1

API DE OPENAI

Ejecutar los modelos de ChatGPT en tu propio sitio web con JavaScript, PHP u otro lenguaje, o dentro de una aplicación que has desarrollado, permite dar un salto de calidad, al mejorarla o brindar una mejor manera de interacción con el usuario. En este libro, hasta el lector más principiante comprenderá los pasos para usar la API de OpenAI para GPT y liberar el potencial para sus aplicaciones, productos o servicios.

1.1 ¿QUÉ ES LA API DE OPENAI?

Figura 1.1. Integrar la IA mediante su API, usando uno de varios lenguajes de programación compatibles, es una solución muy simple para entregar nuevas funcionalidades con modelos GPT a usuarios o clientes. Pero, además, es mucho más rápido y económico de lo que la mayoría podría imaginar.

ChatGPT realmente ha evolucionado como la próxima base para crear aplicaciones basadas en IA y ha tomado un gran impulso cuando Microsoft comenzó a invertir en esta tecnología, además de incorporarla a su motor de búsqueda Bing.

Desde la **automatización de tareas**, la ayuda en el ámbito comercial, la atención al cliente, la redacción y la adaptación de textos, ChatGPT ha demostrado su utilidad en una amplia gama de aplicaciones.

Los **modelos** GPT pueden mejorar la experiencia del usuario de sitios y aplicaciones web, traducir, resumir, responder preguntas y realizar muchas otras tareas.

Los **modelos GPT** (transformador generativo preentrenado) de **OpenAI** han sido entrenados para comprender el lenguaje natural y el código. Los GPT proporcionan salidas de texto en respuesta a sus entradas. Las entradas a los GPT también se conocen como **prompts**. Diseñar un prompt es esencialmente como "programar" un modelo GPT, generalmente proporcionando instrucciones o algunos ejemplos referidos al modo de completar una tarea con éxito. Los GPT se pueden

usar en una gran variedad de actividades, incluida la generación de contenido o código, resúmenes, conversaciones, escritura creativa, y más.

La **API de OpenAI** es una interfaz mediante la cual los desarrolladores pueden interactuar con el modelo de GPT a través de la programación. Permite la integración de las capacidades conversacionales de ChatGPT en aplicaciones, plataformas o sistemas, lo que facilita la creación de experiencias dinámicas e interactivas impulsadas por IA.

Figura 1.2. Si bien ambos conceptos se usan de manera indistinta, y se entiende a qué se hace referencia, técnicamente la API es de los modelos de OpenAI, no es la API de ChatGPT, ya que ChatGPT es una implementación en formato de chat conversacional de estos modelos.

La API que vas a usar no es de ChatGPT, sino que se conecta a los modelos que hay detrás de dicha herramienta, y puedes usarla para tu propio chat y para muchas otras tareas.

Usando la API de OpenAI, los desarrolladores pueden aprovechar el poder de la comprensión y generación del lenguaje natural de ChatGPT para crear chatbots, asistentes virtuales, generadores de contenido, y mucho más.

La API de OpenAI funciona administrando solicitudes con el texto de entrada necesario y obteniendo una respuesta que incluye el texto generado. El **endpoint** de la API se encarga de manejar los aspectos computacionales, lo que permite a los desarrolladores concentrarse en utilizar las respuestas generadas para sus casos de uso específicos.

1.1.1 ¿Cómo se usa la API de OpenAI?

El uso de la API de OpenAI implica una serie de pasos que garantizan una integración fluida y un manejo eficaz. A continuación, recorreremos el proceso paso a paso.

PASO 1

Obtener una clave API

Para comenzar a utilizar la API de OpenAI, el primer paso es obtener una clave de API. Puedes adquirir fácilmente las claves API de OpenAI registrándote o iniciando sesión en la plataforma oficial de OpenAI.

Estas claves te otorgarán acceso a la API de OpenAI y te permitirán realizar solicitudes de API.

PASO 2

Elegir un lenguaje de programación

La API de OpenAI proporciona kits de desarrollo de software (SDK) y bibliotecas para varios lenguajes de programación, como Python y Node.js. Dependiendo de tu familiaridad y preferencias, elige el lenguaje de programación que más te convenga. Estos **SDK** y bibliotecas simplificarán el proceso de integración y te facilitarán la interacción con la API.

PASO 3

Trabajar directamente con el endpoint o usar una biblioteca de API

Una vez que tengas tu clave de API y hayas seleccionado un lenguaje de programación, tienes dos opciones para interactuar con la API de OpenAI: puedes usar directamente el endpoint de la API, que te permite realizar solicitudes HTTP a la API o, como alternativa, puedes utilizar una biblioteca en el lenguaje seleccionado de programación, que proporciona una interfaz de nivel superior para acceder a la funcionalidad de la API.

