

Specialized LLM Bot Using Pre-Trained Models

Project Summary

Name: Ashish Rajput

Email : akrajput801@gmail.com

Deep Learning for NLP Workflow Diagram

- **Project Planning and Setup**
 - Objectives definition
 - Industry selection
 - Environment setup
- **Data Collection**
 - Identify data sources
 - Collect and preprocess data
- **Model Selection**
 - Choose pre-trained model
 - Set up training environment
- **Fine-Tuning the Model**
 - Fine-tune model
 - Limit training epochs
- **Chatbot Development**
 - Develop framework
 - Integrate model
 - Test and validate
- **User Interface Design**
 - Create user interface
 - Add features
- **Deployment**
 - Host backend with ngrok
 - Deploy frontend with Streamlit

Github Link:- <https://github.com/Mr-Rajput801/IndustryGPT-Specialized-LLM-Bot-Using-Pre-Trained-Model>

Summary

Overview

In this capstone project, we will embark on an exciting journey to create an Industry-Specific Large Language Model (LLM) Bot using state-of-the-art pre-trained models from sources like Hugging Face. The primary objective is to build an intelligent bot that can effectively engage with users by answering questions and providing insights specific to a chosen industry. This project will not only enhance your technical skills but also provide a deep understanding of the chosen industry's nuances, challenges, and trends.

Project Objectives:

- **Industry Selection:** Select one industry and this industry will be the focus of their project, including data collection and model training.
- **Data Collection:** Gather relevant data specific to the chosen industry. This data will be used to fine-tune the pre-trained model to ensure the LLM Bot is knowledgeable and contextually aware of industry-specific information.
- **Model Selection and Training:** Utilize any pre-trained model from Hugging Face or similar platforms. Fine-tune the model on the collected data using resources like Google Colab with T4 GPUs, limiting the training to a maximum of 25 epochs to ensure feasibility.
- **Bot Development:** Develop the LLM Bot that can interact with users, providing answers and engaging in meaningful conversations specific to the chosen industry. The bot should demonstrate the ability to understand and process industry-related queries effectively.
- **Demonstration:** Create an explanatory video showcasing the working of the LLM Bot. The video should highlight the bot's ability to handle industry-specific questions, demonstrating its practical application.

Problem Statement:

- The aim of this capstone project is to create an industry-specific Large Language Model (LLM) Bot by leveraging pre-trained models available on platforms like Hugging Face.
- Collect data, and fine-tune a pre-trained LLM
- Showcase the bot's ability to interact with users efficiently by delivering precise and contextually relevant responses.

Abstract

This project demonstrates the end-to-end process of developing an industry-specific finance chatbot using pre-trained models from Hugging Face. By fine-tuning the model with finance-specific data and leveraging advanced techniques like PEFT, LoRA, and QLoRA, the chatbot is optimized for efficient and accurate responses. The deployment of the chatbot using Flask and Streamlit ensures a user-friendly interaction, while the detailed demonstration showcases its practical application in the finance industry. Future improvements could further enhance the chatbot's capabilities, making it an invaluable tool for industry-specific queries and insights.

Industry Selection:

Chosen Industry:- Finance

Reasons:

High Demand for AI in Finance:

Automation of Processes: The finance industry increasingly relies on AI for automating routine tasks such as transaction processing, fraud detection, and customer service.

Rich and Diverse Data Sources:

Varied Data Types: The finance sector deals with diverse data types, including transactional data, market data, news articles, and social media sentiment.

Complex and Specialized Language:

Industry-Specific Terminology: Finance has a unique vocabulary and complex concepts that require specialized understanding and handling.

Enhancing Customer Experience:

Personalized Financial Advice: An LLM Bot can provide tailored financial advice and personalized investment recommendations.

Addressing Security and Fraud:

Fraud Detection: AI and LLMs can help in identifying unusual patterns and potential fraudulent activities.

The finance industry's dynamic nature, coupled with the critical need for precision, makes it an ideal choice for developing a specialized LLM Bot. By focusing on finance, this project aims to harness the power of AI to address complex challenges, improve efficiency, and enhance user experiences in a vital sector.

Dataset Collection

In the process of developing our Finance Industry-specific LLM Bot, the data collection phase is crucial for fine-tuning the pre-trained model to ensure it is knowledgeable and contextually aware of finance-related information.

Data Sources and Collection Methods

Publicly Available Datasets:

Hugging Face Datasets: Leveraged pre-existing datasets from Hugging Face to ensure high-quality, industry-specific data.

