

Prompt Engineering 4 Gen. Al



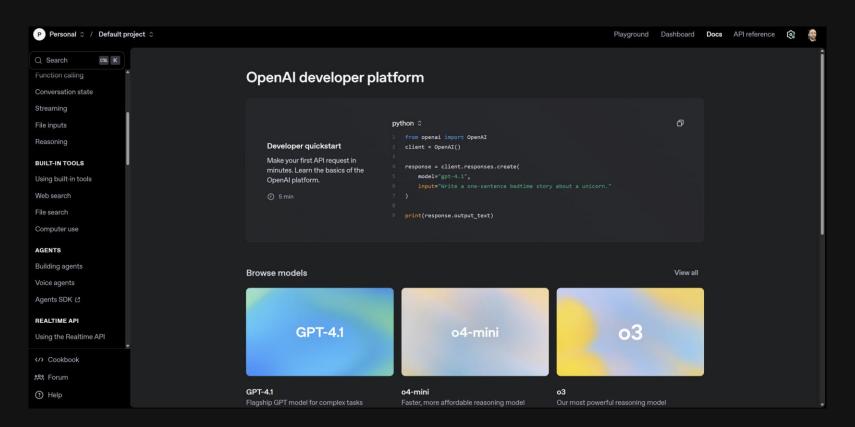


Agentes IA Consola Prompts









Open AI - Models



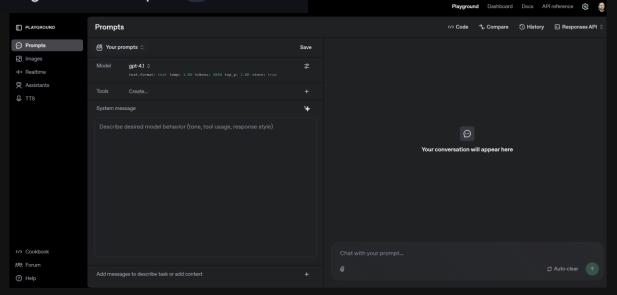
Pausa

Models overview	
The OpenAl API is powered by a diverse set of models with different capabilities and price points. You can also make customizations to our models for your specific use case with fine-tuning.	
MODEL	DESCRIPTION
	Our high-intelligence flagship model for complex, multi-step tasks
	Our affordable and intelligent small model for fast, lightweight tasks
GPT-4 Turbo and GPT-4	The previous set of high-intelligence models
GPT-3.5 Turbo	A fast, inexpensive model for simple tasks
DALL-E	A model that can generate and edit images given a natural language prompt
TTS	A set of models that can convert text into natural sounding spoken audio
Whisper	A model that can convert audio into text
Embeddings	A set of models that can convert text into a numerical form
Moderation	A fine-tuned model that can detect whether text may be sensitive or unsafe
	A set of models without instruction following that can understand as well as generate natural language or code
Deprecated	A full list of models that have been deprecated along with the suggested replacement

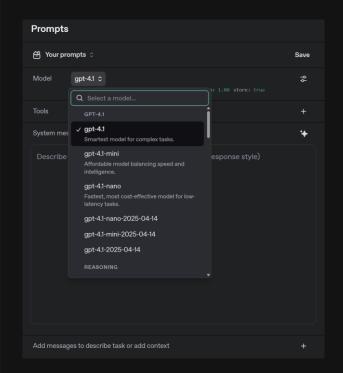
https://platform.openai.com/docs/models

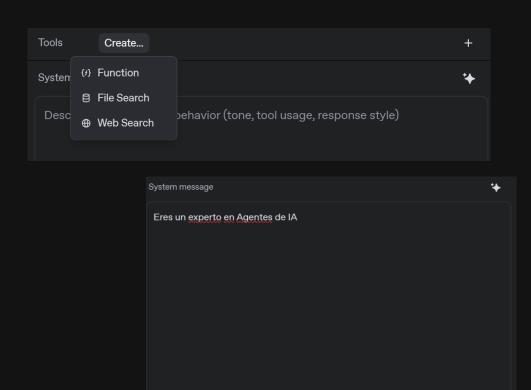


Prompt engineering is the process of crafting or structuring instructions (prompts) to elicit the desired output from a generative AI model. It involves providing context, instructions, and examples to guide the model towards a specific response. Effective prompts can improve the accuracy, relevance, and coherence of AI-generated text, images, or other outputs.

