ROHIT TIWARI

knowrohit.work@gmail.com LinkedIn Github

EXPERIENCE

TartanHQ **July '22 – April '23**

Data Scientist

- Successfully deployed and fine-tuned Document extraction OCR models using LayoutLM v3 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.
- 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 **February '21 – March '22**

Machine Learning Engineer

- Developed an LLM for customer assistance using GPT-NEO(20B); incorporated CI/CD pipelines for model updates and deployment.
- Worked on ML systems using Google Pegasus in production to scrape and learn from the clustered/textual data of streetwear customers.
- Collaborated with data engineering teams to establish robust data pipelines, used MLflow for tracking experiments, reproducibility, and model versioning.

GooseAI **April '21 – July '21**

Data Science: Intern

- Contributed on experimental ML with physics such as Phononics, and federated learning, achieving a 90 percent success rate in simulations.

PROJECTS

Saraswati LLM | *Transformers, Node-js, JavaScript, Docker* |

- A full-stack application allowing students to upload academic documents and engage with my own state-of-the-art LLM, "Saraswati".
- RAG System: Upload academic content (books, PDFs). Uses Sentence Transformers for embedding, FAISS indexing for similarity matching, and integrates with LLM for generating personalized answers based on chunks + query.
- Code Scripting: Generate scripts in various programming languages, trained on 20k high-quality leetcode and stackoverflow infilling samples.
- Logical / IQ Tasks: Perform arithmetic, and chain-of-thought reasoning.
- Model performance: Competing with GPT-3.5 in benchmarks and overall utility.

know-medical-dialogues | *HuggingFace,Datasets,LLM* | HuggingFace

- Dataset of conversational exchanges between patients and doctors on diverse medical topics, covers a wide range of medical queries and advice.
- Beneficial for healthcare professionals, especially in resource-limited regions. Facilitates AI-assisted medical roleplays on affordable devices if trained on quantized LLMs.
- Derived from anonymized patient-doctor seed interactions and then synthetically generated via GPT4.

VirgilAbloh GPT | *Transformers, PyTorch, GLM* | Github

- Virgil engages in fashion-related conversations, provides tailored recommendations, and offers assistance on various tasks like coding, humour.
- Utilized transformers and decoding methods to mimic the user's talking style, also using Text-to-Speech Conversion
- Virgil's heart lies a complex autoregressive language model GLM that utilizes deep learning to produce human-like text.

Fashion RecSys | *Tensorflow, Azure, Docker, HuggingFace, BS4* | GitHub

- AI-powered product recommender system designed to help users find their perfect fashion match with user-friendly UI.
- Implemented feature extraction using Cosine similarities, ResNet hidden layers, and Azure image tags.
- New updates will have video-based recommendations using GCP video intelligence, to be deployed soon.

NFT-GAN | *torchvision, Numpy, PIL, 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.

Open Source Projects | *HuggingFace, Datasets, LLMs* | HuggingFace

- Clean uncensored datasets using Self-instruct fine tuning.
- regular updates with large majority of HF models, in any modality (text, vision, multi-modal, etc.)

TECHNICAL SKILLS

Languages:Python, C/C++, SQL
Technologies: Anaconda, Apache, NVIDIA Cuda, Linux, Jupyter, PostgreSQL, Azure, Shopify, WordPress
Developer Tools: Git, Docker, Microsoft Azure ML, VS Code, AWS Sagemaker, AWS EC2
Libraries: PyTorch, TensorFlow, CUDA, gRPC, Langchain, chromaDB, 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