AMALIA CONF
by COMUNIDAD DOJO

Desplegando aplicaciones de Large Language Models (LLMs) en la nube de Microsoft Azure.

Noviembre 25, 2023, a las 3:00 P.M.

Kevin Knights
Machine Learning Engineer

MICROSOFT CERTIFIED TRAINER
Microsoft
MCT
Trainer
2023 - 2024

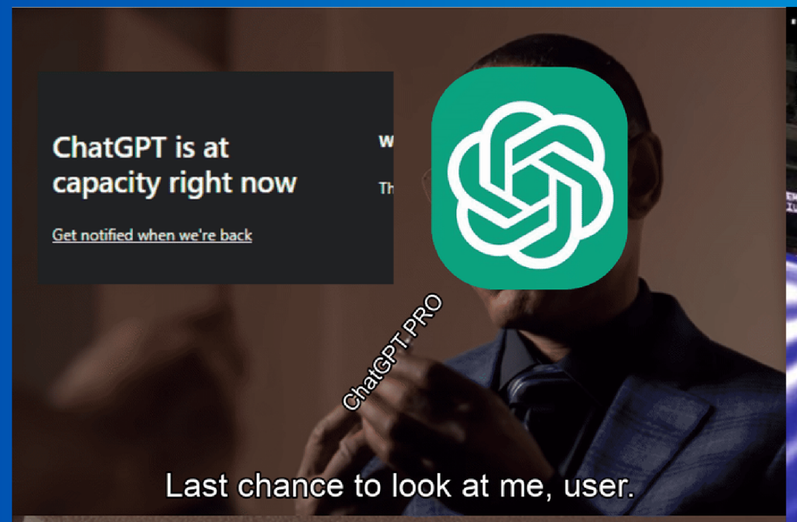
Hola, soy Kevin Knights!