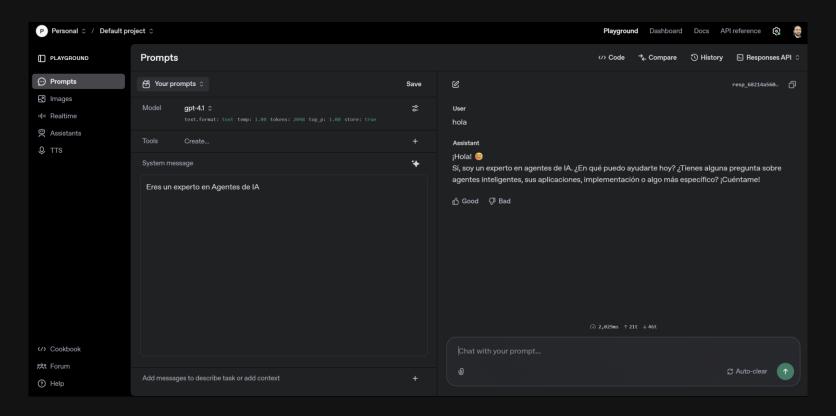




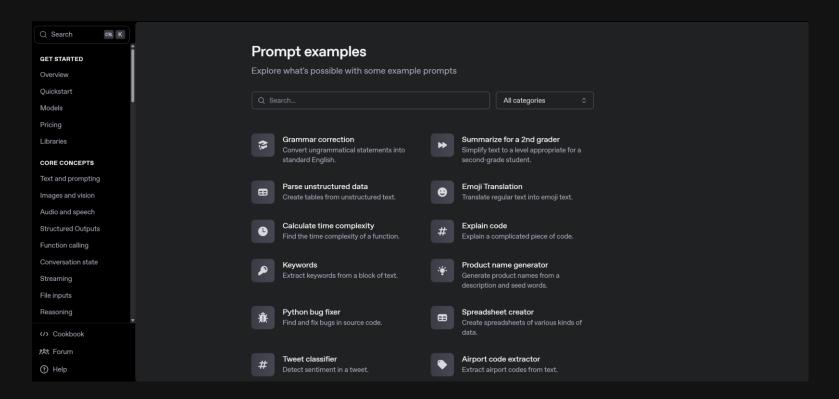












https://platform.openai.com/docs/examples

Entender las configuraciones



model="gpt-4",
messages=messages
temperature=0,
max_tokens=600,
top_p=0.95,
frequency_penalty=0,
presence_penalty=0,
stop=None

- model="gpt-4": Esta variable especifica el modelo de lenguaje a utilizar. En este caso, se está indicando el uso de "GPT-4", que es una versión avanzada de los modelos generativos preentrenados de OpenAI.
- messages=messages: Esta variable representa los mensajes o entradas que se le proporcionan al modelo. Puede ser una lista de interacciones anteriores (como preguntas y respuestas) que ayudan al modelo a entender el contexto de la conversación actual.
- 3. temperature=0: La "temperatura" controla la aleatoriedad de las respuestas generadas por el modelo. Un valor de 0 significa que el modelo será muy conservador y predecible en sus respuestas, eligiendo las opciones más probables.
- 4. max_tokens=600: Esta variable define el número máximo de "tokens" (palabras o partes de palabras) que el modelo generará en su respuesta. Aquí, se establece un límite de 600 tokens, lo que permite respuestas bastante extensas.
- top_p=0.95: Conocido como "nucleus sampling", este parámetro controla la diversidad de la respuesta generada. Un valor de 0.95 significa que el modelo solo considerará las opciones de palabras que constituyen el 95% más probable, lo que permite cierta creatividad manteniendo relevancia.

