www.imarketing.courses

Knowledge: How GPTs Use Files

Bots have been making use of files since before ChatGPT. Remember the "chat with your PDF" craze?

This is accomplished using a method called "retrieval" where the language model pulls part or all of its docs into its context as it answers the user.

And while we cannot know exactly how GPTs interact with files, there are some assumptions we can make that will help you know how to work with uploaded "knowledge".

These are my best practices:

1. Tell the GPT what it has

GPTs have a set of instructions on their capabilities with web browsing, DALL·E image creation, and python/code interpreter. Whenever these plugins are enabled, those instructions are loaded into the top of your prompt. Consider taking this practice to your knowledge files as well.

For example, I recently consulted on a project and the instruction we ended up going with was similar to this:

```
## Data

You are programmed to perform an embedding search to sift through
comprehensive knowledge base documents and retrieve the most relevant
information. You may assume any information you retrieve is 100% true. For
all other knowledge, rely only on facts you have a greater than 95%
confidence level in. If you're unsure, say so. If you don't know
something, let the user know "I don't know" rather than making something
up. Your responses should be concise, accurate, and tailored to the user's
question.
```

2. Tell the GPT to use the knowledge

It sounds obvious, but it isn't necessarily obvious to the GPT.

Just because you uploaded a file into the knowledge, it doesn't know how to use that file.

And while the GPT may automatically pull parts of that file in while answering the user, with no specific instructions, that data is just floating above, hoping to be useful.

www.imarketing.courses

Even 1-2 sentences here can allow the GPT's behavior to shift and be more consistent:

```
## Knowledge

In your knowledge you have a text file that contains numerous articles on
topics the user may want to explore. Refer to these articles to improve
your answers.
```

3. Assume the GPT may retrieve only chunks or excerpts

Based on my research and experiments, it appears that GPTs have multiple methods for retrieving docs. The system itself gets to decide what method to use based on some characteristics of the doc as well as characteristics of the user's input.

We can't see inside the system to understand how this works exactly, but it's helpful to know.

Shorter knowledge files seem to have a higher chance of being fully retrieved.

Files that contain large lists also seem to get fully retrieved more often.

Large files are not surprisingly more likely to only have relevant parts brought back, which is usually the right call anyhow.

One of the takeaways here is that you may be able to influence this behavior from within the GPT's instructions. So, if your GPT isn't behaving as planned, you might try something like this:

```
## Knowledge

Any time the user's inquiry requires you to review files in your
knowledge, always pull in the entire file and review it top to bottom.
```

I can't guarantee this will work every time, but it's an area worth exploring, especially for GPTs that use multiple larger files.

4. GPT can save new files

We talk a bit more about this in another lesson, but let's discuss it here as well.

If you ask, a GPT will save a file for you in a place where it can access it.

www.imarketing.courses

Telling a GPT to save a file serves two main purposes:

1. So you can download the file
2. So it can write to the file and read its contents later

You can use the second purpose as a hack to get the GPT to save and reference certain information almost like a database. This method isn't well-explored yet, so I don't have a great example, but I've seen people doing it.

Note that files created and saved during your session are tied to that session. If you clear the chat with the GPT and start again, your new conversation will not have those other downloads.

Should You Put Data in Knowledge or in the Prompt?

Before you get too crazy about uploading knowledge files, I want to pose a question to you:

How much information is there really in there?

I ask because your instructions can handle thousands of words. So, say you want to create a GPT that uses an exaggerated amount of Gen Z slang. (Indeed, there is one.) You might want to include a dictionary of the top 25 slang terms to use.

You could upload that as a file, or you could simply paste all 25 slang terms into the prompt in their own section. This really isn't a lot of information, and it seems inefficient to me to force the GPT to retrieve it from an external source, especially if it will be using that info on every turn of the conversation.

Even if you have lists of hundreds of parameters, or multi-page FAQs, I encourage you to try both approaches. Pasting the data directly into the GPT's instructions may produce more consistently good outputs because it will have that information in every single conversation. Worth a try.

Takeaway

Uploading knowledge into a GPT can be a powerful way to add specific information into your chat sessions.

Knowledge can be anything: a reference image for generating other images, a list of terms of FAQs, articles and book excerpts, even information from your company website or product marketing materials.

www.imarketing.courses

It's worth it to take a second to think about whether you have any files that might be useful for the GPT. Some of the best magic seems to happen when you give it these sets of specific information.