# **Aplicaciones empresariales con Azure OpenAI**

*La inteligencia Artificial generativa está revolucionando la manera cómo trabajamos en nuestro día a día. No importa en qué área nos desempeñemos, los modelos de lenguaje grandes (LLMs) están mostrando capacidades para desarrollar tareas que en ocasiones es casi imposible diferenciar del trabajo hecho por humano. Sin embargo, muchas de las preocupaciones que tienen las empresas para implementar soluciones basadas en IA generativa están asociadas a la confiabilidad de los contenidos generados por los modelos de lenguaje y la privacidad y seguridad de sus datos cuando usan estos modelos.*

*En Azure OpenAI Service vamos a encontrar una plataforma con características empresariales que aseguran no solo el cumplimiento de diferentes estándares y normativas internacionales sobre el manejo de los datos en la nube, sino que también Microsoft provee un plataforma que garantiza que los datos, prompts y completions analizados por los modelos no van a ser usados ni por Microsoft ni por un tercero para entrenar otros modelos de IA generativa, así como también que los modelos que han pasado por un proceso fine-tuning tampoco estarán disponibles para Microsoft, OpenAI u otros clientes. Por otro lado, la confiabilidad de los contenidos es una preocupación constante en los usuarios y stakeholders de una aplicación basada en IA generativa, es por eso que Microsoft nos provee una serie de principios de IA generativa Responsable que nos ayudan a diseñar aplicaciones que controlen la generación de contenido perjudicial o dañino a través de diferentes herramientas.*

## **OpenAI Studio**

*Cuando desplegamos Azure OpenAI Service tendremos acceso a Azure OpenAI Studio, una plataforma diseñada para desplegar y probar diferentes versiones de modelos de IA generativa, desplegar rápidamente asistentes virtuales, manejar datasets que sean usados para procesos de fine-tuning, consultar y gestionar las cuotas asignadas a nuestros modelos y crear filtros personalizados de contenido para aplicar a los modelos, de tal manera que podamos controlar el tipo de contenido que podrían generar.*

*A screenshot of a computer

Description automatically generated*

### **Modelos**

*En la sección de modelos podemos ver todo el catálogo disponible para desplegar en nuestro recurso de Azure OpenAI service.*

*A screenshot of a computer

Description automatically generated*

*Azure OpenAI también nos permite aplicar fine-tuning a algunos modelos base para crear modelos personalizados que hayan sido ajustados y validados utilizando nuestros propios datos.*

*A screenshot of a custom model

Description automatically generated*

### **Deployments**

*En esta sección visualizaremos todos los modelos desplegados en nuestro recurso de Azure OpenAI y también tendremos la posibilidad de desplegar nuevos modelos base o modelos que han sido ajustados.*

*A screenshot of a computer

Description automatically generated*

### **Datos**

*Con el fin de mantener seguros nuestros datos y garantizar características de encriptación, en esta sección Azure OpenAI Service nos provee un mecanismo para almacenar los datasets usados durante el entrenamiento o validación de los modelos que pasan por procesos de fine-tuning.*

*A screenshot of a computer

Description automatically generated*

### **Filtros de Contenido**

*La IA Generativa Responsable nos recomienda una serie de principios que están enfocados en evitar que nuestras aplicaciones basadas en LLMs generen contenido que pueda llegar a ser perjudicial. Existen diferentes técnicas que nos permiten ayudar a controlar este tipo de situaciones, como por ejemplo el prompting o metaprompting. Sin embargo, en algunos escenarios es indispensable clasificar el contenido que se analiza (prompts) y el contenido generado (completions) por parte de los modelos con el fin de evitar que se genere contenido con alto nivel de severidad en estas categorías:*

