

Project description:

Create an AI assisted resource (a conversational bot) that efficiently collates finance and economy related current news within a specified time period (eg. 1 week) from a YouTube news channel (eg. Yahoo Finance) where people can ask about finance related current news that simultaneously shows the sentiment associated with the context of the given query. The knowledge base for this would be a YouTube News channels (eg. Yahoo Finance) generated within 1 week:

The core technologies used in this project would be :

1. pytube library or Youtube API to download YT videos and generate audio.
2. Whisper openAi for Audio-to-Text conversion for generating text from audio
3. Transcribing + Refining of text
4. Vectorization of YT videos- Creating a vectorized representation of the transcribed text from the YT video using an embedding model like OpenAi embedding model
5. Use a pre-trained hugging face sentiment analysis model to do sentiment analysis on the context of the query.
6. Implementing a RAG framework to retrieve context for the query (Langchain).

Data Description:

1. We have generated audio files from a specified YT news channel (eg. Yahoo Finance) for 1 day using **YouTube API**.
2. We then converted audio to text and stored them in text files by using **OpenAI's Whisper model**.
3. We are converting the text to vectors/embeddings using the OpenAi embedding model either **text-embedding-3-small** model or more powerful **text-embedding-3-large** model (we are still experimenting).
4. And then storing those embeddings in the **Pinecone vector database**. Pinecone is a cloud-native vector database that handles high-dimensional vector data. It is designed to store, index, and retrieve high-dimensional vectors, making it ideal for machine learning applications such as text classification.

Potential Stakeholders

1. End Users: Individuals who use the chatbot to get news summaries and answers to their questions. Their satisfaction and engagement are crucial for the chatbot's success.
2. Analytics and Research Teams: Monitor and analyze the chatbot's performance, user interactions, and feedback to inform improvements and strategic decisions.
3. Advertisers and Sponsors: Businesses or organizations that may partner with the chatbot platform for advertising opportunities, benefiting from its user base and engagement.

Potential KPIs

1. User Satisfaction: Gauged through surveys or feedback mechanisms where users rate their satisfaction with the chatbot's responses and summaries.
2. Accuracy/Error Rate: Measures the correctness of the summaries and answers provided by the chatbot and satisfaction of users with the answers. This could be assessed by comparing the AI's output with trusted sources or user feedback.
3. User Growth Rate: Tracks the rate at which new users are signing up and using the chatbot, reflecting its popularity and market penetration.