Entender las configuraciones



model="gpt-4",
messages=messages
temperature=0,
max_tokens=600,
top_p=0.95,
frequency_penalty=0,
presence_penalty=0,
stop=None

- frequency_penalty=0: Esta variable reduce la probabilidad de que el modelo repita el mismo texto. Un valor de 0 significa que no hay penalización por frecuencia, permitiendo que el modelo repita palabras o frases.
- 7. **presence_penalty=0**: Similar al anterior, pero se enfoca en penalizar la presencia reiterada de palabras o frases. Un valor de 0 indica que no hay penalización, permitiendo que ciertas palabras o temas se repitan en la respuesta.
- 8. stop=None: Este parámetro permite definir un "token de detención" que le indica al modelo cuándo detener la generación de texto. "None" significa que no hay un token de detención específico, y el modelo seguirá generando texto hasta alcanzar otro criterio de detención, como el límite de tokens.

Cada una de estas variables juega un papel importante en cómo el modelo de lenguaje genera respuestas, permitiendo ajustar la longitud, originalidad, y coherencia de las mismas.



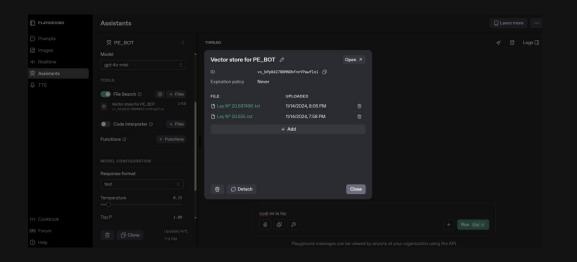
Agentes IA Consola Assistants







Un almacén de vectores (vector store) es un sistema especializado diseñado para almacenar y recuperar eficientemente datos representados como vectores, que son representaciones matemáticas de datos. Son fundamentales en aplicaciones de inteligencia artificial y aprendizaje automático, permitiendo búsquedas eficientes basadas en similitud, en lugar de búsquedas exactas de palabras clave.

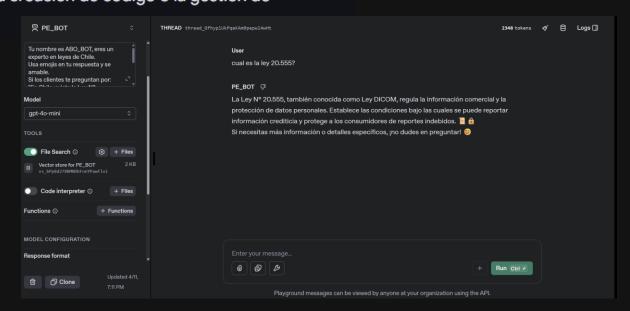


https://platform.openai.com/playground/assistants



Un asistente de IA (Inteligencia Artificial) es un software que utiliza la inteligencia artificial para ayudar a los usuarios a realizar tareas, responder preguntas y gestionar información. Estos asistentes pueden variar desde asistentes virtuales personales como Siri o Alexa, hasta herramientas más especializadas en tareas como la creación de código o la gestión de

proyectos.



https://platform.openai.com/playground/assistants





https://platform.openai.com/log



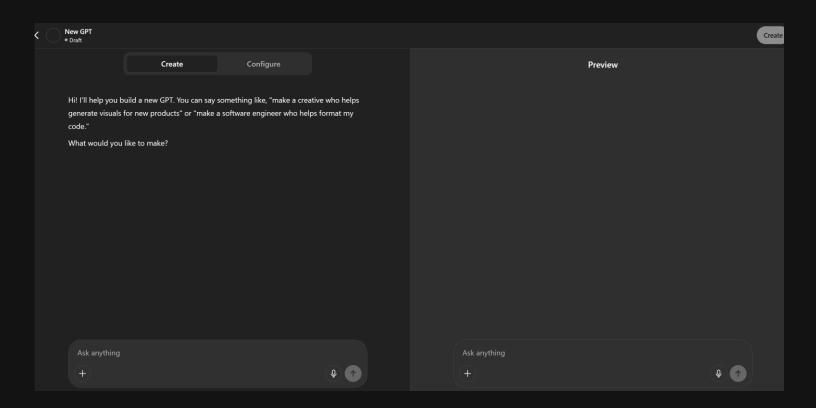
Agentes IA Consola GPTs





OpenAI - GPTs





https://chatgpt.com/gpts/





Prompt Engineering 4 Gen. Al

