GPT Article Summarizer Project

Solo or team?: Solo
Personal or School?: Personal
Approximate date: December 2023

Why did I do this?

When ChatGPT first became available to the general public, I didn't quite know what to make of it. I thought it was interesting that it responded to conversation like a human, but I didn't see it as a revolutionary technology. That all changed when I realized that it was effectively capable of thinking. Anything I could do with my mind– so long as it was prompted well– it could do too, sometimes better than me. As soon as I found out there was an API that could make LLMs like ChatGPT available in python programs, I was intrigued, and decided to give it a try. AI is great for automating complex tasks that you yourself do not want to undertake, such as reading long texts. This project aimed at summarizing a piece of text into a bullet point summary, drastically lessening the amount of time it would take to understand a given article. You could simply copy and paste text into the ChatGPT web interface, but there is a text/token limit, and all of your text may not be considered. In order to overcome this obstacle, I created this project.

Description

This program takes as input a PDF of text (of any length), and outputs a bullet point summary using OpenAI's GPT API. It breaks the text into manageable chunks, providing a summary of each section. It then summarizes the collection of summaries into one, concise bullet point list. This can be very useful for understanding text you don't have the time to read in full.

Overview

Feel free to visit my portfolio website and check out the code I used to complete this project! In the code there are markdown comments explaining each section. However I will briefly go over some key elements of the project for the purposes of summarization.

- I imported the model and context I wanted from the OpenAI API, as well as used an environment variable to store my API key (the program only works with my permission)
- I separated text into manageable sections (small enough for the GPT model to analyze) using python text-editing and NLP libraries
- I calculated the approximate price of the summary using OpenAI's pricing catalog and a tokenizer. Using OpenAI's API is not a free service, so I want to ensure that the text isn't going to cost a lot to analyze.
- I iteratively prompt the GPT model to summarize the given sections of text in paragraph form, using proper prompting technique to ensure the LLM efficiently carries out the task
- I take approximately the token limit of the GPT model and feed that amount of text to the GPT model at a time to produce bullet point summarizes of the paragraph summarizes, effectively condensing the whole article into 10 or more bullet points