Esta biblioteca abstrae los detalles de hacer llamadas API y simplifica el proceso de desarrollo. Por lo general, proporciona funciones bien detalladas que facilitan su uso.

Hay varios kits de desarrollo de software (SDK) que puedes utilizar para integrar la API de ChatGPT en un proyecto. OpenAI proporciona SDK oficiales para una variedad de lenguajes de programación, incluidos Python, Java y JavaScript, pero hay muchas comunidades de desarrollo conocidas que han generado excelentes SDK para muchos otros lenguajes.

PASO 4

Configurar el entorno de desarrollo

Antes de comenzar a usar la API de OpenAI, debes configurar el entorno de desarrollo. Esto implica configurar su SDK seleccionado e instalar las dependencias necesarias. También es recomendable configurar un entorno virtual para asegurar que tu proyecto API de OpenAI esté aislado de otras tareas en el sistema.

PASO 5

Realizar solicitudes de API

Ahora que has configurado el entorno de desarrollo, puedes comenzar a realizar solicitudes de API a la API de OpenAI. Puedes usar la API para generar texto, responder preguntas e, incluso, crear chatbots.

Una vez que pudiste hacer uso de la API, hay mucho por realizar, dependiendo del uso y la integración que quieras lograr con ella.

1.1.2 Usar lenguaje de marcado de chat (ChatML)

Si usas la API para generar algún tipo de asistente o bot, es común utilizar el lenguaje de marcado de chat (ChatML), que permite formatear la entrada y salida de las solicitudes a la API de OpenAI. Puedes usar ChatML para agregar formato, incorporar imágenes e incluir otros elementos en las respuestas de tus chatbots, por ejemplo.

1.1.3 Experimentar con la API de OpenAI

La API de OpenAI abre un mundo de posibilidades para tus aplicaciones. Depende de tu imaginación y creatividad explorar formas innovadoras de utilizar y ajustar la API de OpenAI para satisfacer las necesidades específicas de tu industria o aplicación.

Figura 1.3. La integración de tecnología de IA mediante una API permite hacer uso de nuevas funcionalidades y potenciar un software existente con las características de los modelos GPT. Es una tarea relativamente simple para un programador en la actualidad.

Los beneficios de usar la API en tus proyectos son varios, pero pueden resumirse en dos principales:

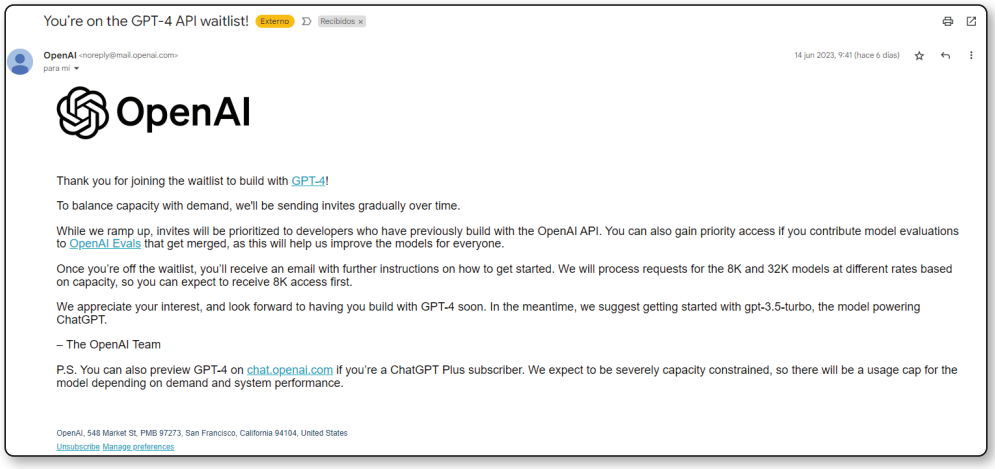
1. **Experiencia de usuario mejorada**: al incorporar la API de OpenAI en tus aplicaciones o servicios, puedes ofrecer experiencias de usuario más atractivas e interactivas. Los usuarios pueden comunicarse con tu producto o plataforma utilizando lenguaje natural, haciendo que las

interacciones sean más intuitivas y personalizadas. Ya sea un chatbot de atención al cliente, un asistente virtual o un generador de contenido impulsado por IA, la API de OpenAI agrega un toque conversacional que mejora la satisfacción del usuario.