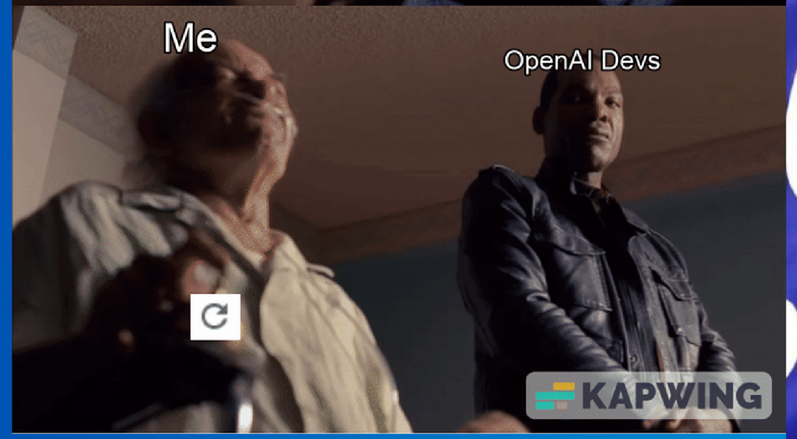
Machine Learning Engineer

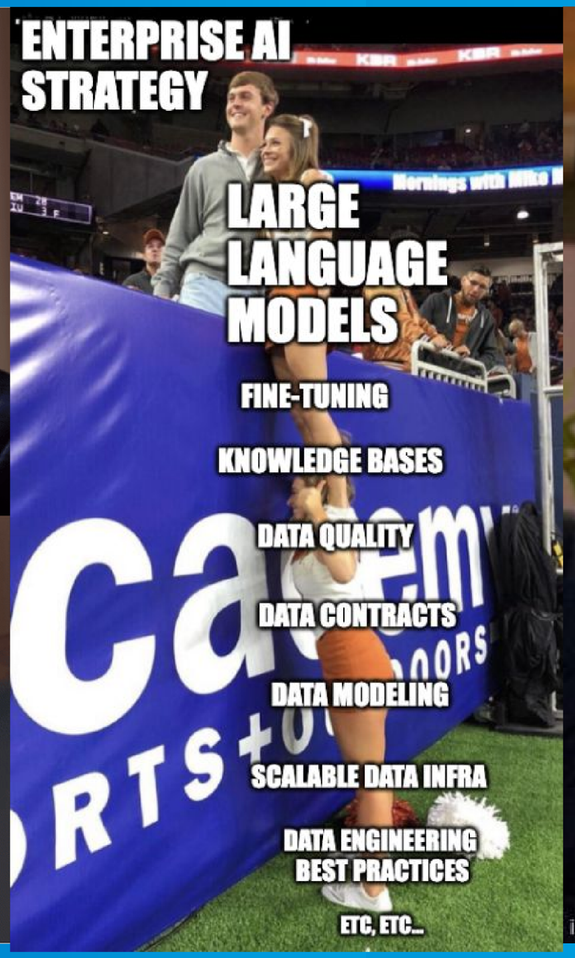
- He desarrollado varios productos GenAI.
- He participado en conferencias globales sobre IA.
- Certificado en AI de Stanford, Microsoft, entre otros...

LinkedIn: Kevin Knights
https://www.linkedin.com/in/knightsk/

¡Todo ahora es con LLMs (GenAI)!





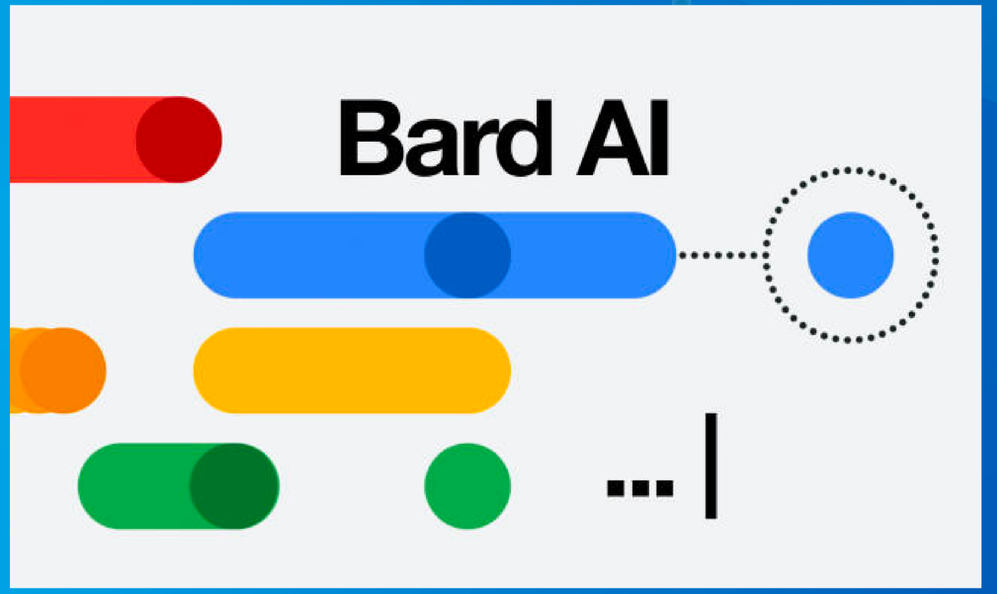


Google dijo 'AI' más de 140 veces en su discurso de apertura del Google I/O de 2 horas

¿Qué es GenAI?

