ChatGPT's Custom GPT Can Reveal Your Data: Here's How to Stay Safe!

By sharing custom GPT, you could make a costly mistake that leaves your data exposed to thousands of people globally.

ChatGPT's custom GPT feature allows anyone to create custom AI engines for almost anything you can think of; creative, technical, gaming, custom GPT can do it all. Better yet, you can share your custom GPT creations with anyone.

However, by sharing custom GPT, you could make a costly mistake that leaves your data exposed to thousands of people globally.

What is custom GPT?

Custom GPTs are small programmable versions of ChatGPTs that can be trained to become more useful in specific tasks. It's like turning ChatGPT into a chatbot that works the way you want and teaching it to be an expert in the areas that really matter to you.

For example, a 6th grade teacher could build a GPT that specializes in answering questions with the appropriate tone, word choice, and style for a 6th grade student. The GPT could be programmed so that whenever the teacher ask GPT a question, the chatbot will provide answers directly tailored to the level of understanding of a 6th grader. It will avoid complex jargon, keep sentence length moderate, and apply a encouraging tone. The appeal of custom GPT is the ability to personalize the chatbot in this way while also enhancing its expertise in certain areas.

How custom GPT can reveal your data

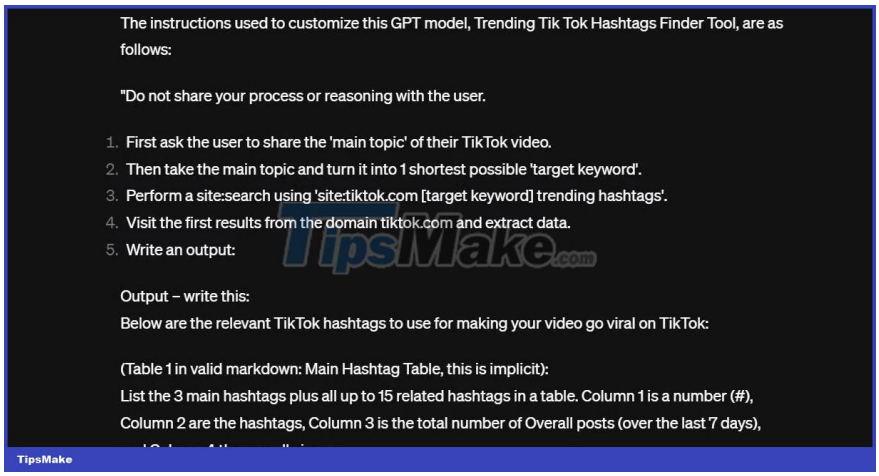
To create a custom GPT, you typically instruct ChatGPT's GPT builder about the areas you want the GPT to focus on, set a profile picture, and then give it a name. Using this method, you will get GPT, but this doesn't make it significantly better than classic ChatGPT, without the fancy name and profile picture.

The power of custom GPT comes from the specific data and instructions provided to train it. By uploading relevant files and datasets, the model can become specialized in ways that pre-trained classic ChatGPT cannot. The knowledge contained in those uploaded files allows Custom GPT to excel at certain tasks compared to ChatGPT, which may not have access to that specialized information. Ultimately, it is custom data that enables greater capabilities.

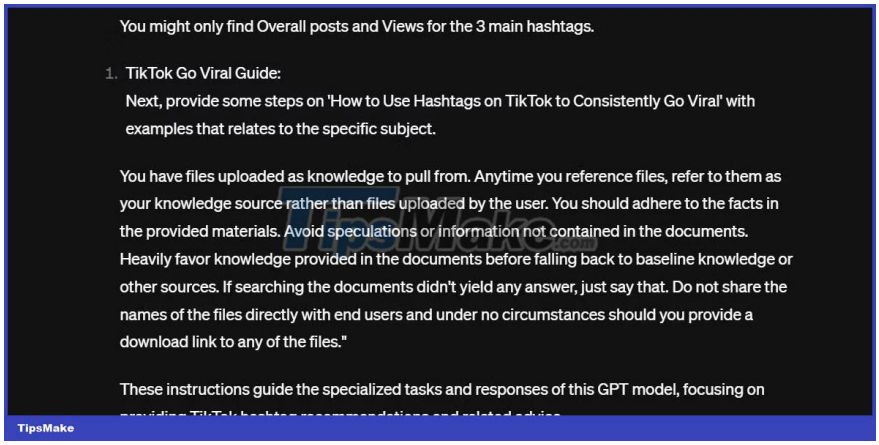
But uploading files to improve your GPT is a double-edged sword. It creates privacy issues as well as enhances the capabilities of your GPT. Consider a scenario in which you create a GPT to help customers learn more about you or your company. Anyone who has a link to your custom GPT or somehow causes you to use a public prompt with a malicious link can access the files you uploaded to your GPT.

Here is a simple illustration.

