



SA Azure OpenAI

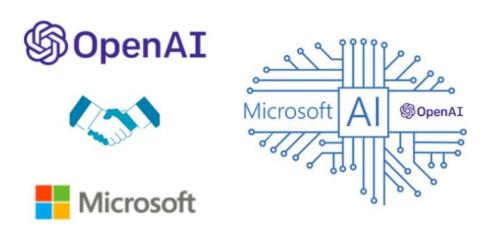
Cómo aprovechar el poder de la IA en tu proyecto

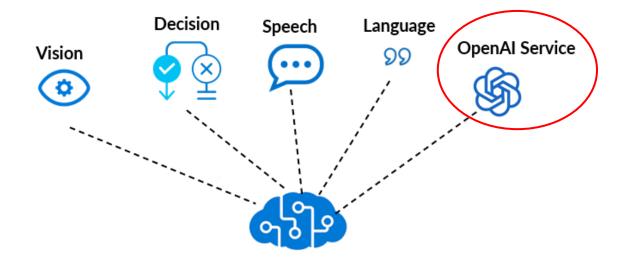
#Somoshiberus

#Lascosasocurrenaqui

#WeAreDifferent

1. INTRODUCTION





Azure Cognitive Service

Azure OpenAl es el resultado de la colaboración entre Microsoft y OpenAl

Es un servicio basado en la nube que provee acceso a modelos de lenguaje avanzados desarrollados por OpenAI.

Azure OpenAl vs OpenAl

Factors	Azure OpenAl	OpenAl
Privacidad de los datos		
No emplea tus datos para entrenar al modelo		
Residencia de los datos		
Puedes elegir entre diferentes regiones		8
Infrastructura de seguridad		
Redes virtuales Link privado Control del acceso		⊗ ⊗ ⊗

1. Acceso a Azure OpenAl

Request Access to Azure OpenAl Service

* Required

Please read all instructions carefully and complete form as instructed

Thank you for your interest in Azure OpenAl Service. Please submit this form to register for approval to access and use Azure OpenAl's Limited Access text and code and/or DALL·E 2 text to image models (as indicated in the form). All use cases must be registered. Azure OpenAl Service requires registration and is currently only available to approved enterprise customers and partners. Learn more about limited access to Azure OpenAl Service here.

Home > Azure Al services

Azure Al services | Azure OpenAl 🖈 …

+ Create / Manage deleted resources 🍪 Manage view 🗸 🖰 Refresh Filter for any field... Subscription equals Microsoft Partner Network - D Showing 1 to 6 of 6 records. Name ↑↓ Kind ↑↓ C-openai-demo OpenAl OpenAl FNMTAzureOpenAl OpenAl noured OpenAl openai-elecciones OpenAl westdemoopenai OpenAl

Create Azure OpenAl 2 Network 3 Tags 4 Review + submit Basics Enable new business solutions with OpenAl's language generation capabilities powered by GPT-3 models. These models have been pretrained with trillions of words and can easily adapt to your scenario with a few short examples provided at inference. Apply them to numerous scenarios, from summarization to content and code generation. Project details Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources. OpenAl Test Subscription Subscription * ① Resource group * ① test-resource-group Create new Instance details South Central US Region * (i) Name * ① azure-openai-test-001 Pricing tier * ① Standard S0 View full pricing details Content review policy

To detect and mitigate harmful use of the Azure OpenAl Service, Microsoft logs the content you send to the Completions and image generations APIs as well as the content it sends back. If content is flagged by the service's filters, it may be reviewed by a Microsoft full-time employee.

Configuración de la seguridad de red

Create Azure OpenAl ...









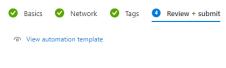
Review + subm



Type *

- All networks, including the internet, can access this resource.
- Selected networks, configure network security for your Azure Al services resource.
- Disabled, no networks can access this resource. You could configure private endpoint connections that will be the exclusive way to access this resource.

- •Opción 1:Todas las redes, incluida Internet, pueden acceder a este recurso.
- •Opción 2: Redes seleccionadas, configurar la seguridad de la red para su recurso de servicios Azure Al.
- •Opción 3: Deshabilitado, ninguna red puede acceder a este recurso. Puede configurar puntos de conexión privados que serán la forma exclusiva de acceder a este recurso.



TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the Azure Marketplace Terms for additional details.

Basics

Subscription Microsoft Partner Network - DIFERENCIALES

Resource group OpenAI_demo_M

 Region
 East US

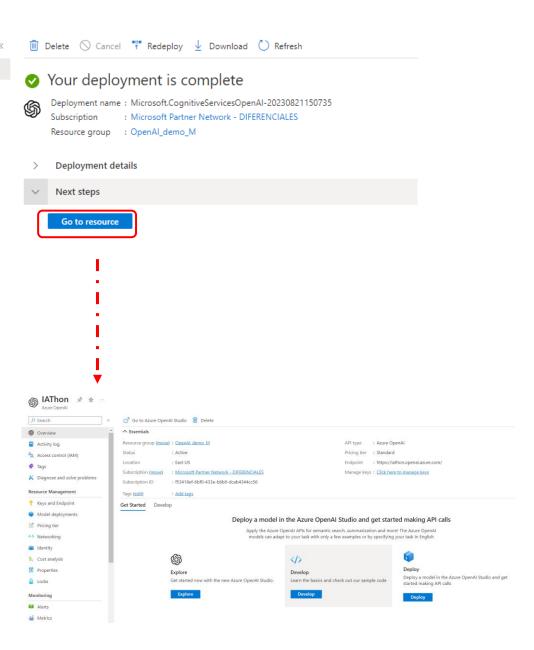
 Name
 IAThon

 Pricing tier
 Standard SO

Network

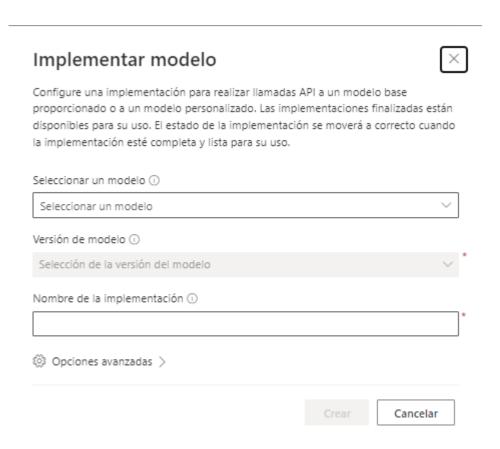
Type All networks, including the internet, can access this resource.