* *Odio*
* *Sexual*
* *Autolesión*
* *Violencia*

*A screenshot of a computer

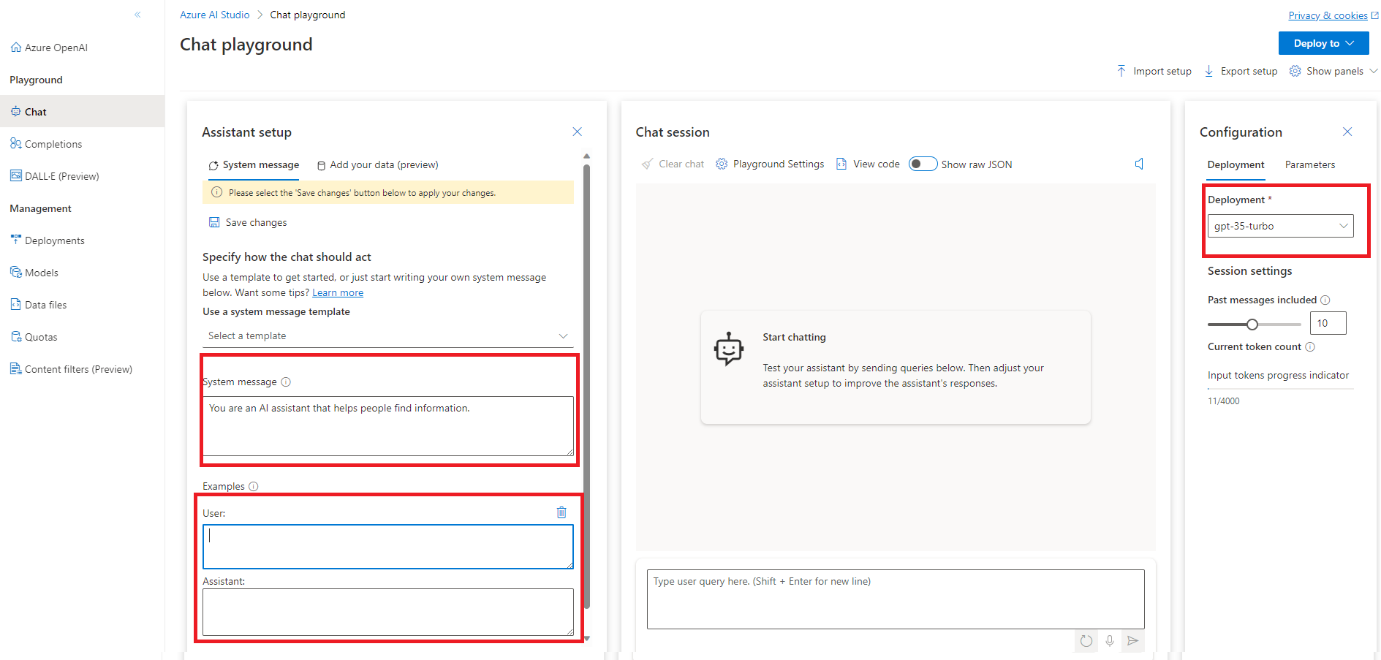
Description automatically generated*

## **Playground**

*Azure OpenAI service nos ofrece 3 platygrounds que nos permiten interactuar rápidamente con los modelos base o modelos ajustados que se despliegan.*

### **Chat Playground**

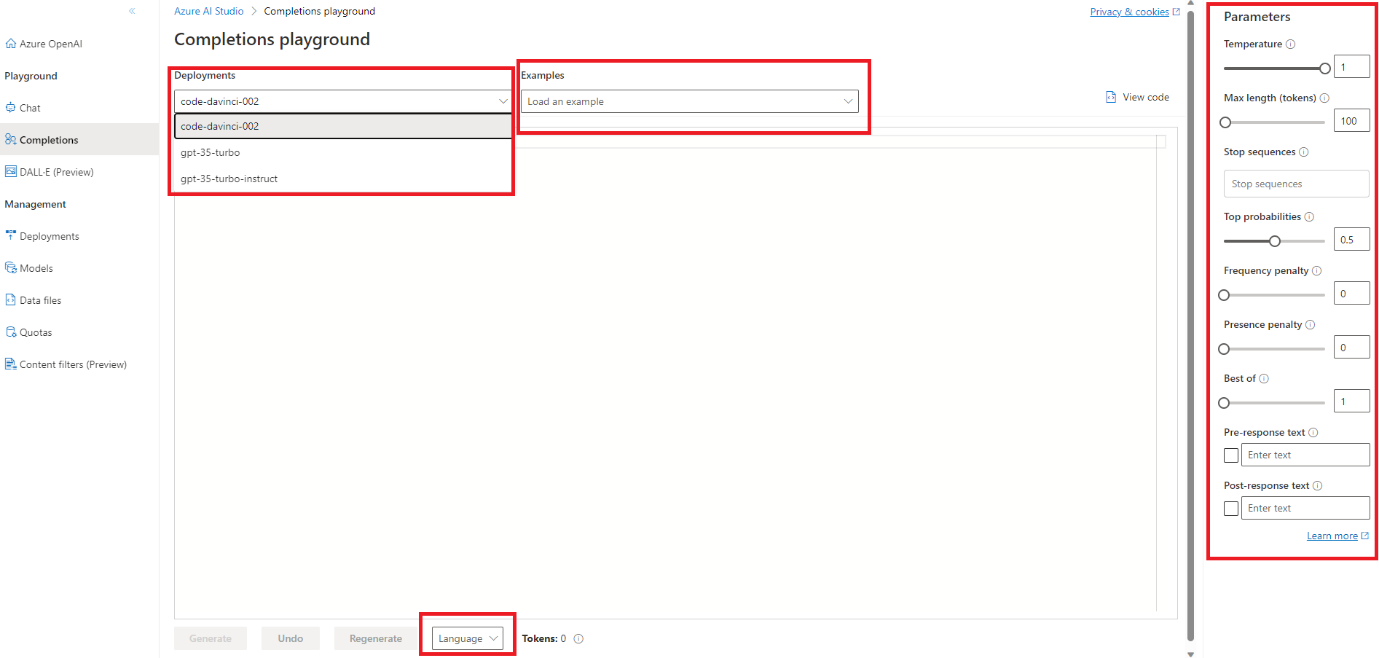
*Este playground nos permite interactuar con los modelos conversacionales (Chat Completions) que hemos desplegado.*