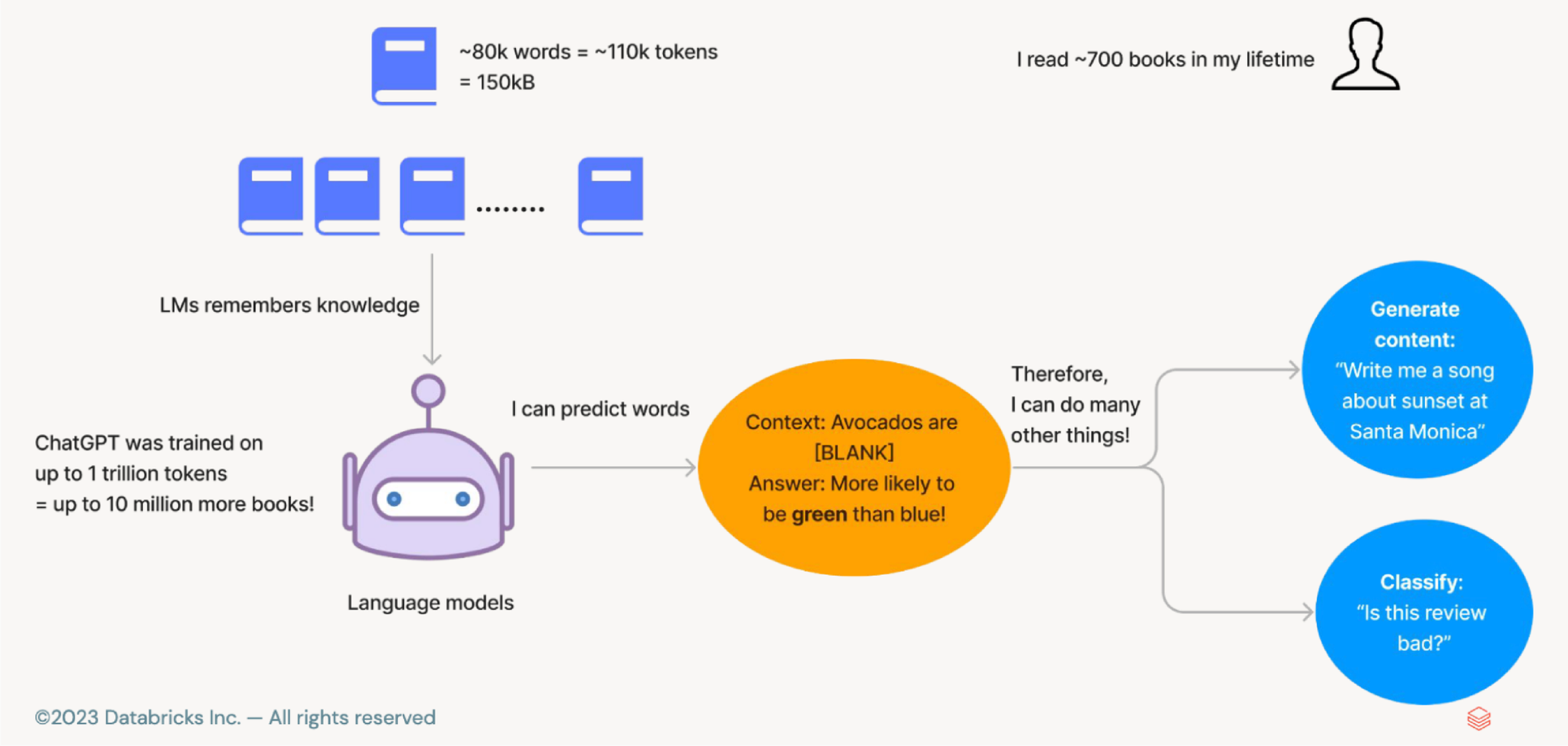
La inteligencia artificial generativa (GenAI) describe algoritmos (como ChatGPT) que se pueden utilizar para crear contenido nuevo, incluido audio, código, imágenes, texto, simulaciones y videos.

Fuente: McKinsey – What is generative AI?



¿Qué son los Large Language Models?

Los modelos de lenguaje grande (LLM) son modelos de GenAI que pueden comprender y generar texto en lenguaje humano.



¿Por qué implementar LLMs?

McKinsey estima que GenAI (LLMs) agregará entre
2,6 y 4,4 billones de dólares en valor anual a la economía global.

