

Chat Completion and Function Calling

Take Home Exercise Review

Use either Dall E 3 or Suno to generate an image or audio. Use prompt engineering techniques to get the most quality response.



What is an API?

- API = Application Programming Interface
- Used by developers to access data or take other actions in an application or website.
- In this case, we are going to make an API call (or request) to OpenAI with a prompt and a few other parameters.
 - We will get our generated AI response back from the API call.



API Keys

Most APIs require an API key to make a call.

- To control the number of calls made and ensure security.
- Obtained from the provider's website
- Sometimes there will be a usage cost

It is important to keep your API keys private

- Connected to your account with your personal information
- Store it in a `.env` file in the same directory as your code.
 - Using python libraries, you can load the `.env` file as part of your environment variables.
 - More organized when working with multiple projects and multiple API keys



OpenAI API

- OpenAI's API is fortunately easy to navigate and well documented. Here we can check the documentation and make sure the OpenAI python library is installed on our machines.



Chat Completion API

Chat Completion is an OpenAI API endpoint

- Similar responses to Chat GPT
- Chat section of OpenAI's playground but within code



Chat Completion Parameters

- **Model**
 - Takes a string that signifies the specific model that we intend to use. (gpt-4, gpt-3.5-turbo, etc.)
- **Messages**
 - Takes a list of messages in dictionary format.
 - The “role” key signifies where the message is coming from. (system, user, tool, or assistant)
 - The “content” key signifies the message content (system message, user prompt, or assistant response)
 - The message list can be thought of as the context window.
- **Temperature**
 - Creativity factor (0 strict to training data, 2 creative and sometimes gibberish)



Chat Completion Parameters

- **Max_tokens**
 - Maximum tokens slider in Playground
 - Limits the response to the set amount of tokens.
 - Helps us control the response length
- **Stream**
 - Takes a boolean value
 - Determines whether the response will be returned in a real time stream of data instead of waiting for the full response to return the data.
 - When set to True, there will be a slightly different way of obtaining and displaying the response within your script.



Exercise

Write a simple chat completion API call that prompts GPT 3.5 Turbo to respond with a joke.



Function Calling

Function calling is a way to connect an AI model with external tools.

- With all models, there are limitations to real time information and risks of hallucinations.
- Function calling allows you to connect your own functions within the chat completion API call by using the “tools” and “tool choice” parameters.



Tool Choice

Directs the model on which (if any) function to choose.

- “none” or not defining this parameter makes the model not call a function and acts as a normal chat completion API call.
- “auto” makes the model decide on which function to call based on the context of the prompt.
- For a specific function call, you have to pass the function name in a dictionary format.
 - This will force the model to choose that function every time.



Tools

When “tool choice” is set to “auto” or a specific function name, the model will refer to the “tools” parameter.

- Holds the list of available tools to choose from.
- Formatted in a dictionary
 - Natural language descriptions of the overall function and each argument with its relationship to the function.



Response Object

The API response object will hold the function name and a dictionary of arguments.

- Both are generated based on the context of the prompt and descriptions within the “tool choice” parameter.
- If functions are chosen incorrectly, focus on making better distinctions between functions in “tool choice”.



Parallel Function Calling

Function calling also now supports parallel function calling.

- If the prompt contains multiple actions for the model to take, it will call multiple functions within the same API call.



Exercise

Write three functions and format them properly to be understood by the model as if they were being passed in as the “tool choice” parameter.



Take Home Exercise

Create a function calling chatbot script that supports parallel function calling. Have at least 3 functions listed in your tool calls and try to call them all in one prompt.