**

*Esta interfaz nos provee un mecanismo para hacer la configuración que permite establecer el comportamiento del asistente virtual, así como un mecanismo rápido para agregar ejemplos que sirvan de contexto para el modelo y una sección donde se puede escoger el modelo conversacional a usar.*

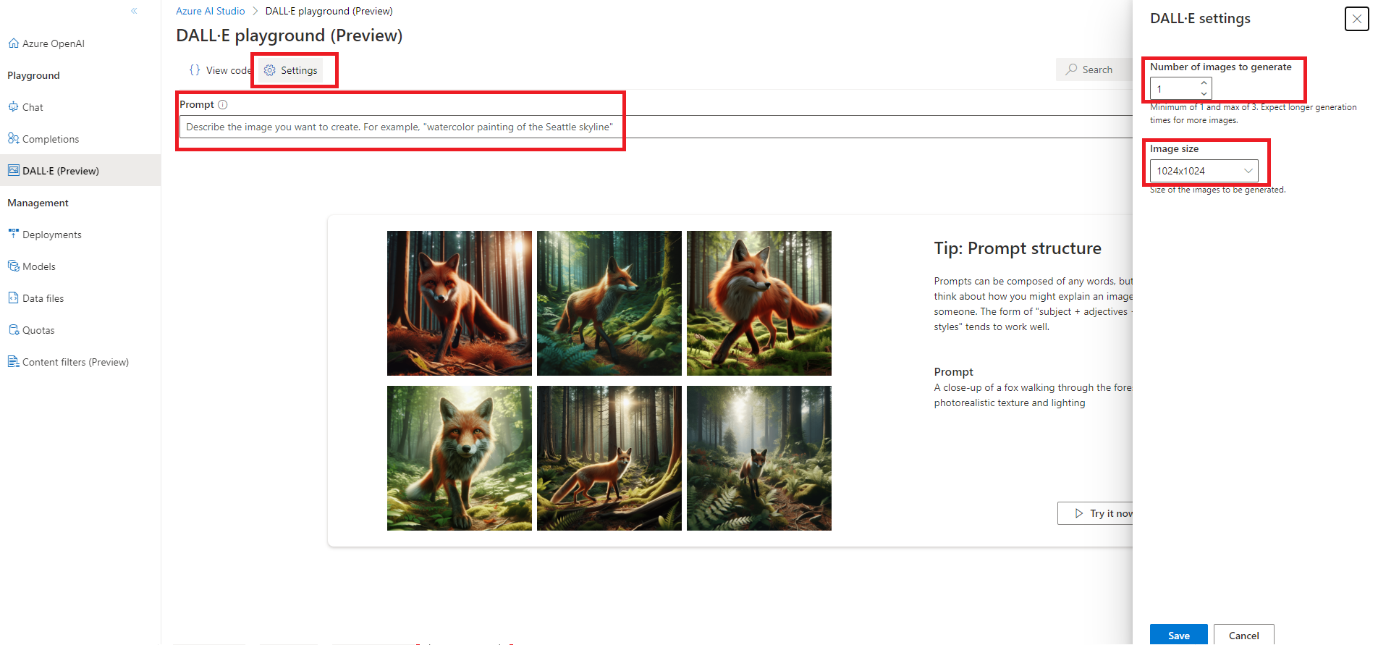
### **Completions Playground**

*Este playground nos permite interactuar con modelos que no tienen la API de chat completions, como por ejemplo Codex o solo la API de Completions. En esta interfaz vamos a poder especificar el modelo desplegado que queremos usar, el lenguaje que se debe considerar para generar los completions, realizar ajustes de los hiperparámetros e incluso revisar y probar ejemplos de prompts que podríamos usar para determinados escenarios.*