Implementar un modelo

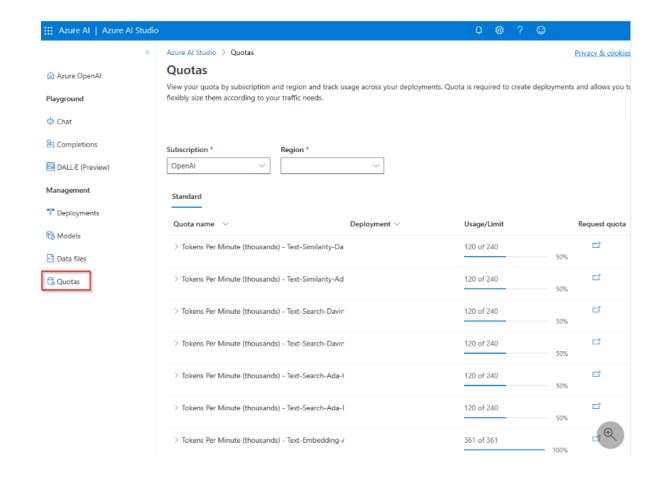
- 1.Inicie sesión en Azure OpenAl Studio.
- 2.Seleccione la suscripción y el recurso de Azure OpenAl con el que quiere trabajar y seleccione Usar recurso.
- 3.En Administración, seleccione Implementaciones.
- 4. Seleccione Crear nueva implementación y configure los siguientes campos:



Implementar un modelo

Deploy model X Set up a deployment to make API calls against a provided base model or a custom model. Finished deployments are available for use. Your deployment status will move to succeeded when the deployment is complete and ready for use. Select a model (1) text-similarity-ada-001 Deployment name (i) text-similarity-ada-001 Advanced options ∨ Content Filter (1) Default i) 120K tokens per minute quota available for your deployment Tokens per Minute Rate Limit (thousands) (1) 120K Corresponding requests per minute (RPM) = 720

Al crear una implementación de modelo, tiene la opción de asignar Tokens por minuto (TPM) a esa implementación



Los modelos de OpenAI tienen dos aplicaciones principales: chat y completado.

En escenarios de **Chat**, los usuarios participan en conversaciones naturales e interactivas con el modelo, ya sea usando ChatGPT o modelo GPT-4. Esta interacción puede incluir la realización de **preguntas**, la búsqueda de **consejos o simplemente disfrutar de una charla informal**.

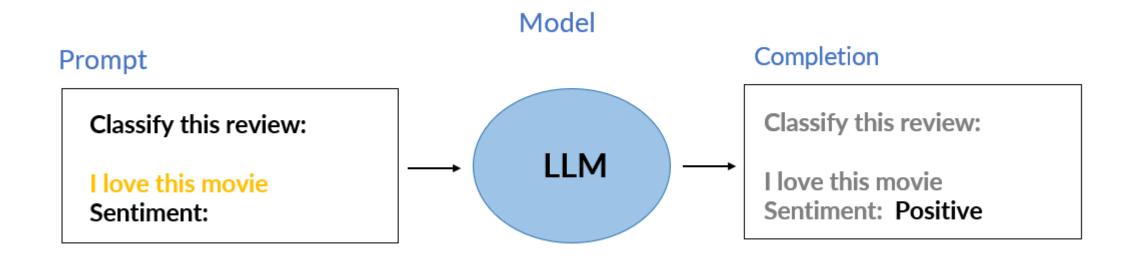
Los escenarios de **completado** implican desafiar al modelo a crear algo extraordinario y único, como crear poemas, historias, código o canciones.

En Azure OpenAI los usuarios pueden explorar y experimentar con los modelos de OpenAI en varias interfaces, incluyendo completions y chats.

Conceptos clave

Entradas (prompts) & salidas (completions)

El completion endpoint es el componente principal del servicio API. Este servicio API provee acceso a la interfaz de entrada y salida del modelo. Los usuarios solo necesitan proporcionar un mensaje de entrada conteniendo el texto del comando. El modelo generará una respuesta.



Conceptos clave

```
openai.api_type = "azure"
openai.api_base = "<u>https:/</u>
openai.api_version = "2023
openai.api_key = "29bc1acf
```

El presidente de la Generalitat de la Cataluña, Pere **Aragonès, acudirá el próximo jueves a la Comisión General de las Comunidades Autónomas del Senado** para defender "los grandes consensos" de la sociedad catalana en torno a la <u>amnistía</u> y la autodeterminación, frente a la "derecha reaccionaria".

Así lo ha anunciado este martes en rueda de prensa la portavoz del Govern, Patrícia Plaja, que ha argumentado que el Ejecutivo catalán "no dejará nunca ningún espacio vacío" para reivindicar "la resolución del conflicto y el consenso de la sociedad catalana en torno a una cuestión fundamental como el fin de la represión".

La Comisión General de las Comunidades Autónomas del Senado se reunirá el día 19 a propuesta del PP "para debatir sobre la aplicación efectiva de los principios de igualdad ante la ley de todos los españoles" y de "solidaridad" entre los distintos territorios. En principio está previsto que asistan todos los presidentes autonómicos del PP, pero no los del PSOE ni el lehendakari vasco, Íñigo Urkullu.

Kataluniako Generalitateko presidenteak, Pere Aragonès, etorriko da datorren ostegunean Senatuko Autonomia Komunitate Nagusiko Batzorde Orokorrera, kataluniar gizarteak "amsnestia eta autodeterminazioaren kontuan diren adostasun handiak" defendatzeko, "erreakzionario eskortik" aurre eginez.

Horrela adierazi du asteartean argitaratutako prentsaurrekoan Gobernuaren portavoz Patrícia Plaja, Kataluniako Gobernuak "inoiz ere ez du hutsuneik utziko" gatazkaren ebazpena eta kataluniar gizartearen adostasuna aldarrikatzeko "konponbiderik gabe" aldarrikatzeko.

Senatuko Autonomia Komunitate Nagusiko Batzorde Orokorrak PPren proposamenez bilduko da ekainaren 19an "espainiar guztion artean legearen aurrean berdintasunaren eta lurralde ezberdinen arteko elkartasunaren printzipioak efektiboki aplikatzeari buruz eztabidatzeko". Une honetan, PPko autonomiengo presidente guztiak eta ez PSOEkoak dutenak edo euskal lehendakaria, Íñigo Urkullu.

Conceptos clave

Tokens

Azure OpenAl procesa el texto dividiéndolo en tokens. Los tokens puede ser palabra o, simplemente, fragmentos de caracteres.

Por ejemplo, la palabra "hamburger" se divide en los tokens "ham", "bur" and "ger", mientras que una palabra corta y común como "pear" es un único token. Muchos tokens comienzan con un espacio en blanco, por ejemplo, "hello" y "bye".

El número total de tokens procesados data una solicitud dependen de la longitud de la entrada, salida y parámetros de solicitud. La cantidad de token procesos también afecta al tiempo de respuesta y rendimiento de los modelos.

Conceptos clave

Tokens

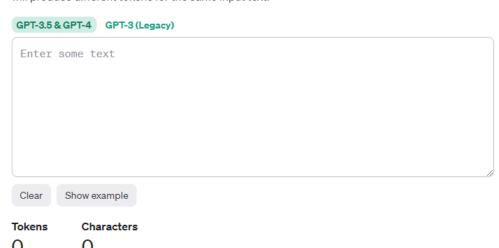
Tokenizer

Learn about language model tokenization

OpenAl's large language models (sometimes referred to as GPT's) process text using **tokens**, which are common sequences of characters found in a set of text. The models learn to understand the statistical relationships between these tokens, and excel at producing the next token in a sequence of tokens.

You can use the tool below to understand how a piece of text might be tokenized by a language model, and the total count of tokens in that piece of text.

It's important to note that the exact tokenization process varies between models. Newer models like GPT-3.5 and GPT-4 use a different tokenizer than our legacy GPT-3 and Codex models, and will produce different tokens for the same input text.