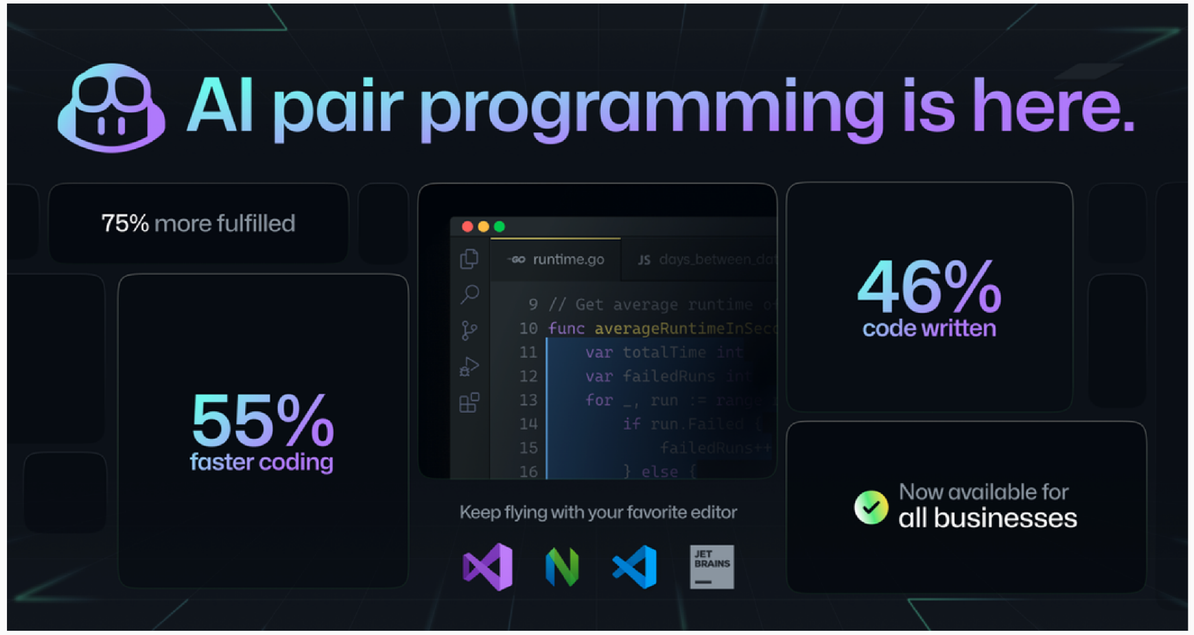
40%

El nuevo impacto económico de la IA
como un todo, gracias a GenAI (LLMs).

No puedo pensar en nada que haya sido más poderoso desde la computadora de escritorio. – Michael Carbin, MIT, MosaicML.

Fuente: MIT Tech Report – Generative AI Report

LLMs y la transformación digital

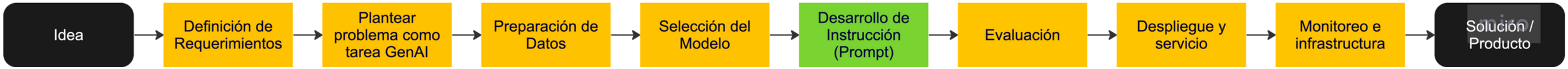


Github Copilot, herramienta de generación de código

- GenAI y los LLM están democratizando el acceso a la IA, provocando los inicios de una IA verdaderamente empresarial.
- Ahora, una gran cantidad de datos no estructurados, pueden ser aprovechados para generar valor empresarial.
- Los LLMs liberarán a la fuerza laboral del trabajo que requiere mucho tiempo, para centrarse en áreas de conocimiento, estrategia y tareas de alto valor comercial.

Fuente: MIT Tech Report – Generative AI Report

¿Cómo construir aplicaciones con LLMs?



Marco para construir soluciones de GenAI.

1. Definición de requerimientos.
2. Plantear el problema como una tarea de ML.
3. Preparación de datos.
4. Selección del modelo.
5. Desarrollo de instrucción (prompt)
6. Evaluación.
7. Despliegue y servicio.
8. Monitoreo e infraestructura.