**

### **DALL-E Playground**

*El playground de DALL-E nos ofrece una interfaz para generar imágenes a partir de un prompt en lenguaje natural, así como también establecer la configuración que permite especificar los parámetros del modelo encargado de generar la imagen: número de imágenes a generar y tamaño de las imágenes.*

**

## **Chatea con tus datos**

*Uno de los escenarios en dónde los modelos de IA generativa pueden tener un alto impacto es aquellos en dónde los asistentes virtuales puedan ayudar a los usuarios a entender los datos, los cuales pueden venir desde fuentes estructuradas o no estructuradas. Por ejemplo, las organizaciones usualmente generan grandes volúmenes de datos que están en fuentes no estructuradas y que por lo general no son procesadas debido a la complejidad de sus formatos, sin embargo, este tipo de datos podrían generar valor cuando hablamos de escenarios en dónde se tienen documentos de tipo reportes o presentación de resultados. En este escenario los asistentes virtuales podrían ayudar a responder preguntas sobre la información contenida en los documentos.*

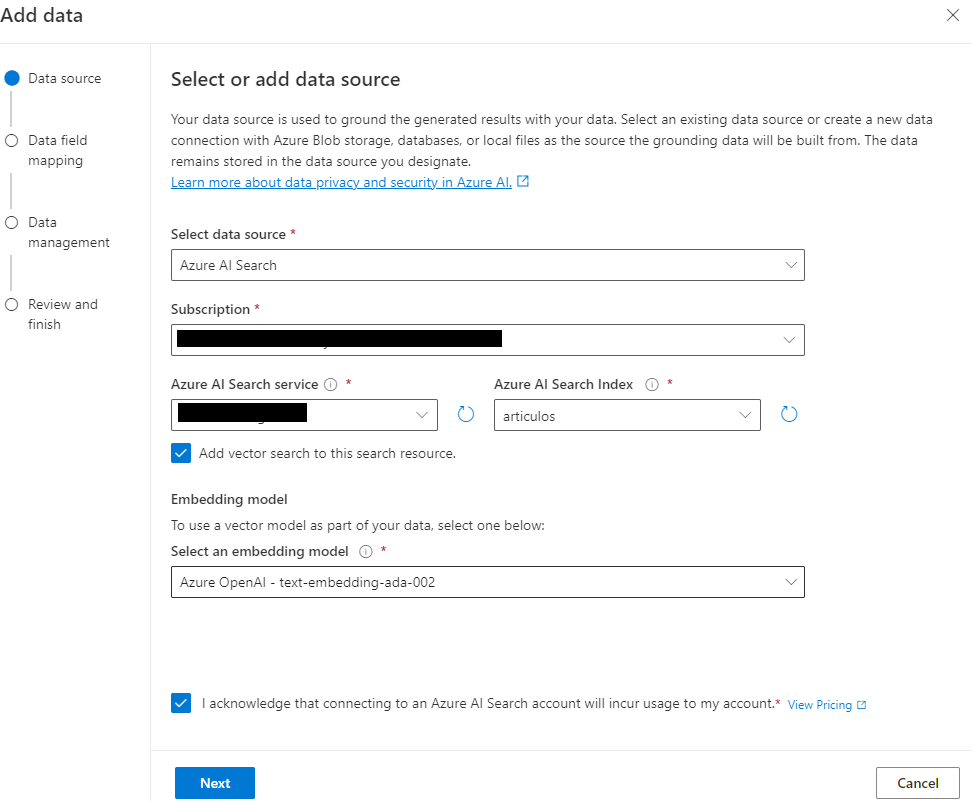
### **Agregar fuente de datos**

En el chat playground vamos a tener un mecanismo que nos permitirá agregar fuentes de datos no estructuradas e indexarlas a través de Azure AI Search.