2. **Eficiencia de tiempo y costos**: desarrollar IA conversacional desde cero puede ser un proceso que requiere mucho tiempo y recursos. Sin embargo, al aprovechar la API de OpenAI, puedes reducir significativamente el tiempo de desarrollo y los costos asociados. El modelo y la infraestructura previamente entrenados proporcionados por la API de OpenAI te permitirán concentrarte en ajustar el modelo para tu caso de uso específico, en vez de construir todo desde cero.

1.1.4 ¿Y la API de GPT-4?

GPT-4 de OpenAI es el último modelo de lenguaje innovador, capaz de generar texto similar al humano y de ayudar en una variedad de tareas. Muchos usuarios están ansiosos por acceder a su poder para uso comercial, pero es importante comprender el proceso para hacerlo.


You're on the GPT-4 API waitlist! Externo Recibidos ×

OpenAI <noreply@mail.openai.com> 14 jun 2023, 9:41 (hace 6 días)
para mí

OpenAI

Thank you for joining the waitlist to build with GPT-4!

To balance capacity with demand, we'll be sending invites gradually over time.

While we ramp up, invites will be prioritized to developers who have previously build with the OpenAI API. You can also gain priority access if you contribute model evaluations to OpenAI Evals that get merged, as this will help us improve the models for everyone.

Once you're off the waitlist, you'll receive an email with further instructions on how to get started. We will process requests for the 8K and 32K models at different rates based on capacity, so you can expect to receive 8K access first.

We appreciate your interest, and look forward to having you build with GPT-4 soon. In the meantime, we suggest getting started with gpt-3.5-turbo, the model powering ChatGPT.

– The OpenAI Team

P.S. You can also preview GPT-4 on chat.openai.com if you're a ChatGPT Plus subscriber. We expect to be severely capacity constrained, so there will be a usage cap for the model depending on demand and system performance.

OpenAI, 548 Market St, PMB 97273, San Francisco, California 94104, United States
Unsubscribe Manage preferences

Figura 1.4. Al unirte a la lista de espera, es crucial proporcionar detalles correctos y precisos. OpenAI se basa en esta información para priorizar el acceso y garantizar un proceso de incorporación fluido para todos los usuarios.

Para acceder a GPT-4 a través de la API, debes unirte a la lista de espera. OpenAI está trabajando diligentemente para proporcionar acceso a tantos usuarios como sea posible, pero ten en cuenta que puede pasar algún tiempo antes de que todos tengan acceso. Si estás interesado en utilizar GPT-4 con fines comerciales, es esencial agregar tu nombre a la lista de espera de la API.

Antes de unirte a la lista de espera, tienes que crear una cuenta de OpenAI, si no la tienes ya. Después de hacerlo, busca la opción para unirte a la lista de espera de la API GPT-4.

1.1.5 ChatGPT Plus: acceso GPT-4 con un límite de uso

Si eres suscriptor de ChatGPT Plus, ya tienes acceso a GPT-4, aunque con ciertas limitaciones. ChatGPT Plus es un servicio de suscripción que permite a los usuarios acceder a GPT-4 a través del chatbot ChatGPT. Los suscriptores pueden disfrutar de los beneficios de GPT-4 en **chat.openai.com** con un límite de uso. Si deseas experimentar las capacidades de GPT-4 mientras esperas el acceso para la API, puedes explorar GPT-4 Playground. OpenAI proporciona un entorno de juegos donde puedes interactuar con GPT-4 y probar sus capacidades de generación de lenguaje.

