Lalit Kumar

SUMMARY

A skilled GenAI Engineer adept in leveraging Azure services to develop innovative solutions for NLP-related tasks and data science challenges. with strong foundation in software development and passion for advancing technologies.

CONTACT

Phone:
+91 8867400741

Email:
LLTKUMAR69@GMAIL.COM

LinkedIn:
https://www.linkedin.com/in/lalit-sheoran

GitHub:
https://github.com/LalitSheoran-repo

SKILL HIGHLIGHTS

➢ **Programming Language**
Python, JavaScript, HTML

➢ **Tools and Software**
OpenAI GPT Models, Langchain, ChromaDB, PostgreSQL, MongoDB, PyTorch, Pandas, NumPy, Bootstrap

Proficient In

- GenAI
- Prompt Engineering
- Vector Database
- Machine Learning
- NLP
- Microsoft Azure
- CI/CD Pipeline
- PySpark

EDUCATION

Master of Science, Information Technology Aug 2020 – 2022 Aug 2020 – May 2022
Dhirubhai Ambani Institute of Information and Communication Technology (DAIICT) **CGPA: 7.87/ 10**

EXPERIENCE

Fractal Analytics (Jun22 - Current)

Automated Proforma Generation

- Automated Proforma Generation for mortgage-based loans involves processing rental and operating statements of buildings.
- Utilizes **OpenCV** for image preprocessing, including Otsu thresholding, skew correction, and denoising, and leverages **Pandas** and **NumPy** for processing Excel and CSV files.
- Employs **Azure Document Intelligence** and **Azure GPT Vision** models for content extraction from documents.
- Organizes extracted content using a **Pytorch** implementation of a Transformer-based model, specifically the **table-transformer**, to differentiate between tables and paragraphs.
- Recategorizes extracted content into a structured JSON object that outlines loan amount eligibility, utilizing **Prompt Engineering** with **Azure OpenAI**.

BrainTrain:

- Implemented smooth content retrieval from a localized internal database, utilizing **LangChain**, **GPT-4**, and **ChromaDB**.
- Developed a file search service on **Azure Blob Storage** by indexing files with Lucene library, with indexes saved in **Azure Cognitive Services** for efficient file searching.
- Automated the generation of PPT skeleton and content from a localized database using a **RAG**-based approach supported by ChatGPT, restructured the PPT to fit specified templates via the python-pptx library.
- Enabled real-time web search functionality using Bing Search and Perplexity LLM.

Microsoft Responsible AI Framework Accelerator:

- Generated a comprehensive test dataset utilizing Large Language Models (LLM) to simulate real-world news scenarios,
- Developed a news classification pipeline employing **Transformers, XGBoost, SVM,** and **Decision Trees**.
- Implemented a multimodal functionality integrating traditional algorithms and LLM,
- Evaluated outputs using the Microsoft Responsible AI framework.

ACADEMIC PROJECTS

Publication:

Summary: Published a research paper titled "Deep Learning Based Automated Localization of Anterior Commissure and Posterior Commissure Landmarks in 3D Space from 2D Three-Plane MRI Localizer Slices of the Brain," accepted at the International Conference on Machine Learning and Data Engineering (ICMLDE) 2022.
Guide: Dr. Bakul Gohel
Link: PDF

Sentiment Analysis of YouTube Videos and Tweets:

Summary: Sentiment analysis of YouTube videos as well as Live speech was accomplished. The aim was to predict emotion out of a spectrum of 5 emotions. In this project we implemented Ensemble method of Vector Space Model (VSM)(Implemented from a research paper) and LSTM. Also, we designed Web-App to fulfil the aim of building an end-to-end application.
Guide: Prof. Ahlad Kumar
Link: GitHub Link

CERTIFICATIONS

➢ AZ-204 (Azure Developer Associate)
➢ Associate Pyspark Developer (Fractal Certified)
➢ AZ - 900(Microsoft Azure Fundamentals)
➢ DP - 900 (Microsoft Azure Data Fundamentals)
➢ Databricks Accredited Lakehouse Platform Fundamentals