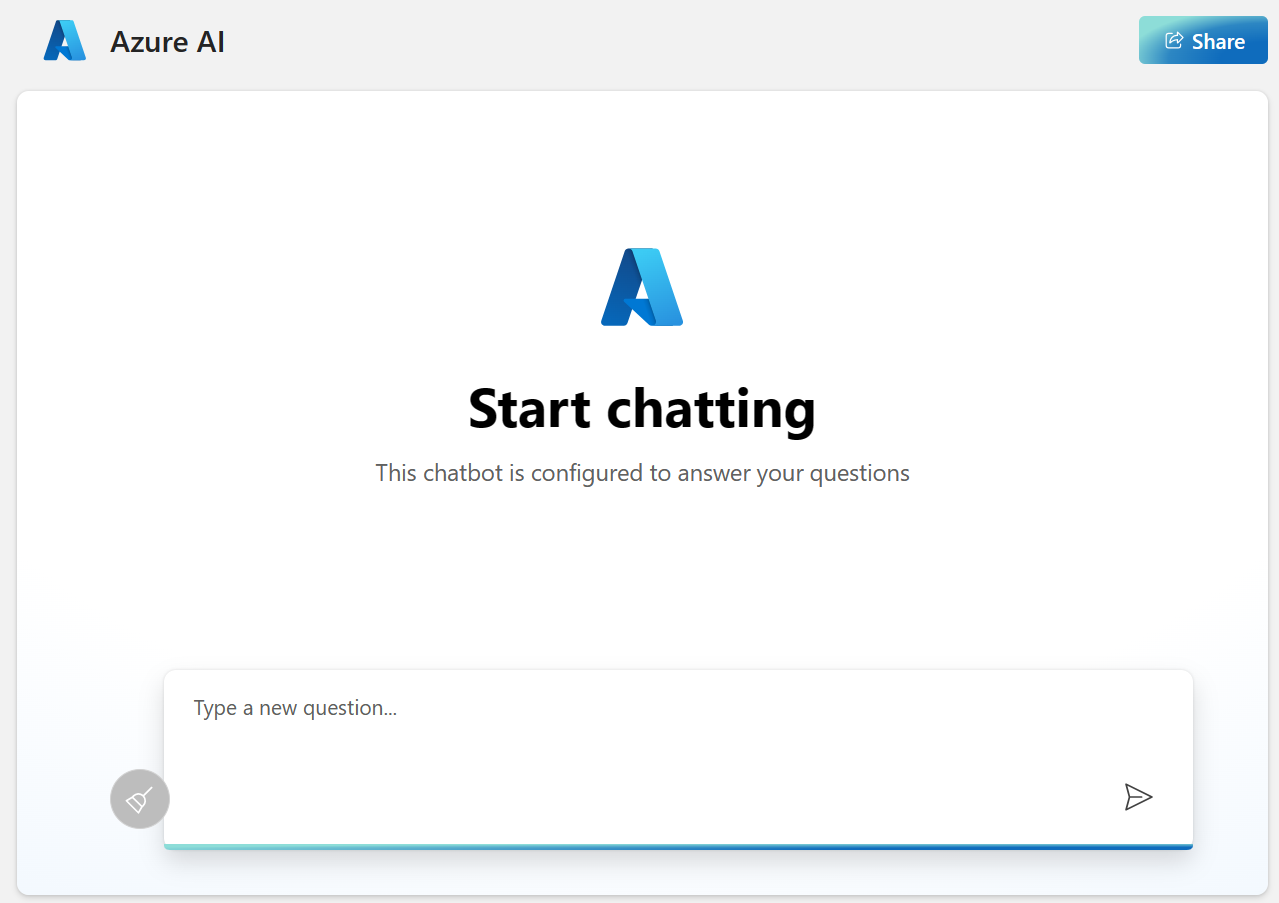
A screenshot of a computer

Description automatically generated

Con AI Search tendremos la capacidad de implementar la búsqueda híbrida, de tal manera que usemos un modelo de Azure OpenAI para generar embeddings a partir de los datos no estructurados y posteriormente soportar consultas que usen estos embeddings para implementar una búsqueda híbrida.



Una vez hemos integrado nuestros datos al asistente virtual, podemos desplegar en cuestión de minutos una aplicación web que soporta autenticación y ofrece una interfaz para hacer consultas que serán resueltas por los modelos de IA generativa utilizando nuestros datos.



*Ahora que ya sabes integrar tus datos para que interactúen con los modelos de IA generativa, desarrolla una aplicación en la que uses datos corporativos y aprovecha todas las características empresariales que te brinda el ecosistema Azure cuando usas Azure Open AI Service.*