Tokens Characters 280 1024

El presidente de la Generalitat de la Cataluña, Pere Aragonès, acudirá el próximo jueves a la Comisión General de las Comunidades Autónomas del Senado para defender "los grandes consensos" de la sociedad catalana en torno a la amnistía y la autodeterminación, frente a la "derecha re accionaria".

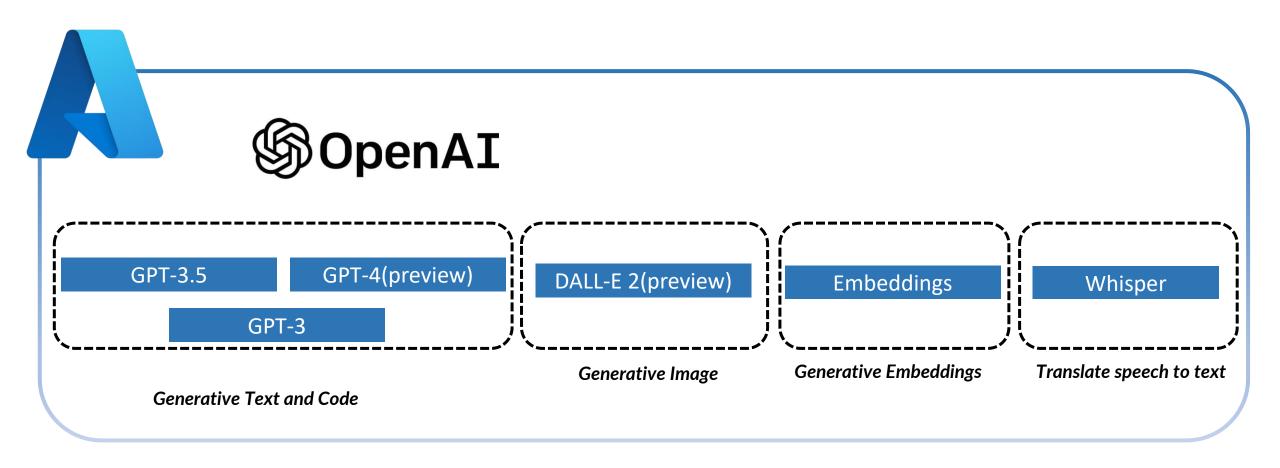
Así lo ha anunciado este martes en rueda de prensa la portavoz del Govern. Patrícia Plaja, que ha argumentado que el Ejecutivo catalán "no dejará

ASI 10 ha anunciado este martes en rueda de prensa la portavoz del Govern
, Patrícia Plaja, que ha argumentado que el Ejecutivo catalán "no dejará
nunca ningún espacio vacío" para reivindicar "la resolución del conflic
to y el consenso de la sociedad catalana en torno a una cuestión
fundamental como el fin de la represión".

La Comisión General de las Comunidades Autónomas del Senado se reunirá el día 19 a propuesta del PP "para debatir sobre la aplicación efectiva de los principios de igualdad ante la ley de todos los españoles" y de "solidaridad" entre los distintos territorios. En principio está previsto que asistan todos los presidentes autonómicos del PP, pero no los del PS OE ni el lehendakari vasco, ��ñigo Urkullu.

TEXT TOKEN IDS

OpenAl Models on Azure



OpenAl Models en Azure

GPT-4

GPT-4 puede resolver difíciles con mayor precisión que cualquier modelo previo de OpenAI. Como GPT-3.5 Turbo, GPT-4 está optimizado para chatear y funciona bien en tareas de completado tradicionales. Para usar GPT-4 emplea la API de Chat Completions.

Se ofrecen dos versiones del modelo GPT-4:

- 1. GPT-4: admite un máximo de 8,192 tokens de entrada.
- 2. GPT-4-32k: admite un rango ampliado de hasta 32,768 tokens.

Para obtener acceso a GPT-4, los clientes de Azure OpenAl puede solicitar el acceso completando el siguiente formulario:

https://customervoice.microsoft.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHbR7en2Ais5pxKtso Pz4b1 xURjE4QlhVUERGQ1NXOTINT0w1NldTWjJCMSQlQCN0PWcu

OpenAl Models en Azure

GPT-3

Los modelos GPT-3 de OpenAI son considerados unos de los modelos de lenguaje más avanzados del mundo ya que son capaces de producir textos que, a menudo, se asemejan al estilo de escritura humana.

LATEST MODEL	DESCRIPTION	MAX TOKENS	TRAINING DATA
text-curie-001	Very capable, faster and lower cost than Davinci.	2,049 tokens	Up to Oct 2019
text-babbage-001	Capable of straightforward tasks, very fast, and lower cost.	2,049 tokens	Up to Oct 2019
text-ada-001	Capable of very simple tasks, usually the fastest model in the GPT-3 series, and lowest cost.	2,049 tokens	Up to Oct 2019
davinci	Most capable GPT-3 model. Can do any task the other models can do, often with higher quality.	2,049 tokens	Up to Oct 2019
curie	Very capable, but faster and lower cost than Davinci.	2,049 tokens	Up to Oct 2019
babbage	Capable of straightforward tasks, very fast, and lower cost.	2,049 tokens	Up to Oct 2019
ada	Capable of very simple tasks, usually the fastest model in the GPT-3 series, and lowest cost.	2,049 tokens	Up to Oct 2019

OpenAl Models en Azure

GPT-3.5

Los modelos GPT-3.5 pueden entender y generar lenguaje natural y código. El modelo más capaz y rentable de la familia GPT-3.5 es GPT-3.5 Turbo, conocido como CHATGPT, que ha sido optimizado para chatear y también funciona bien en tareas de completado tradicionales.

Los modelos GPT-3.5 Turbo incluyen:

GPT-3.5 Turbo

GPT-3.5 Turbo 16k

El modelo GPT-3.5 Turbo admite un máximo de 4,096 tokens de entrada, mientras que el modelo GPT-3.5 Turbo 16k extiende la admisión a hasta 16,384 tokens.

Para utilizar el modelo GPT-3.5 Turbo, similar a GPT-4, simplemente ejecuta la API Chat Completions.

OpenAl Models en Azure

GPT-4

With broad general knowledge and domain expertise, GPT-4 can follow complex instructions in natural language and solve difficult problems with accuracy.

Learn about GPT-4

Model	Input	Output
8K context	\$0.03 / 1K tokens	\$0.06 / 1K tokens
32K context	\$0.06 / 1K tokens	\$0.12 / 1K tokens

GPT-3.5 Turbo

GPT-3.5 Turbo models are capable and cost-effective.

gpt-3.5-turbo is the flagship model of this family and is optimized for dialog.

gpt-3.5-turbo-instruct is an Instruct model and only supports a 4K context window.

Learn about GPT-3.5 Turbo ↗

Model	Input	Output
4K context	\$0.0015 / 1K tokens	\$0.002 / 1K tokens
16K context	\$0.003 / 1K tokens	\$0.004 / 1K tokens

OpenAl Models en Azure

Fine-tuning models

Create your own custom models by fine-tuning our base models with your training data. Once you fine-tune a model, you'll be billed only for the tokens you use in requests to that model.