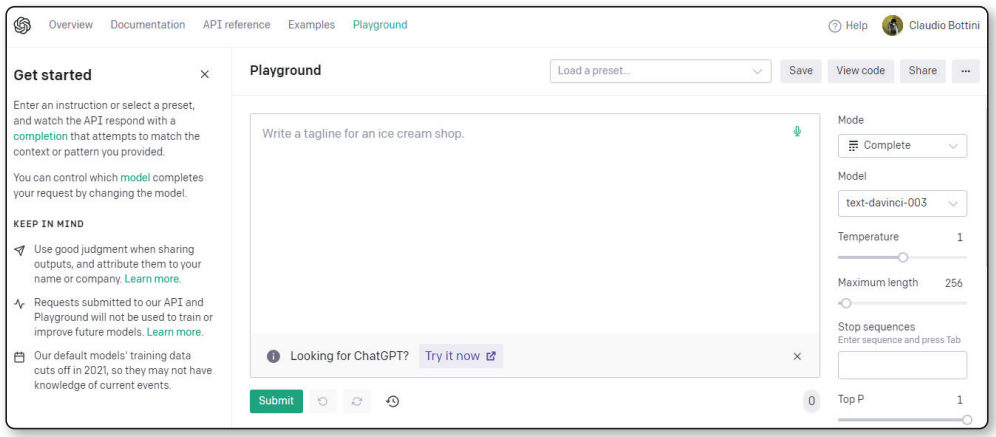


Figura 1.5. Para acceder a GPT-4 Playground, visita el sitio web de OpenAI y ve a la sección de juegos. Allí puedes ingresar indicaciones y observar las respuestas de GPT-4. Aunque GPT-4 Playground no ofrece el mismo nivel de personalización e integración que la API, es una herramienta valiosa para echar un vistazo a las capacidades de GPT-4 y generar texto para uso personal.

1.2 CONCEPTOS IMPORTANTES: PROMPTS, TOKENS, INCRUSTACIONES Y MODELOS

Existen cuatro conceptos que verás continuamente si trabajas con la IA de OpenAI. Conocer qué significa cada uno y sus variaciones te ayudará a hacer un mejor uso de la herramienta. OpenAI ofrece una gama de modelos con diferentes niveles de potencia adecuados para distintas tareas, así como la capacidad de ajustarlos y crear tus propios modelos personalizados.

1.2.1 Prompts

La entrada o información que le ofreces a los sistemas GPT es la forma de "programar" qué quieres que haga el modelo, generalmente proporcionando algunas instrucciones o ciertos ejemplos. Por la manera en que funcionan estos modelos de IA, el modo, la redacción o el formato de información que les brindes en los prompts serán fundamentales para que este "comprenda" tus necesidades, y se llegue a un resultado dentro de lo esperado.

1.2.2 Tokens

Los modelos de OpenAI entienden y procesan el texto dividiéndolo en tokens, que pueden ser palabras o simplemente fragmentos de caracteres. Por ejemplo, la palabra "hamburguesa" se divide en las fichas "ham", "bur" y "guesa", mientras que una palabra corta y común como "pera" es un solo token. Además, en una oración, el primer token de cada palabra generalmente comienza con un carácter de espacio.

OpenAI ofrece una herramienta en línea para hacer cálculos y comprender el uso de **tokens** que puede ayudarte a saber cuántos hay dentro de un texto dado, la encuentras en **esta dirección**.

La cantidad de tokens procesados en una solicitud de API determinada depende de la longitud de sus entradas y salidas. Como regla general, 1 token tiene aproximadamente 4 caracteres o 0,75 palabras para texto en inglés. Una limitación para tener en cuenta es que la solicitud de texto y la finalización generada, combinadas, no deben superar la longitud máxima de contexto del modelo (para la mayoría de los modelos, esto es 2048 tokens, o alrededor de 1500 palabras).

1.2.3 Incrustaciones (embeddings)

Una incrustación es una representación vectorial de una parte de los datos (por ejemplo, un texto) que pretende preservar aspectos de su contenido y/o su significado. Los fragmentos de datos que son similares de alguna manera tenderán a tener **incrustaciones** más cercanas entre sí que los datos no relacionados. OpenAI ofrece modelos de incrustación de texto que toman como entrada una cadena de texto y producen como salida un vector de incrustación. Las incrustaciones son útiles para la búsqueda, agrupación, recomendaciones, detección de anomalías, clasificación y toda tarea de "comparación". Están presentes como parte de la mayoría de los modelos que veremos a continuación.

1.2.4 Modelos

La API está impulsada por un conjunto de modelos con diferentes capacidades y puntos de precio. GPT-4 es el último modelo, y el más potente. GPT-3.5-Turbo es el modelo que impulsa ChatGPT y está optimizado para formatos conversacionales.

1.2.4.1 ELEGIR EL MODELO DE LENGUAJE ADECUADO

Los modelos están clasificados primero en su serie, que los distingue entre sí según la tarea principal para la que fueron creados. Es así que la primera clasificación que tenemos es:

Serie del modelo	Características
GPT-4 (en beta limitada)	Conjunto de modelos que mejoran GPT-3.5 y pueden comprender y generar lenguaje natural o código.
GPT-3.5	Conjunto de modelos que mejoran GPT-3 y pueden comprender y generar lenguaje natural o código.
DALL·E (en beta libre)	Modelo que puede generar y editar imágenes con un mensaje de lenguaje natural como entrada.
Whisper (en beta libre)	Modelo que puede convertir un audio en su transcripción en texto.
Embeddings	Conjunto de modelos que pueden convertir texto en una forma numérica.
Moderation	Modelo perfeccionado que puede detectar si el texto puede ser confidencial o indebido de acuerdo con las condiciones de uso de la API.
GPT-3	Conjunto de modelos que pueden comprender y generar lenguaje natural.

