Name: Leland Ling
Email: lling@bu.edu
Status (junior, senior, graduated): Graduated
Link to Project Github:https://github.com/lelandling/eightball

Project Name: Oh all powerful Eightball

1. Brief Explanation of Project: Enter in a question via command line. Metaphor API pulls URLs using the question as the prompt, where then beautifulsoup scrapes webdata and preprocesses it. Individual webdata is then processed into a list using GPT4, where it analyzes the sentiment, key points, and condenses the information. The resulting list is then fed into GPT4 with the original question to summarize overall points. If there are further questions, another round can occur, where the overall summary and a new question is used as context for further summarization/elaboration

2. How you Built it: I started by seeing how the metaphor API worked. I saw that I had a resulting amount of URLs relating to the prompt. Then, I considered what kind of project I wanted to make. I wanted to make some kind of assistant tool so my goal was to pull digested information off the internet. I then planned out my project - pull URLs, scrape URLs, process URL webdata, condense webdata, return summary.

First, I had to figure out how to scrape information off the URLs. I hadn't used beautiful soup before. But through consulting their documentation and using ChatGPT, I figured out how to scrape and process the URL websites.

After, it was a matter of how to use the resulting scraped information to condense into readable information.I decided that GPT4 would be able to do so. I ran into two issues, one being that OpenAI's API only takes a certain number of tokens per minute, and two being that some websites had too much raw text to process due to the same limitation above. As such I just added a wait and truncated the text if it was too long as a work around. This does make the program run pretty long but this can be fixed by changing organization settings through OpenAI support. After I had relevant information I ran it through GPT4 again and then printed the output.

I originally wanted the program to be dictatable, but I ran into some problems with the BARK voice transformer where the text to speech output was subpar and not understandable.

3. Challenges/Feedback on the API: Maybe adding some kind of relevancy parameter to each of the URLs would be useful?

Also if the user wants to continue searching (i.e. because they're not satisfied with the first 10 websites provided by google, they go to the next set of pages) it's kind of hard to continue with Metaphor. With the current API I can imagine a workaround where you omit the previously pulled URLs and call the API again.

4. Why you're interested in Metaphor :)
My prior experience with NLP, LLMs and other machine learning fields excite me day to day to work on relevant projects. I learned a lot during this project, had a lot of fun, and more importantly made something I thought was pretty cool. If being at Metaphor is anything like this project I think I will have a great time there.