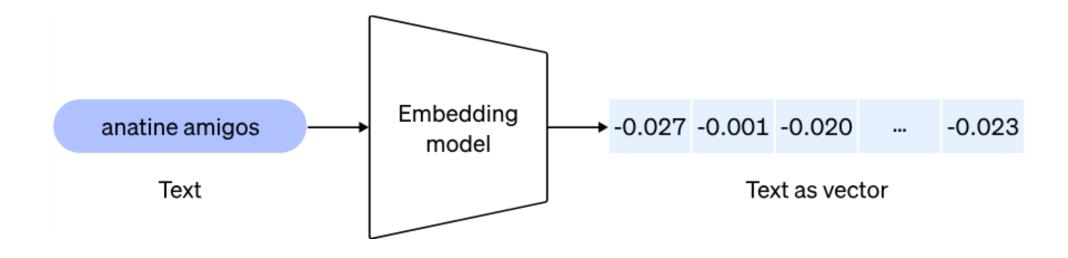
Learn about fine-tuning ↗

Model	Training	Input usage	Output usage
babbage-002	\$0.0004 / 1K tokens	\$0.0016 / 1K tokens	\$0.0016 / 1K tokens
davinci-002	\$0.0060 / 1K tokens	\$0.0120 / 1K tokens	\$0.0120 / 1K tokens
GPT-3.5 Turbo	\$0.0080 / 1K tokens	\$0.0120 / 1K tokens	\$0.0160 / 1K tokens

OpenAl Models en Azure

Modelos de embedding

Los modelos de embeddings de OpenAI son un conjunto de modelos que pueden convertir texto en una forma numérica que captura el significado semántico y la relación del texto. Los modelos de embeddings de OpenAI son útiles para varias tareas, como búsqueda, agrupación, recomendación, detección de anomalías y clasificación.



OpenAl Models en Azure

Embedding models

Build advanced search, clustering, topic modeling, and classification functionality with our embeddings offering.

Learn about embeddings ↗

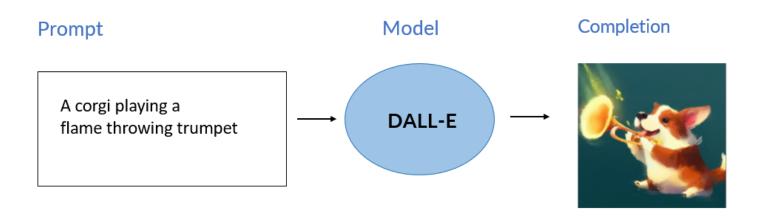
Model	Usage
Ada v2	\$0.0001 / 1K tokens

OpenAl Models en Azure

Modelo DALL-E

DALL-E es un modelo generativo de aprendizaje profundo que toma una entrada de texto y genera una imagen novedosa que ilustra el texto. Fue introducido en enero de 2021 por OpenAI.

DALL-E es un modelo de difusión que emplea el proceso de difusión para fusionar múltiplos conceptos visuales para crear imágenes novedosas que no se han visto antes.



OpenAl Models en Azure

Image models

Build DALL-E directly into your apps to generate and edit novel images and art. Our image models offer three tiers of resolution for flexibility.

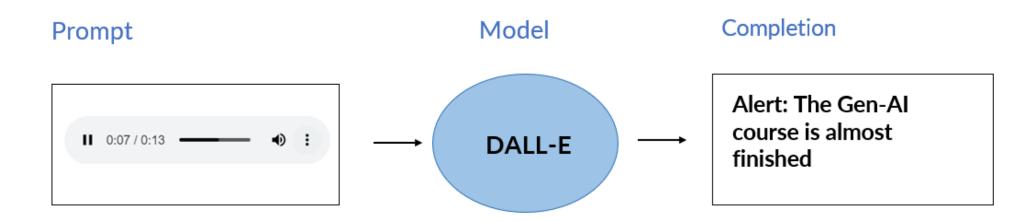
Learn about image generation ↗

Resolution	Price
1024×1024	\$0.020 / image
512×512	\$0.018 / image
256×256	\$0.016 / image

OpenAl Models en Azure

Modelo Whisper

Un modelo transformer secuencia a secuencia es entrenado en varias tareas de procesamiento del lenguaje, incluyendo el reconocimiento de múltiples idiomas, la traducción de textos, la identificación del lenguaje hablado y la detección de actividad de voz.



OpenAl Models en Azure

Audio models

Whisper can transcribe speech into text and translate many languages into English.

Learn about Whisper ↗

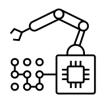
Model	Usage
Whisper	\$0.006 / minute (rounded to the nearest second)