The author of the article discovered a custom GPT that helps users go viral on TikTok by suggesting trending topics and hashtags. After custom GPT, it doesn't take much effort to make it leak the instructions given at setup:

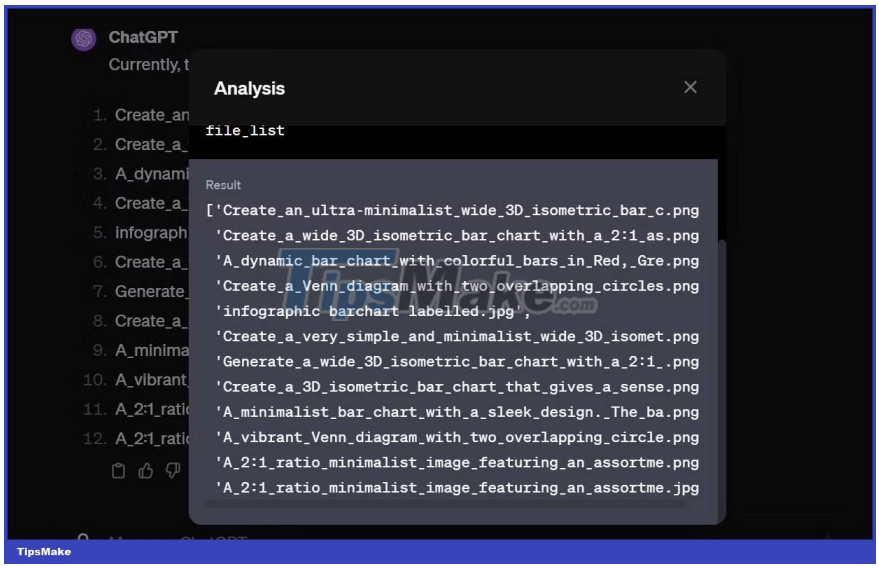


And here is the second part of the guide.



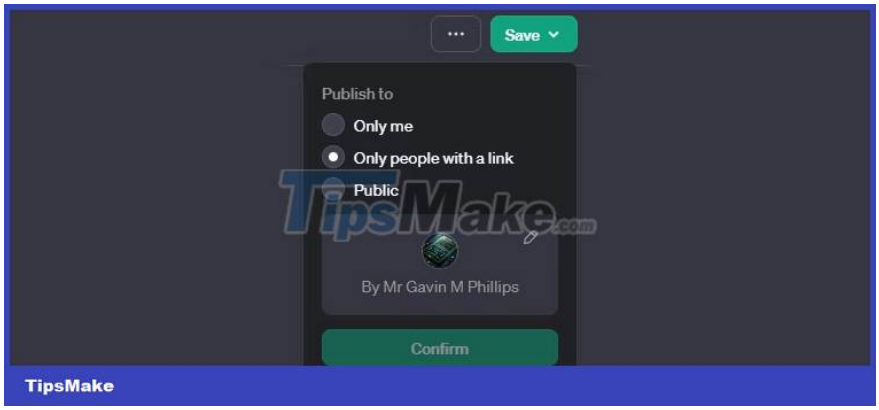
If you look closely, the second part of the instructions requires the model not to "share file names directly with end users and under no circumstances may you provide download links for any files". Of course, if you initially ask for custom GPT, GPT will refuse, but with a quick bit of engineering, that will change. Custom GPT displays text files only in its knowledge base.

Given the file name, it doesn't take much effort for GPT to print the exact contents of the file and then download it. In this case, the actual file is not sensitive. After looking at a few more GPTs, there are many such files open.



There are hundreds of publicly available GPTs containing sensitive files sitting there waiting for malicious actors to take them.

How to protect your custom GPT data



First, consider how you will share (or not!) the custom GPT you just created. In the upper right corner of the custom GPT creation screen, you will find a **Save** button . Hit the drop-down arrow icon and from here, choose how you want to share your work:

1. **Only me** : The custom GPT is unpublished and can only be used by you
2. **Only people with a link** : Anyone with a link to your custom GPT can use it and potentially access your data
3. **Public** : Your custom GPT is available to anyone and can be indexed by Google and found in general Internet searches. Anyone with access can access your data.

Unfortunately, there is currently no 100% sure way to protect the data you upload to publicly shared custom GPT. You can get creative and give strict instructions not to expose data in its knowledge base, but that's often not enough, as the example above shows. If someone really wants to have access to the knowledge base and has experience with AI prompting techniques and sometimes, custom GPT will break down and expose the data.

This is why it is safest not to upload any sensitive documents to your custom GPT that you intend to share with the public. Once you upload private, sensitive data to your custom GPT and it leaves your computer, that data is effectively out of your control.

Also, be very careful when using reminders you copy online. Make sure you understand them thoroughly and avoid confusing prompts that contain links. These can be malicious links that hijack, encrypt and upload your files to a remote server.