LangChain

Streamlit

qdrant

Python

MS Azure

OpenAI

¿Cómo construir LLMs apps en Azure?

Administrar imágenes de Docker

Despliegue de aplicaciones contenerizadas con k8s

Servicios y APIs de LLMs

Construyendo una solución de GenAI

Idea:
Aplicación para generar contenido de redes sociales.

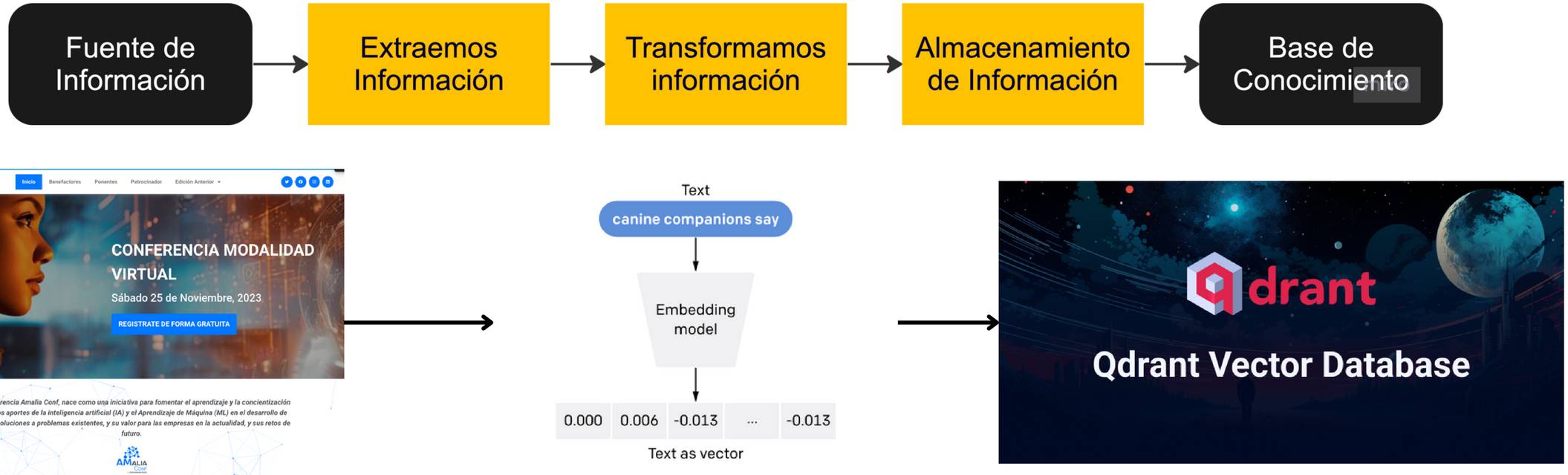
Objetivo de Negocio:
Posicionar productos de forma personalizada.

Requerimientos:
- Interfaz.
- Subir documentos.
- Personalizar contenido.