Prompt Engineering



Fine-Tuning



- > Ada
- Babbage
- Curie

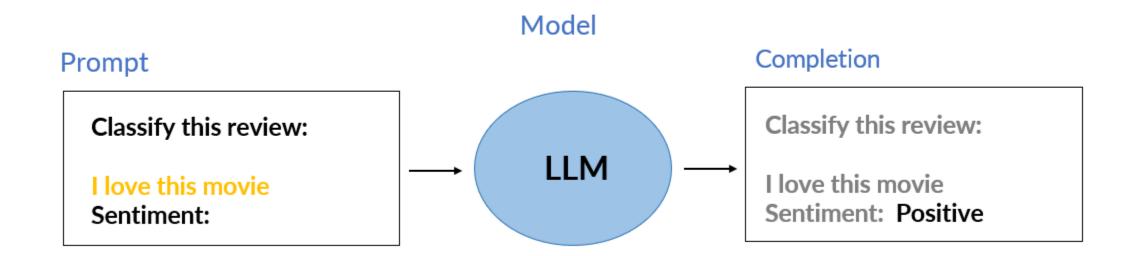
Azure OpenAl On Your Data



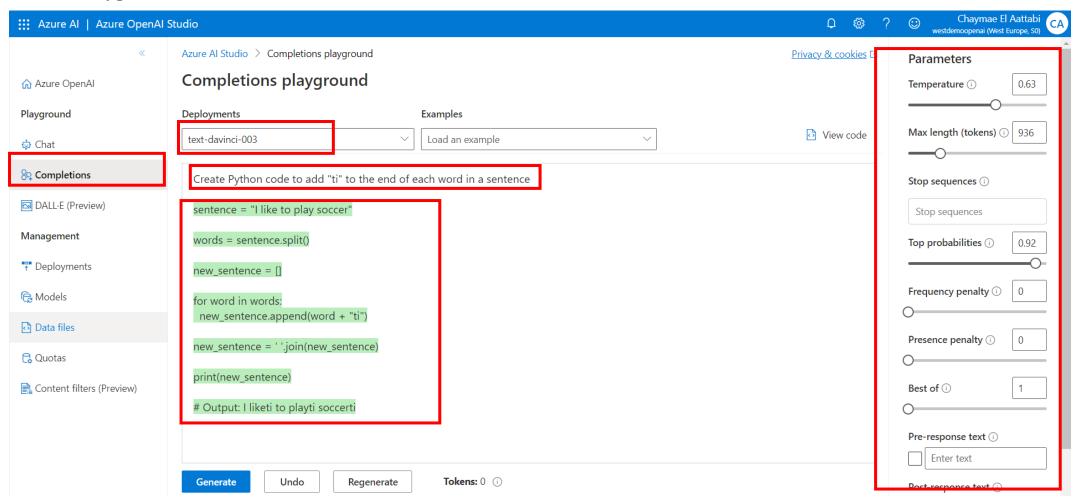
➤ GPT-3.5 Turbo

➤ GPT-4

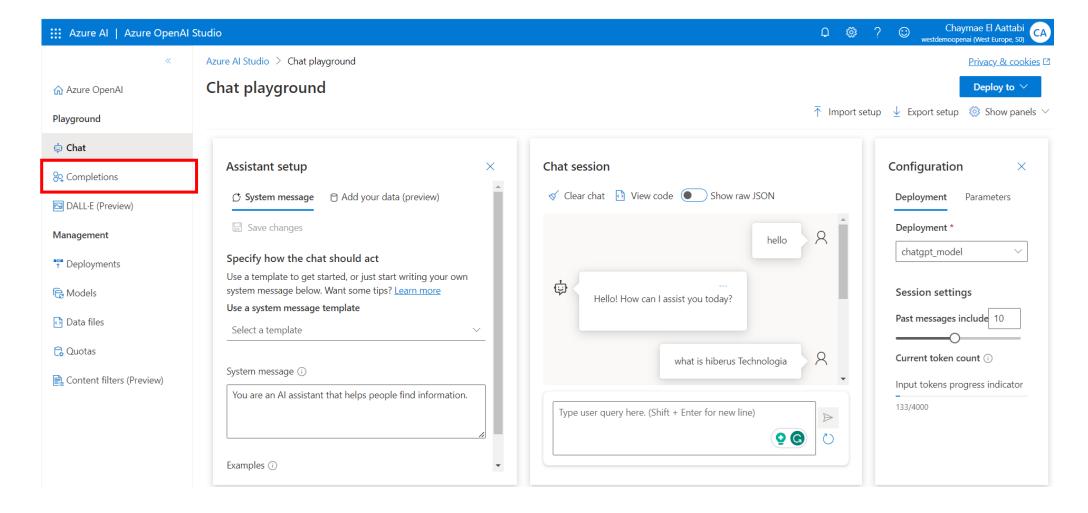
Prompt Engineering



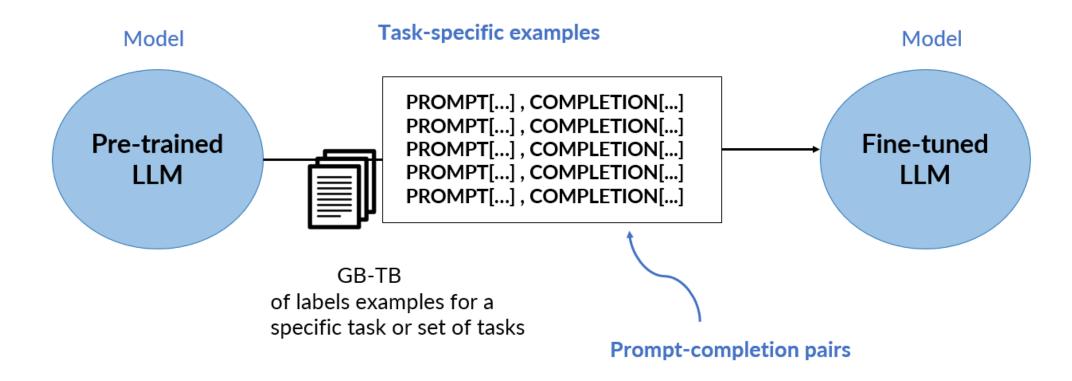
Completions Playground



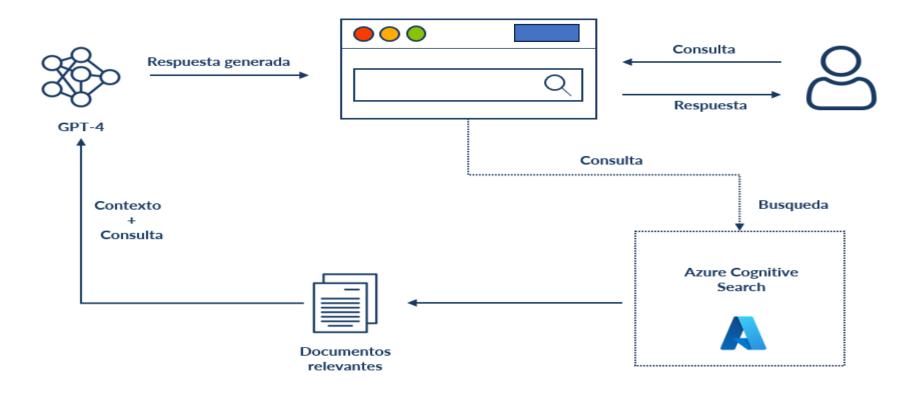
Chat Playground



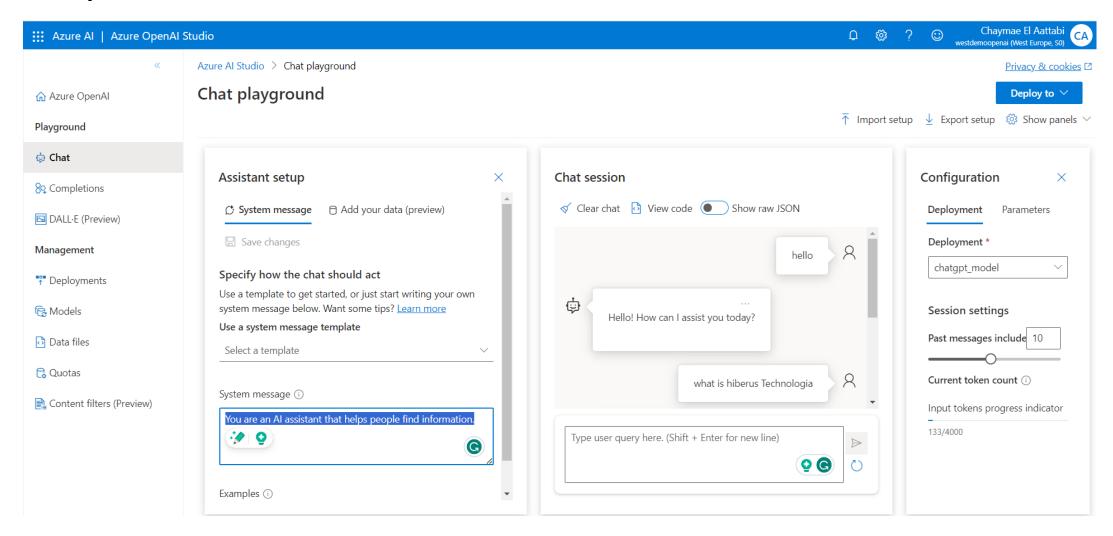
Fine-tuning

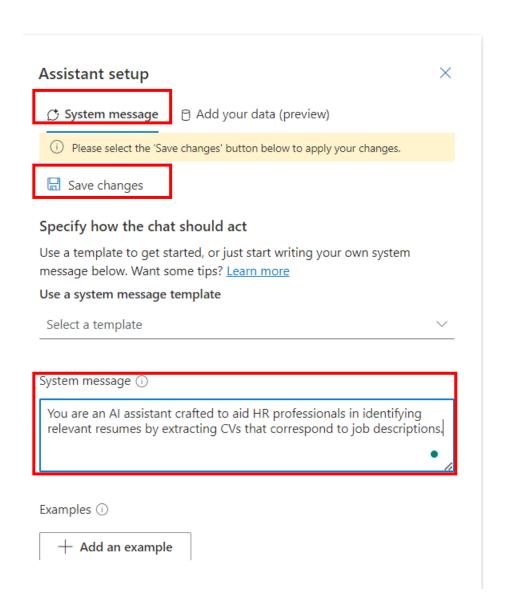


Azure OpenAl On Your data



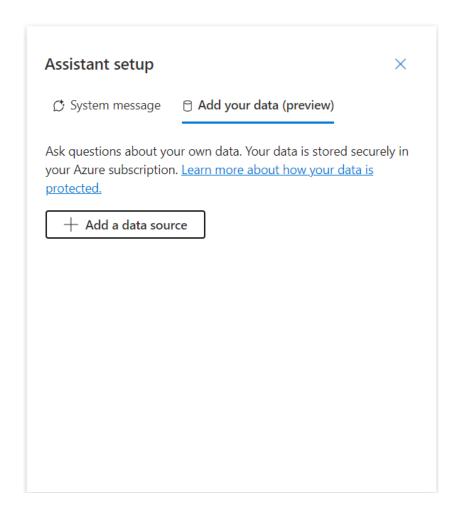
Azure OpenAl On Your data

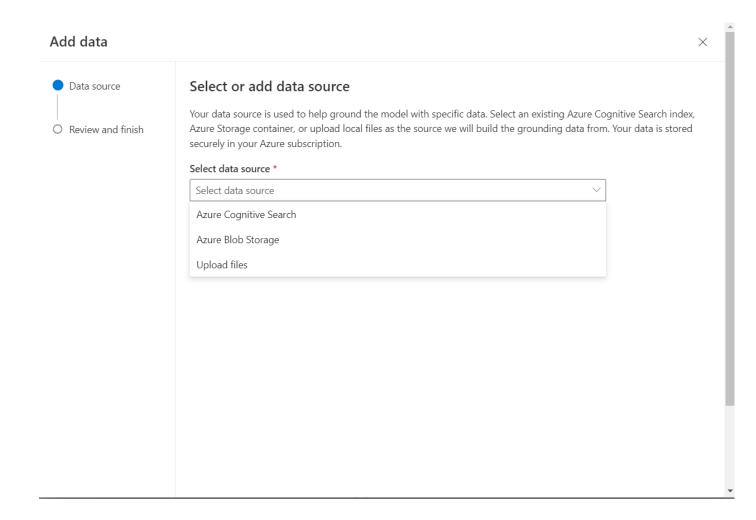


