

UN BUEN DATA SET

Para crear un GPT realmente efectivo, es fundamental entender cómo funciona la sección "Conocimiento" en la interfaz de ChatGPT y cómo crear un buen data set. Piensa en el "Conocimiento" como una mochila que tu GPT lleva consigo, llena de información que puede consultar cuando lo necesite.

Mochila de Documentación

Esta "mochila de documentación" es lo que técnicamente se conoce como Retrieval Augmented Generation (RAG). Sin embargo, OpenAI no nos da demasiada información sobre qué tipo de archivos funcionan mejor para los GPTs. Un truco es intentar preguntarle al constructor de GPTs en ChatGPT qué tipos de archivos son aceptados, aunque la información que proporciona no es exhaustiva. A través de la experimentación, hemos descubierto que el GPT puede tener hasta 20 archivos en su "mochila", con un límite de 512 MB por archivo. La elección de qué tipo de archivos usar dependerá del tipo de información que queremos que nuestro GPT maneje.

Tipos de Archivos

- **PDF o Word:** Si bien estos archivos son muy comunes en el día a día, no son la mejor opción para alimentar a los GPTs. Esto se debe a que contienen mucha información superflua que puede dificultar la extracción del texto relevante. Lo ideal es que la información esté limpia y organizada. Una buena práctica es separar el texto de las imágenes, guardar el texto en formato TXT y nombrar las imágenes de forma descriptiva para que el GPT pueda encontrarlas fácilmente.
- **TXT / Markdown:** Los archivos TXT son archivos de texto plano, sin formato. El lenguaje Markdown nos permite darle cierta estructura a estos archivos, lo cual ayuda al GPT a interpretar la información de forma más eficiente. Se pueden usar archivos Markdown cortos directamente en ChatGPT, pero si son muy extensos, lo mejor es convertirlos a TXT para evitar que sean ignorados. Un truco útil es copiar el texto en formato Markdown, pegarlo en un archivo TXT y luego subir el TXT a ChatGPT.
- **Imágenes:** OpenAI permite subir imágenes a los GPTs, aunque su uso es limitado. Funcionan como una base de datos visual a la que el GPT puede acceder. Por ejemplo, si quieres crear un GPT que recomiende recetas de cocina, puedes subirle fotos de platos e indicarle al GPT que solo proponga recetas basadas en las fotos que tiene.
- **Archivos ZIP:** ChatGPT puede procesar archivos ZIP si se habilita la opción "Interpreta el código y analiza datos" en la sección "Funciones". Sin embargo,

el GPT necesita descomprimir el archivo para acceder a su contenido, lo que puede ralentizar su funcionamiento. Un truco interesante es crear varios GPTs, cada uno en un archivo TXT con un tema específico, y luego comprimirlos en un archivo ZIP. De esta forma, el GPT principal puede actuar como un distribuidor de tareas, eligiendo el GPT más adecuado para responder a la pregunta del usuario.

La importancia de un buen dataset

Un buen data set es fundamental para un GPT efectivo. La clave está en que la información sea:

- **Bien estructurada:** Utiliza el lenguaje Markdown o TXT para organizar la información de forma clara y concisa.
- **Elaborada:** No tengas miedo de darle mucha información al GPT. Cuanto más sepa, mejor responderá.
- **Rica en ejemplos:** Proporciona al GPT ejemplos concretos de lo que quieres que haga.
- **Incluye ejemplos de lo que NO se debe hacer:** Esto le ayudará al GPT a entender mejor tus expectativas.

Tip: Una buena estrategia es iterar con el GPT a medida que lo entrenas. Observa sus respuestas, guarda las que te parezcan buenas y las que no, e inclúyelas en la documentación para que el GPT aprenda de sus propios aciertos y errores.

Recuerda, crear un GPT personalizado es un proceso continuo. A medida que le proporcionas más información y lo entrenas con ejemplos, su capacidad para realizar tareas complejas y generar outputs precisos irá mejorando.