URVIL PANCHAL

Ahmedabad | 9016760721| urvilpanchal13@gmail.com | LinkedIn | Github | Portfolio

Work Experience

DATA SCIENCE TRAINEE @ NEXUSLINK SERVICES PVT. LTD. Jun 2024 – Jan 2025

- Developed an end-to-end object detection system using the yolo-v8, with deployment on **RunPod** serverless.
- Built a **RAG** system leveraging large language models (LLMs) for enhanced information retrieval. Created and deployed scalable APIs using **FastAPI** and containerized applications with **Docker**.
- Managed cloud infrastructure and deployments on **AWS**, optimizing for performance and cost, while handling version control and code management through **GitHub** to streamline team collaboration.

Skills

Programming Languages: Python
Core Competencies: Deep Learning, Computer Vision, Natural Language Processing, Machine Learning
AI/ML Frameworks: PyTorch, TensorFlow, Langchain, Scikit-Learn
Tools & Platforms: Git, Github, Docker, Amazon Web Services (AWS)
Additional Skills: Flask, Streamlit, FastAPI, LLM Fine-tuning, Ollama, Transformers, RAG Techniques

Projects

Fine-Tuned DeepSeek-R1 on Bhagwad_Geeta | Github

- Fine-tuned **DeepSeek-R1-Distil-Qwen-1.5B** on the Bhagwad Geeta using **LoRA**, optimizing for efficient adaptation with low-rank updates using parameter-efficient fine-tuning (**PEFT**) technique.
- Leveraged PEFT techniques to reduce memory usage while maintaining model performance.
- Developed a Streamlit-based chat UI, enabling interactive conversations.

NSFW Detection System

- Trained a **YOLOv8** object detection model to detect explicit content by identifying adult imagery. Data was manually annotated to ensure high accuracy.
- Created a **Docker** application for the detection system, enabling seamless integration and scalability.
- Hosted the system on **RunPod**, providing efficient and scalable cloud-based detection for NSFW content.

LLM-Powered Document Q&A System | Medium article

- Developed a **Retrieval-Augmented Generation (RAG)** application that accepts text and PDF files as inputs, and let users to ask questions and get responses based on the content of the files.
- Used **Llama 3.1-70B** as the Large Language Model (LLM) with **Groq API**, and **Qdrant Vector Store** for efficient data storage and document retrieval.
- Used **Langchain** as the framework to seamlessly integrate the LLM, document retrieval, and user interaction for an interactive Q&A experience.

Object Detection Using Detection Transformer

- Annotated over **8,000 doors** in floor-plan images for object detection
- Trained **Facebook's DeTr** model on custom annotated data to detect doors in floor plans.
- Developed and deployed the model in a **Docker image** for scalable use.

Fine-Tuned BERT for Sentiment Analysis

- Fine-tuned BERT for binary **sentiment analysis** (Positive/Negative) using the Transformer library**.**
- **Tokenized** input text and applied **positional encoding** to handle word order.
- Used **Masked Attention** to ensure the model focused on relevant parts of the input during training.

Implemented GraphRAG

- Implemented **GraphRAG** using Langchain framework for retrieval-augmented generation.
- Integrated **Llama 3.1-70B** as the large language model with **Groq API** for enhanced performance.
- Utilized **Neo4j Vector Database** for efficient storage and retrieval of embeddings.

Education

B.E in Information Technology • Gujarat Technological University 2020-2024