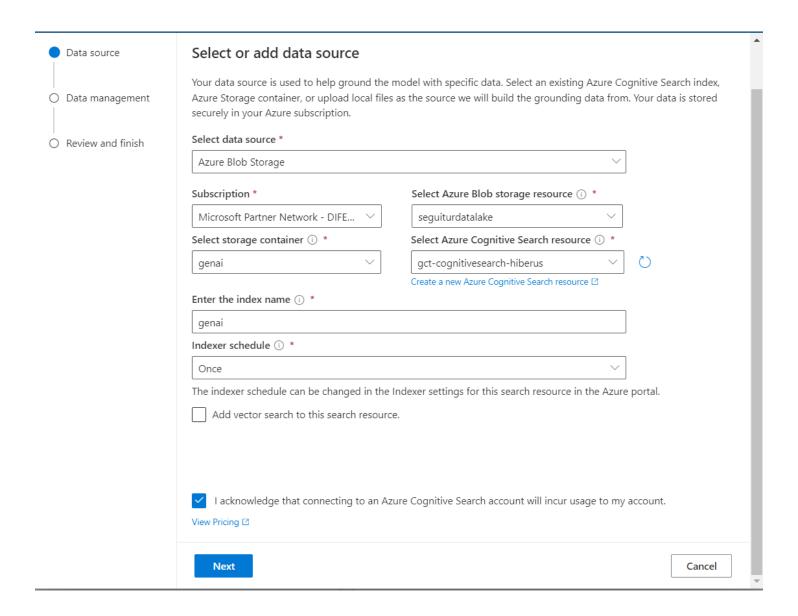


#somoshiberus

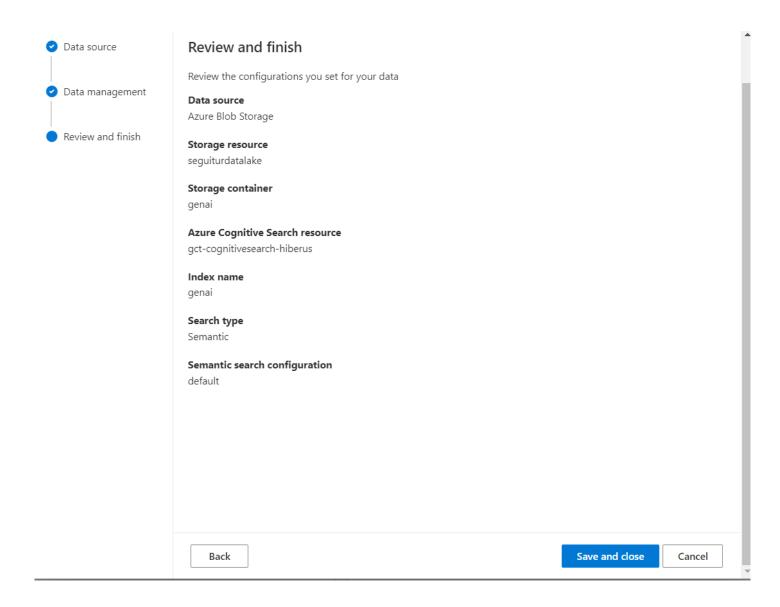
2. Características y casos de uso de Azure OpenAl

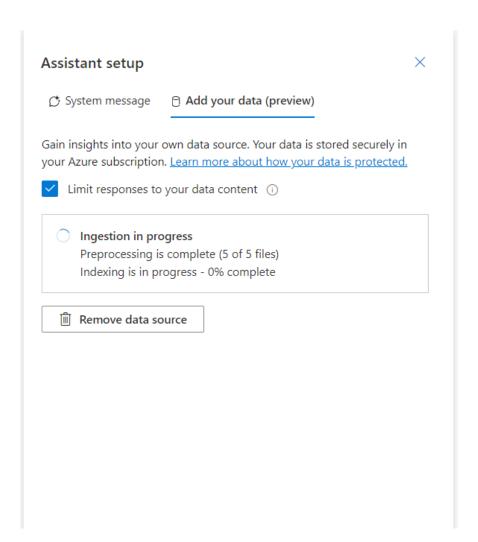


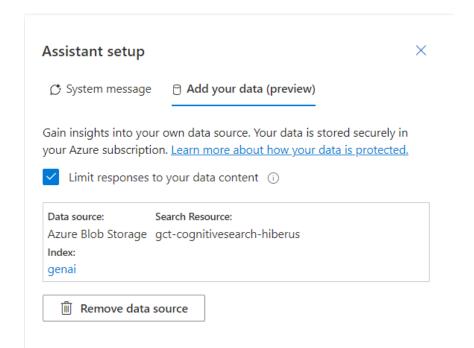


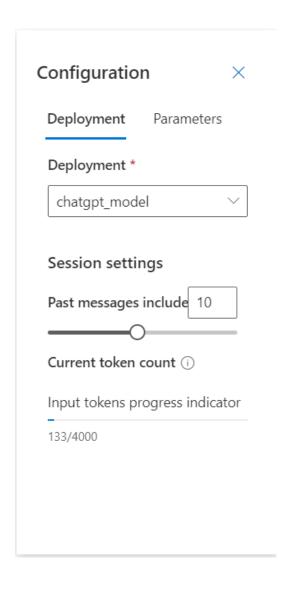


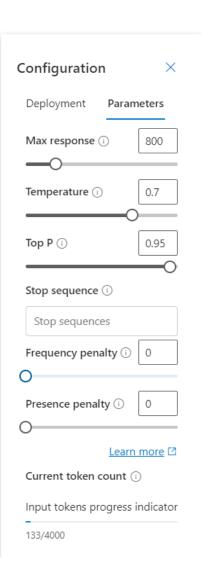
✓ Data source	Data management	
Data management	Set up specific configurations for your data and how the model will respond to requests.	ı
	Search type (i) *	ı
Review and finish	Search type	
	Keyword	ı
	Semantic	ı
		ı
		ı
		ı
		ı
		ı
		ı
		ı
		ı
		ı
	Back Next Cancel	