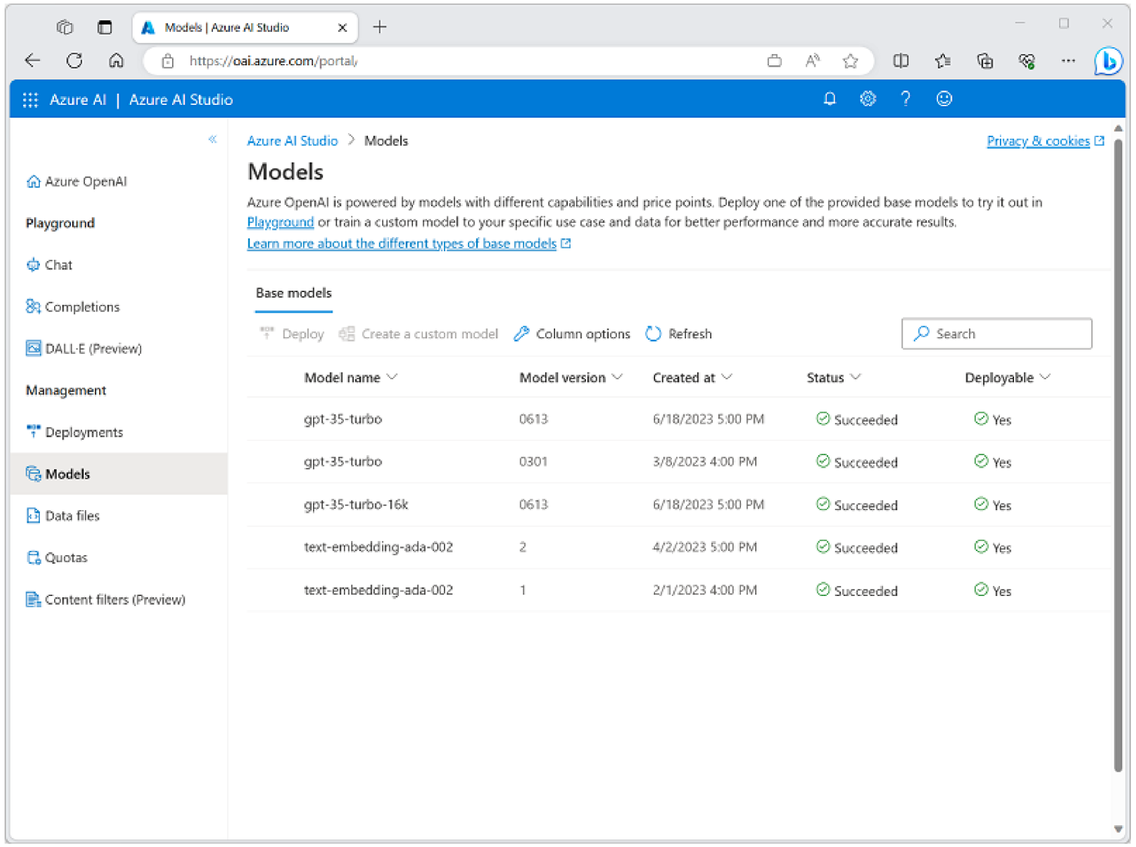
Objetivo de LLM:
Maximizar la personalización del contenido.

Construyendo una solución con LLMs

Preparación de datos:



Construyendo una solución con LLMs



Selección del modelo:

gpt3.5-turbo de OpenAI [1]
text-embedding-ada-002 [2]

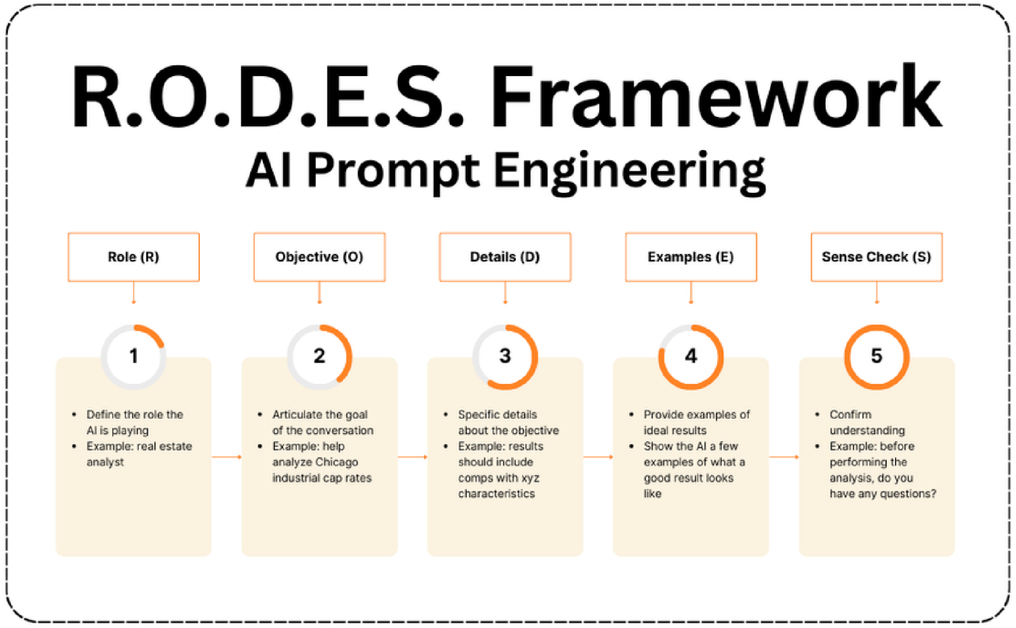
- Modelo con alta precisión.
- Buena calidad/precio.
  - ~ 750 palabras = 0.002 USD [1]
  - ~750 palabras = 0.0001 USD [2]
- Fácil integración