La serie GPT de modelos de lenguaje de OpenAI ofrece varias opciones para que los desarrolladores elijan, con diferentes capacidades y tamaños.

Si bien GPT-3 y GPT-4 brindan funciones más avanzadas, también requieren más potencia y recursos computacionales. Elegir el modelo de lenguaje adecuado implica evaluar los requisitos de tu proyecto, los recursos disponibles y el rendimiento deseado. OpenAI ofrece varios modelos dentro de su serie GPT, con diferentes capacidades y tamaños. Al seleccionar un modelo, ten en cuenta factores como el rendimiento, la complejidad y el costo.

Dentro de la clasificación antes vista, suele optarse por usar modelos dentro de dos ramas principales en la actualidad: GPT-3.5 y GPT-3, siempre hablando de aquellos que son útiles para las tareas más requeridas. Obviamente, si lo que quieres es generar imágenes, elige el modelo Dall-E sin dudarlo.

Modelos de la serie GPT-3.5: Los modelos GPT-3.5 pueden comprender y generar código o lenguaje natural. El modelo más capaz y rentable en la familia GPT-3.5 es gpt-3.5-turbo, que se ha optimizado para el chat, pero también funciona bien para las tareas tradicionales.

Modelo	Características	Máximo de tokens
gpt-3.5-turbo	El modelo GPT-3.5 más capaz y optimizado para chat a 1/10 del costo de text-davinci-003.	4096
gpt-3.5-turbo-16k	Mismas capacidades que el gpt-3.5-turbomodelo estándar pero con 4 veces el contexto.	16384
texto-davinci-003	Puede realizar cualquier tarea de lenguaje con mejor calidad, resultados más prolongados y seguimiento de instrucciones constante que los modelos curie, babbage o ada. También admite algunas funciones adicionales, como la inserción de texto.	4097
texto-davinci-002	Capacidades similares a text-davinci-003pero entrenadas con ajuste fino supervisado en lugar de aprendizaje por refuerzo.	4097
código-davinci-002	Optimizado para tareas de finalización de código.	8001

Modelos de la serie GPT-3: Los modelos GPT-3 pueden comprender y generar lenguaje natural. Fueron reemplazados por los modelos de generación GPT-3.5 más potentes. Sin embargo, los modelos base originales de GPT-3 (davinci, curie,

ada y babbage) son los únicos modelos actuales que están disponibles para entrenar a medida para tus necesidades:

text-davinci-002	Este es el modelo GPT-3 más capaz, pero también el más caro y tiene la latencia más alta. Es mejor para aplicaciones que requieren comprensión y rendimiento avanzados.
text-curie-002	Este modelo GPT-3 es un buen equilibrio entre capacidad y costo. Es adecuado para la mayoría de los casos de uso.
text-babbage-002	Este modelo GPT-3 es menos costoso que curie y davinci pero aun así ofrece un buen rendimiento. Es una buena opción para proyectos con presupuesto limitado.
text-ada-002	Este modelo GPT-3 es el menos costoso y tiene la latencia más baja, pero también es el menos capaz. Es mejor para aplicaciones que no requieren una comprensión del lenguaje muy sofisticada.

En la práctica, la mejor manera de elegir el modelo de lenguaje adecuado es experimentar con diferentes opciones y evaluar su rendimiento en función de cada caso de uso específico.

Recuerda que la API de OpenAI tiene límites de velocidad y costos de uso, así que ten en cuenta la cantidad de solicitudes que envías durante la experimentación. Consulta los precios de la API de OpenAI para obtener más detalles.

1.3 EL COSTO DE USAR LA API

La API de ChatGPT se destaca cada día más porque es una excelente opción para los desarrolladores. Con su gama de planes de precios y opciones de uso flexibles, proporciona una forma eficiente de aprovechar las capacidades de procesamiento de lenguaje natural de última generación.

El precio de la API de OpenAI está configurado para revolucionar el mercado, ofreciendo una reducción de diez veces en comparación con otros modelos GPT-3.5. A los desarrolladores se les cobrará en función de la cantidad de tokens utilizados, que corresponden a secuencias de mensajes con metadatos consumidos por el modelo. El precio es de $0,002 por cada 1000 tokens, equivalente a aproximadamente 750 palabras en la mayoría de los idiomas.