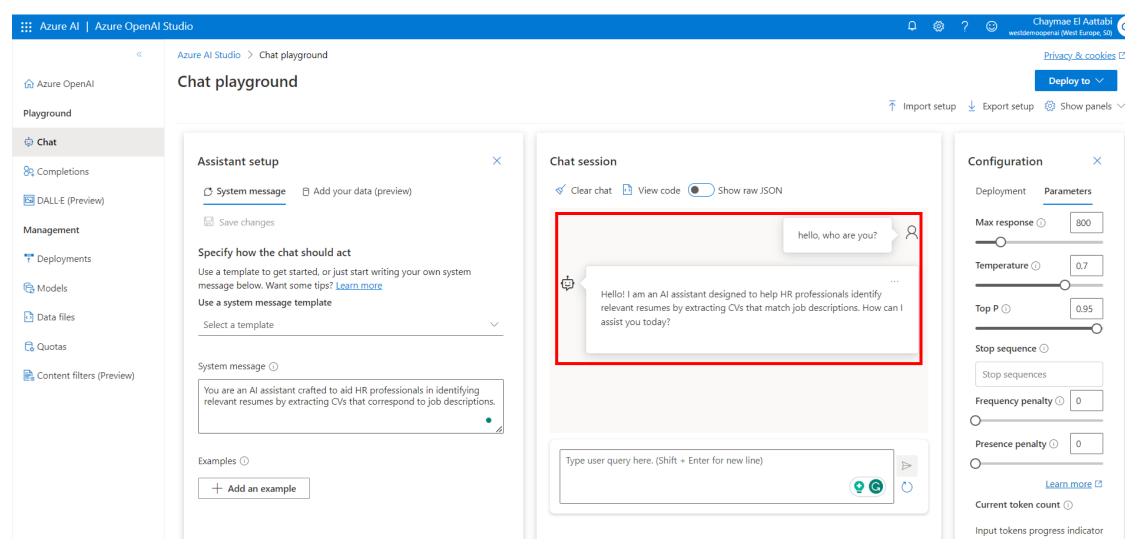




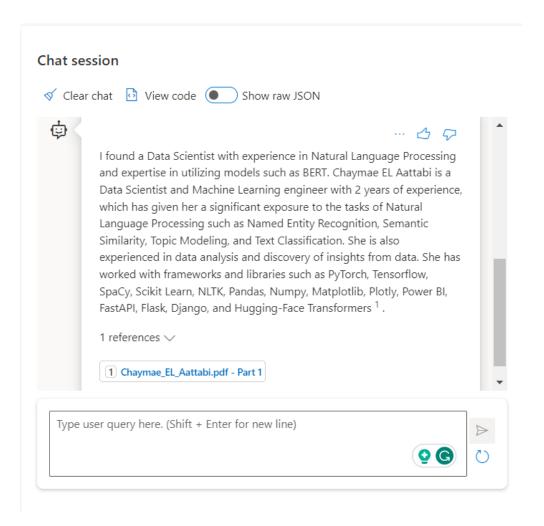


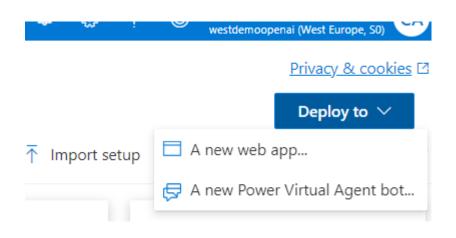


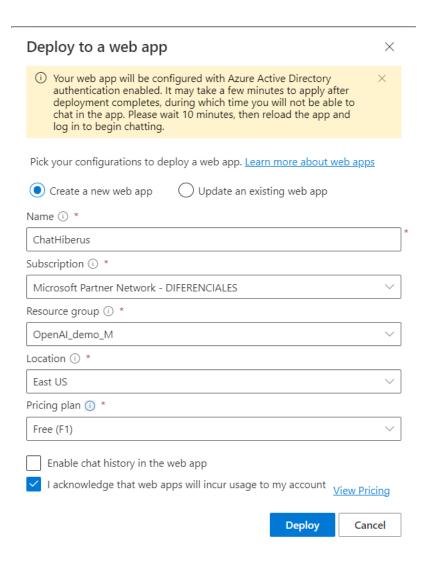
#somoshiberus

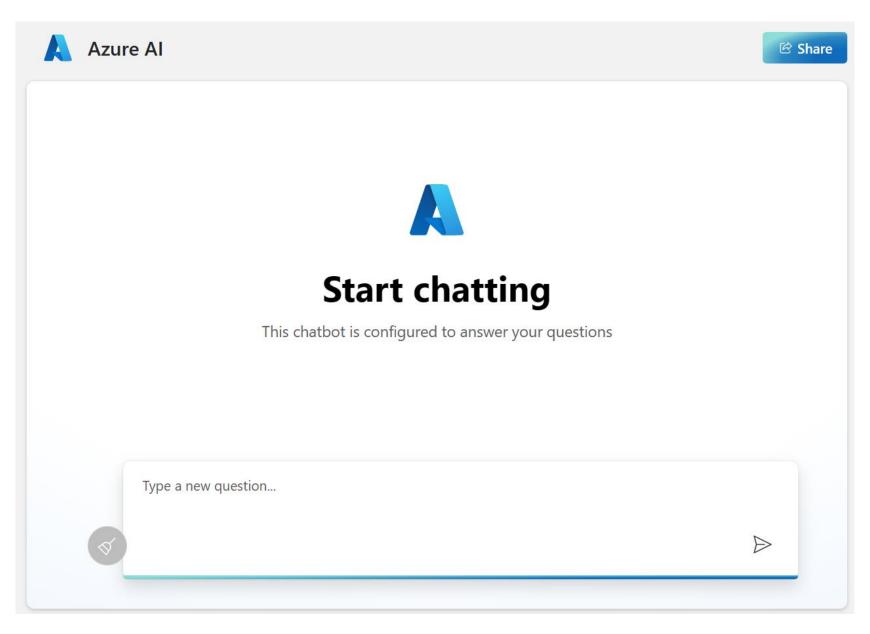


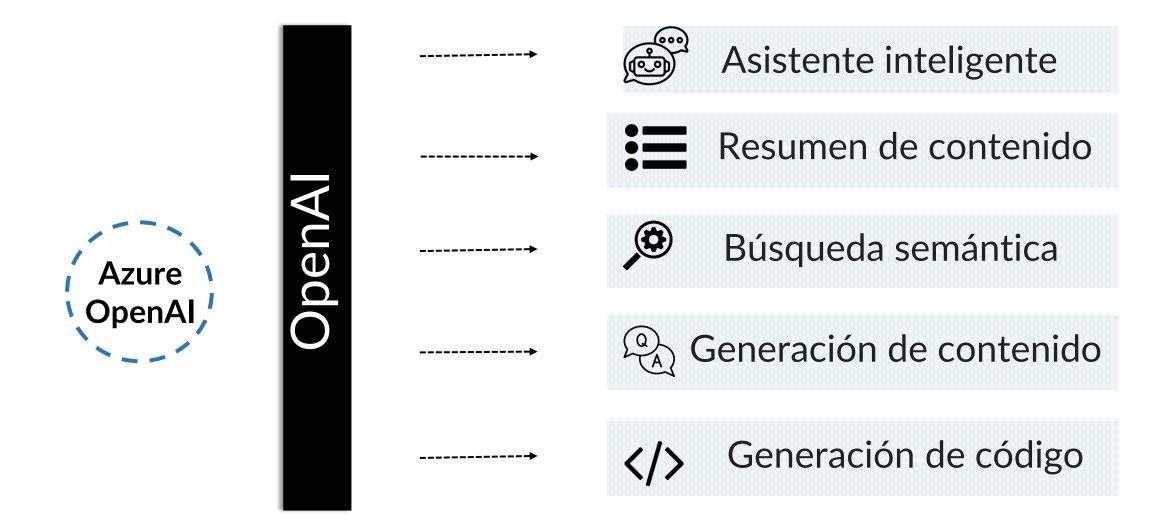
Chat session Show raw JSON hello, who are you? ф Hello! I am an AI assistant designed to help HR professionals identify relevant resumes by extracting CVs that correspond to job descriptions. How can I assist you today? I'm seeking a data scientist who has specialized knowledge in machine learning, specifically in natural language processing. They should have a profound grasp of transformer architectures and expertise in utilizing models such as BERT





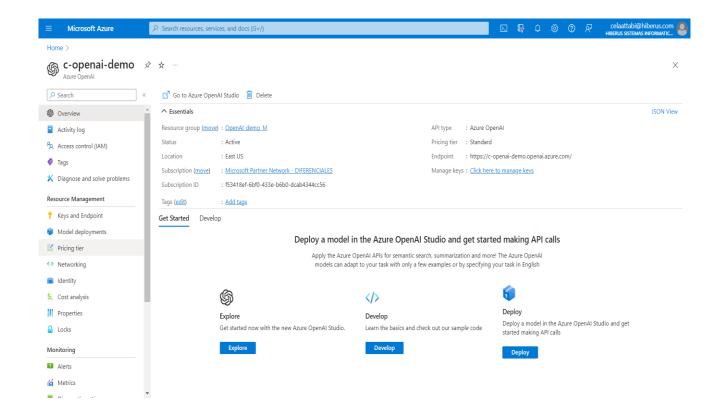






#somoshiberus

Como acceder e interactuar con la API de OpenAI



```
https://c-openai-demo.openai.azure.com/
                                                         python
                                                        json
 1 #Note: The openai-python library support for Azure
preview.
                                                       python
 2 import os
 3 import openai
                                                        C#
 4 openai.api type = "azure"
 5 openai.api_base = "https://c-openai-demo.openai.azu
                                                        curl
 6 openai.api_version = "2022-12-01"
 7 openai.api_key = os.getenv("OPENAI_API_KEY")
 9 response = openai.Completion.create(
      engine="chagpt model",
      prompt="",
      temperature=1,
13
      max_tokens=100,
      top p=0.5,
      frequency penalty=0,
      presence_penalty=0,
      stop=None)
```

Categorías de filtrado de contenido

Category	Descripción
Odio	La categoría de odio describe los ataques o usos del lenguaje que incluyen un lenguaje peyorativo o discriminatorio con referencia a una persona o grupo de identidad sobre la base de ciertos atributos diferenciadores de estos grupos, incluidos, entre otros, la raza, la etnia, la nacionalidad, la identidad y expresión de género, la orientación sexual, la religión, el estatus migratorio, el estado de capacidad, la apariencia personal y el tamaño corporal.
