Debasish Pradhan

+91 xxxxxxxxxx | pradhandebasishxxxx@gmail.com | Hyderabad, India

Linkedin | Github | Medium

WORK EXPERIENCE

Generative AI Engineer

MLAI Digital | Hyderabad   Mar 2024 - present

Employee Bot

- Developed a RAG-based conversational **chatbot**, ingesting data from **45,000+** URLs and **10,000+** documents, to create a comprehensive private knowledge base.
- Optimized retrieval quality through preprocessing, content-aware **chunking**, and custom **contextual** retrieval using **BM25s**, **Qdrant** vector DB with **hybrid** search, and metadata filtering.
- Applied Re-ranking, Sliding Window, and RAG fusion to generate high-precision answers using **Llama 3** and **3.1** family models with a **96%** reliability rate.
- **Updated** the knowledge base **hourly** to reflect any changes in URLs, ensuring the chatbot provides the most current and accurate information.

Synthetic Data Generation

- Generated over **1.2 million** synthetic data points for diverse email types using the **Llama 3.1 405b** model.
- Performed EDA, data cleaning, data transformation, and **clustering** techniques to enhance intent prediction accuracy.
- Improved model performance through iterative testing and **chain-of-thought** prompting, ensuring high-quality intent generation.
- Collaborated with the business team to align synthetic data generation with specific intent prediction needs.

Personalized Email Bot

- Built an email bot using **RAG** (Retrieval-Augmented Generation) and **few-shot** prompting to automatically retrieve the most relevant policies and templates out of **120** sources for generating emails based on user-provided topics.
- Integrated template matching in prompts, enhancing the bot's ability to produce consistent and professional email outputs.
- Incorporated data security measures by masking user input with Python regex, ensuring data privacy and integrity.

Code and Document Comparison

- Created a Python-based tool to detect changes between two PDF or SQL code files, ensuring accurate and efficient comparison.
- Integrated **LLM** models for SQL code explanation, providing detailed insights into detected changes.
- Displayed and highlighted the changes using **PyMuPDF**, enhancing clarity and user experience.
- Collaborated with the hardware team to strengthen infrastructure, achieving a **50%** improvement in response latency.

Voice Call Summarization and Extraction

- Built a pipeline to translate customer-agent conversations from **Hinglish** to **English**.
- Extracted key information such as customer queries, products, and issue resolution status using **function calling**, reducing **90%** of manual work.
- **Summarized** entire conversations using prompt engineering to provide detailed insights.

AI Engineer   Aug 2023 - Jan 2024

Prizmato | Hyderabad

- Created **100+** client-specific training materials on Machine Learning, Deep Learning, NLP, CV, and Generative AI, to facilitate employee learning.
- Helped clarify doubts and guided students at the client location, ensuring a clear understanding of the concept.
- Trained and mentored over **20** internal team members in data science and machine learning, leading to a notable improvement in team expertise.

Data Scientist Intern   May 2023 - Sep 2023

Techweirdo | Remote

- Implemented a robust data extraction solution for **50+** different types of bank statements and ITR documents.
- Formulated a **Large Language model(LLM)** that can extract non-tabular data and integrated computer vision techniques for tabular data extraction with an accuracy of **95%**.
- Scraped raw data from **20+** food and e-commerce websites with **selenium** and extracted structure data with the help of **Open AI GPT-3.5 turbo.**
- Identified and resolved complex edge cases and bugs within existing projects, optimizing model performance by **5-10%** through troubleshooting and iterative testing.

SKILLS

Programming Languages: Python, SQL, Java
Data Analysis: Data Visualization techniques, Data Preprocessing techniques, Exploratory Data Analysis(EDA), Feature Engineering, Unstructured and Structured Data handling, Scaling & Normalization
Machine Learning: Supervised Learning, Unsupervised Learning, Regression, Classification, PCA, SVD, Decision Tree, Bagging, Boosting, Clustering, Collaborative Filtering, and Recommendation techniques
Deep Learning: Artificial Neural Networks (ANN), Deep Neural Networks (DNN), Recurrent Neural Networks (RNN)
Computer Vision: Convolution Neural Networks (CNN), Object Detection, Image Classification, Image Segmentation, OCR
Natural Language Processing(NLP): LSTM, GRU, Text Classification, Text Preprocessing, NER, Text Embeddings, Sentiment Analysis, Tokenization, Stemming, Lemmatization
Generative AI: Auto Encoders, VAE, Embeddings, BERT, Transformers, Attention, Encoding & Decoding, Open AI GPT 3/4, LLM, Prompt Engineering, VectorDB, RAG, PEFT, LoRA, QLoRA, AI Agents
Cloud: AWS, Amazon Sagemaker, GCP, Vertex AI
Python/ML Libraries: Numpy, Pandas, Matplotlib, Seaborn, Plotly, scikit-learn, TensorFlow, Keras, PyTorch, OpenCV, Langchain, Hugging Face
Model Deployment/ML Ops: MLOps, CI/CD, GitHub Actions, Docker, DVC
Database: MySQL Workbench, MongoDB, MongoDB Compas, MongoDB Atlas
Development Tools: Jupyter, Vs Code, Linux, Bash, Git, Anaconda, Pycharm, Spyder

PROJECTS

Indian Coin Detection   View in GitHub
- Developed and integrated a state-of-the-art **YOLO-V8** model for the detection and classification of Indian coins, resulting in an overall map of **92%** after training with 100 epochs.
- Implemented a modular and well-organized Python codebase for the entire project, ensuring maintainability and scalability.
- Containerized the application using **Docker** and established a **CICD** pipeline to build and push a new image every time the application is modified.
- Leveraged **GitHub Actions** and **DVC** to automate the syncing of highlights whenever updates are made.
- Deployed the project on an **AWS EC2** instance, enabling seamless accessibility as a real-world application.

Sentiment Analysis with Fine-Tuned LLM   View in GitHub
- Fine-tuned **TinyLlama** with crypto news dataset for sentiment analysis and subject prediction of a given news.
- Trained by reducing **92%** of trainable parameters through efficient fine-tuning using **PEFT** and **LoRA** methods on a single GPU, achieving an impressive loss of **0.13**.
- Deployed the fine-tuned model using **AWS Sagemaker** and **Lambda Function**, achieving optimal scalability and real-time inference capabilities

Bosch Production Line Performance   View in GitHub
- Managed and analyzed a large-scale dataset with over **1.1 million** rows to optimize production line performance.
- Explored **10+** Machine Learning Models with feature engineering, feature selection, and hyperparameter tuning; Resulting in **XGBoost** achieving the highest **MCC** score of **0.30** on Kaggle test data.
- Integrated the final model into a flask-based webpage, enhancing the overall user experience.

CERTIFICATIONS

Generative AI with Large Language Models (Coursera)   View Credentials
Applied AI Course (Applied AI)   View Credentials
Introduction To Generative AI (Google)   View Credentials

EDUCATION

Bachelor of Technology, Mechanical   July 2019 - May 2023
Parala Maharaja Engineering College - (8.36/10)   Berhampur, Odisha, India

Resonance Residential College   2017 - 2019
Percentage: 74%   Cuttack, Odisha, India