Con la integración de la API OpenAI del modelo gpt-3.5-turbo, la asequibilidad alcanza nuevas cotas. Por ejemplo, si necesitas 1000 ejecuciones de 100 palabras cada una, te costaría aproximadamente $0,30. Esto hace que la API de OpenAI sea una opción muy atractiva para los desarrolladores que buscan capacidades de lenguaje económicas, pero potentes.

Figura 1.6. La estructura de precios de la API de ChatGPT está diseñada para adaptarse a una amplia gama de necesidades de los desarrolladores, lo que garantiza que siga siendo accesible y rentable. A continuación se revisan algunos conceptos importantes sobre los precios de la API de ChatGPT.

Una de las ventajas más significativas del precio de la API de OpenAI es su estructura basada en tokens. Este enfoque permite a los desarrolladores pagar solo por los recursos que utilizan, lo que lo hace increíblemente rentable y flexible. Al alinear los costos con el uso real, es posible optimizar los gastos y asignar recursos de manera eficiente.

Si bien la API de OpenAI ofrece precios asequibles, es esencial realizar un seguimiento de su uso para evitar gastos inesperados. Al igual que con cualquier API, es crucial monitorear su consumo cuidadosamente.

Algunos puntos clave para considerar:

- **Realiza un seguimiento del uso de tokens**: al monitorear la cantidad de tokens que consume tu aplicación, puedes tener una comprensión clara de los gastos. Verifica regularmente el conteo de tokens para asegurarte de que se alinee con tus expectativas y presupuesto.
- **Establece límites de uso**: define límites de uso dentro de tu aplicación para evitar un consumo excesivo. De esta manera, puedes limitar la cantidad de tokens utilizados, manteniendo los costos bajo control.

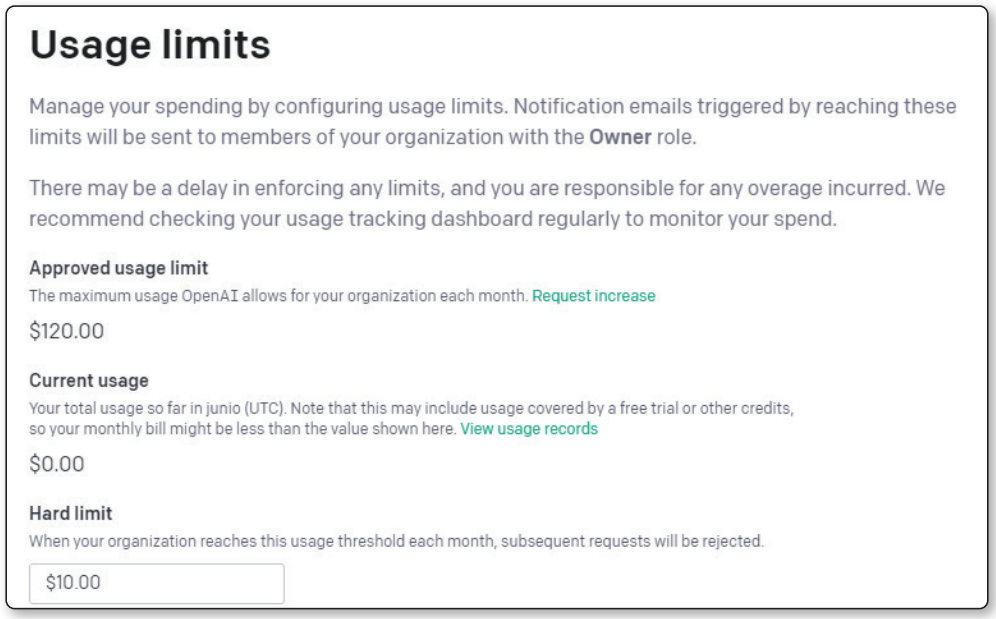


Figura 1.7. Límites de uso.

- **Implementa alertas de costos**: configura alertas o notificaciones para recibir actualizaciones sobre el consumo de la API. Este enfoque proactivo permite identificar rápidamente cualquier pico inesperado en el uso.
- **Optimiza las secuencias de mensajes**: la estructuración eficiente de los mensajes y metadatos puede ayudar a reducir el consumo de tokens. Agiliza tu flujo de comunicación y considera minimizar los mensajes innecesarios para optimizar los costos.
- **Revisa periódicamente la facturación**: controla de manera regular tus estados de cuenta para estar al tanto de los gastos. Esta práctica ayuda a identificar discrepancias o cargos inesperados, lo que garantiza que tengas una visión general clara del uso de la API.