Sexual	La categoría sexual describe el lenguaje relacionado con los órganos anatómicos y los genitales, las relaciones románticas, los actos representados en términos eróticos o afectuosos, los actos sexuales físicos, incluidos los representados como una agresión o un acto violento sexual forzado contra la propia voluntad, la prostitución, la pornografía y el abuso.
Violencia	La categoría de violencia describe el lenguaje relacionado con acciones físicas destinadas a herir, lesionar, dañar o matar a alguien o algo; describe armas, etc.
Autolesiones	La categoría de autolesiones describe el lenguaje relacionado con acciones físicas destinadas a herir, lesionar o dañar intencionadamente el propio cuerpo, o a suicidarse.

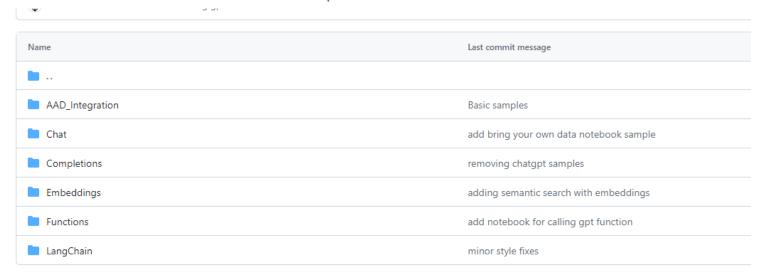
Category	Descripción
Caja fuerte	El contenido puede estar relacionado con categorías de violencia, autolesiones, sexualidad u odio, pero los términos se usan en contextos generales, periodísticos, científicos, médicos y profesionales similares, que son apropiados para la mayoría del público.
Bajo	Contenido que expresa prejuicios, juicios u opiniones, incluyen el uso ofensivo del lenguaje, los estereotipos, los casos de uso que exploran un mundo ficticio (por ejemplo, los juegos, la literatura) y las representaciones con baja intensidad.
Media	El contenido que usa un lenguaje ofensivo, insultante, burlón, intimidatorio o degradante hacia grupos de identidad específicos, incluye representaciones de búsqueda y ejecución de instrucciones dañinas, fantasías, glorificación, promoción del daño con una intensidad media.
Alto	Contenido que muestra instrucciones, acciones, daños o abusos explícitos y gravemente perjudiciales; incluye la aprobación, glorificación o promoción de actos gravemente perjudiciales, formas extremas o ilegales de daño, radicalización o intercambio o abuso de poder no consentido.

#somoshiberus

Welcome *∂*

This repo is a compilation of useful Azure OpenAl Service resources and code samples to help you get started and accelerate your technology adoption journey.

The Azure OpenAI service provides REST API access to OpenAI's powerful language models on the Azure cloud. These models can be easily adapted to your specific task including but not limited to content generation, summarization, semantic search, and natural language to code translation. Users can access the service through REST APIs, Python SDK, or our web-based interface in the Azure OpenAI Studio.



https://github.com/Azure-Samples/openai/tree/main/B asic Samples