Specific Dataset used:

- **Finance Alpaca 1K Test:** A dataset containing finance-related prompts and responses.
- **Content:** Includes 1,000 examples of finance-related questions and answers.
- **Purpose:** Provides a solid foundation for understanding finance-specific dialogues and improving response accuracy
- **Alpaca Finance EN:** Another comprehensive dataset focused on financial dialogues and scenarios.
- **Content:** A comprehensive dataset containing various financial scenarios, user queries, and expert responses.
- **Purpose:** Enhances the model's ability to engage in meaningful financial conversations and provide contextually relevant insights.

Project Workflow for Building the Finance Chatbot

Project Planning and Setup

- **Define Objectives:** Clearly define the objectives of the project, including the desired capabilities of the chatbot.

- **Set Up Environment:** Set up the necessary development environment, including installing required software and libraries such as Python, Hugging Face Transformers, and Streamlit.
- **Data Collection**
 - **Identify Data Sources:** Determine sources for finance-specific data. This can include datasets from Hugging Face, financial news articles, company reports, and market analysis reports.
 - **Collect Datasets:** Gather the selected datasets. For example, you might use datasets like [poornima9348/finance-alpaca-1k-test](#) and [ssbuild/alpaca_finance_en](#) from Hugging Face.
 - **Data Preprocessing:** Clean and preprocess the data to ensure it is well-organized and suitable for training. This includes handling missing values, formatting text, and removing irrelevant information.
- **Model Selection**
 - **Choose Pre-trained Model:** Select a pre-trained language model from Hugging Face that is suitable for fine-tuning on finance data. Popular choices include models based on GPT-3, BERT, or similar architectures.
 - **Set Up Training Environment:** Use platforms like Google Colab with T4 GPUs for model training to ensure efficient processing.

- **Fine-Tuning the Model**

Introduction to Fine-Tuning Techniques:

- **PEFT (Parameter-Efficient Fine-Tuning):** Optimize the model by adjusting only a subset of parameters, making the process more efficient.
- **LoRA (Low-Rank Adaptation):** Decompose the model parameters into lower-rank matrices for efficient fine-tuning.
- **QLoRA (Quantized LoRA):** Combine quantization with LoRA to further reduce memory usage and computation.
- Fine-tune the selected pre-trained model on the finance datasets using PEFT, LoRA, and QLoRA techniques.
- Limit training to a maximum of 25 epochs to ensure feasibility and manage computational resources.
- **Pushing Fine-Tuned Model to Hugging Face:**
 - After fine-tuning, push the updated model to Hugging Face Model Hub for easier access and deployment.
- **Chatbot Development**

Develop Flask Backend:

- Create a Flask application to serve the fine-tuned model and handle user queries.
- Use ngrok to expose the Flask application to the internet for testing and demonstration purposes.

Develop Streamlit Frontend:

- Create a Streamlit application to provide a user-friendly interface for interacting with the chatbot.
- Ensure the Streamlit app can communicate with the Flask backend via ngrok URL.

- **User Interface Design**

- Create User Interface: Design a user-friendly interface using Streamlit that allows users to interact with the chatbot. This should include features like text input, response display, and conversation history.
- Add Features: Implement additional features such as a sidebar for entering the ngrok URL, a button to clear chat history, and prompts to guide user interaction.

- **Deployment**

- Host Backend: Use ngrok to host the Flask backend, making it accessible via a public URL.
- Deploy Frontend: Deploy the Streamlit frontend, ensuring it communicates effectively with the hosted backend.

Conclusion

This project demonstrates the end-to-end process of developing an industry-specific finance chatbot using pre-trained models from Hugging Face. By fine-tuning the model with finance-specific data and leveraging advanced techniques like PEFT, LoRA, and QLoRA, the chatbot is optimized for efficient and accurate responses. The deployment of the chatbot using Flask and Streamlit ensures a user-friendly interaction, while the detailed demonstration showcases its practical application in the finance industry.

The successful development and deployment of this finance chatbot underscore the importance of using advanced AI techniques for efficient fine-tuning. These techniques not only enhance the model's performance but also ensure that it can handle industry-specific queries with precision and contextual relevance.

Future improvements could further enhance the chatbot's capabilities, making it an invaluable tool for industry-specific queries and insights. By continually refining the model and expanding its dataset, the chatbot can stay updated with the latest trends and information, ensuring it remains a reliable and valuable resource in the finance industry. This project lays a strong foundation for the continued advancement of AI-driven conversational agents, highlighting their potential to transform industry-specific communication and support.

Future Scope

- Integrating the chatbot with additional data sources for enhanced accuracy.
- Expanding the bot's capabilities to cover more complex finance-related tasks.
- Adapting the chatbot for other industries to demonstrate its versatility and scalability.