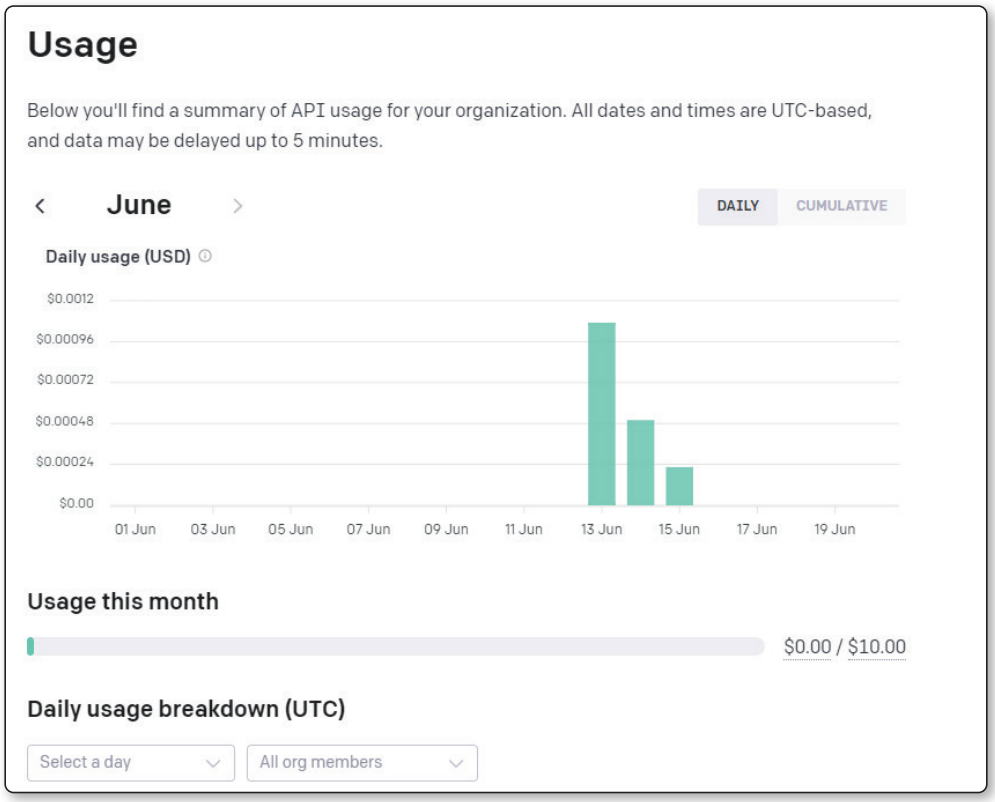


Figura 1.8. Detalle de consumos diarios.

1.3.1 ¿Puedo usar la API de OpenAI con la suscripción de ChatGPT Plus?

No, la suscripción a ChatGPT Plus cubre el uso exclusivamente en **chat.openai.com** y no incluye el acceso a la API de OpenAI.

1.3.1.1 ÚLTIMAS ACTUALIZACIONES DE LA API PARA DESARROLLADORES

A comienzos de junio de 2023, OpenAI lanzó novedades muy sorprendentes. Estas actualizaciones de la API incluyen una memoria de conversación cuatro veces mayor para GPT-3.5 y la funcionalidad de llamada a funciones.

Se anunció una actualización considerable de las ofertas de la API de modelo de lenguaje grande (incluidos GPT-4 y gpt-3.5-turbo), incluida una nueva capacidad de llamada de funciones, reducciones significativas de costos y una opción de **ventana de contexto** de 16.000 tokens para el modelo GPT-3.5-turbo.