Precios de Open AI API: https://openai.com/pricing

Construyendo una solución con LLMs

Desarrollo de instrucción (Prompt)

Cree una publicación en las redes sociales visualmente impactante y emocionalmente atractiva que refleje la experiencia de un especialista en marketing digital.
El contenido debe ser altamente personalizado, fomentando la interacción a través de personas identificables.
Narraciones o preguntas que inviten a la reflexión.
La publicación debe ser adecuada para un público diverso, alineado con los valores de la marca,
y optimizado para una alta participación en las plataformas de redes sociales.



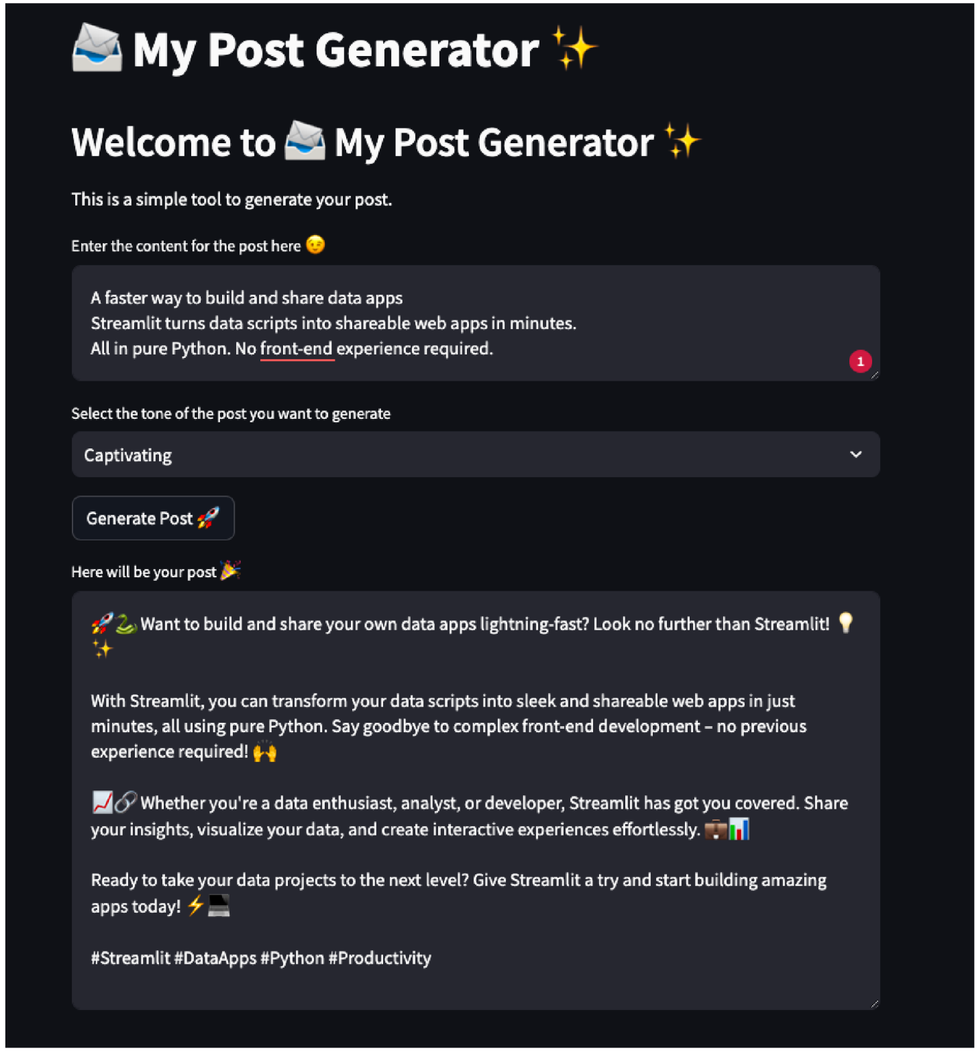
Ingeniería de prompts: https://www.promptingguide.ai/

