions using similarity scores and ensemble learning with VGG16 in computer vision and NCF for combined predictions, deployed on AWS EC2.

NFT-GAN | *torchvision, numpy, PIL, tensorboard, scipy* | GitHub
- Developed an NFT-GAN for creating NFT avatar and collectible projects using Gmapping and Gsynthesis.
- Configured image generation processes to control trait rarity and generate JSON metadata for NFTs in compliance with OpenSea metadata requirements.

TECHNICAL SKILLS

Languages: Python, C/C++, SQL
Technologies: Anaconda, Apache, NVIDIA Cuda, Linux, Jupyter, PostgreSQL, MySQL Shopify, WordPress
Developer Tools: Git, Docker, Microsoft Azure ML, VS Code, AWS Sagemaker, AWS EC2
Libraries: PyTorch, TensorFlow, CUDA, BigSLEEP, Numpy, Pandas, Scikit-Learn, OpenCV, FastAPI, Flask
MLOPS: Kubeflow, MLflow, WeightsBiases

EXTRA-CURRICULAR

- Member of Amity Linux Assistance Sapience Club(ALiAS)
- Zonal-Level Football Champion, BBFS FC