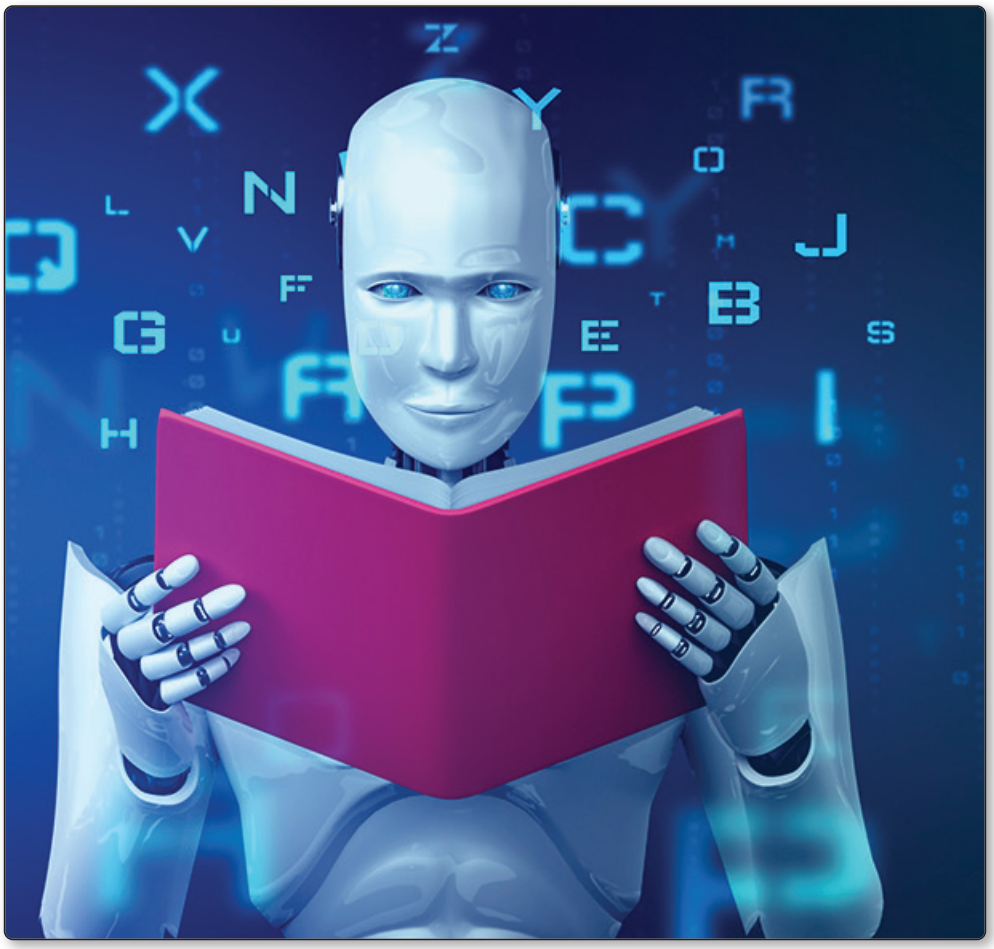


Figura 1.9. En los modelos de lenguaje grande (LLM), la "ventana de contexto" es como una memoria a corto plazo que almacena el contenido del prompt original o, en el caso de un chatbot, todo el contenido de la conversación en curso. En los modelos de lenguaje, aumentar el tamaño del contexto se ha convertido en una carrera tecnológica. OpenAI ha desarrollado una versión de 32.000 tokens de GPT-4, pero aún no está disponible públicamente.

Con cuatro veces la longitud de contexto de la versión estándar de 4000, gpt-3.5-turbo-16k puede procesar alrededor de veinte páginas de texto en una sola solicitud. Este es un impulso considerable para los desarrolladores que requieren que el modelo procese y genere respuestas para fragmentos de texto más grandes.

Además, se anunció un recorte de precio del 75% en el modelo de incrustaciones "ada", una reducción del 25% en el precio de los tokens para gpt-3.5-turbo.

La incorporación de las llamadas a funciones de OpenAI permite a los desarrolladores describir una función y el modelo genera una salida **JSON** que contiene los argumentos. Esta función no llama a ninguna función en sí misma, pero genera el JSON que se puede utilizar para llamar a una función desde el código.

Los desarrolladores definen las funciones como parte de la llamada de completado de chat. Luego, el modelo genera una salida JSON que se puede utilizar para llamar a la función específica desde el código.

Por ejemplo, puede convertir prompts como "Envía un correo electrónico a Eduardo para ver si puede reunirse el próximo viernes", en una llamada de función como "send_email(to: string, body: string)". En particular, esta función también permitirá una salida con formato JSON consistente, que los usuarios de la API tenían dificultades para generar anteriormente.

Cabe destacar que los costos del uso de la API están reduciéndose más con cada actualización, lo que hace pensar que esta tecnología seguirá siendo más que accesible para los desarrolladores a futuro, aunque (como todo en el mundo de la Inteligencia Artificial) la velocidad y lo cambiante de este fascinante mundo hace que resulte difícil predecir el rumbo comercial que pueda tomar en los próximos meses o años.

1.4 ACTIVIDADES

A continuación verás las preguntas que deberías saber responder para considerar aprendido el capítulo.

1.4.1 Test de autoevaluación

1. *Explica brevemente qué es una API y cuáles son sus principales usos en programación.*
2. *¿Qué es ChatML?*
3. *Indica qué diferencia hay entre usar ChatGPT Plus y la API de GPT de OpenAI.*
4. *Realiza un cuadro donde detalles los principales modelos de GPT que pueden usarse desde la API.*
5. *Define qué son las incrustaciones y cómo se relacionan con los conceptos de tokens y los prompts enviados a la API.*