Construyendo una solución con LLMs

Evaluación:

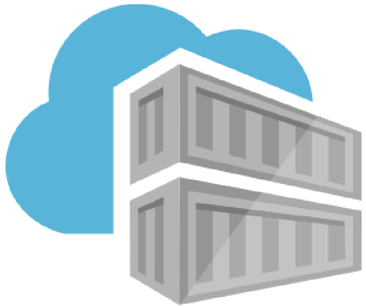
El modelo es capaz de responder basándose en la base de conocimiento creada.

Si quisiéramos que el modelo sea más “conversón” podemos ajustar parámetros dentro del modelo.



Construyendo una solución de GenAI

Despliegue:

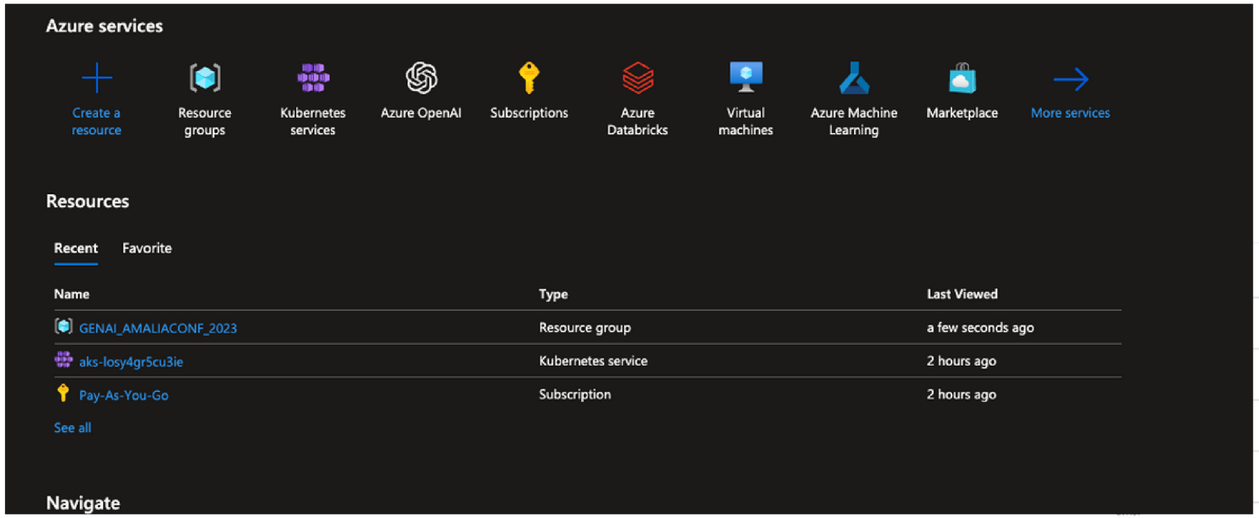


Azure Container Registry

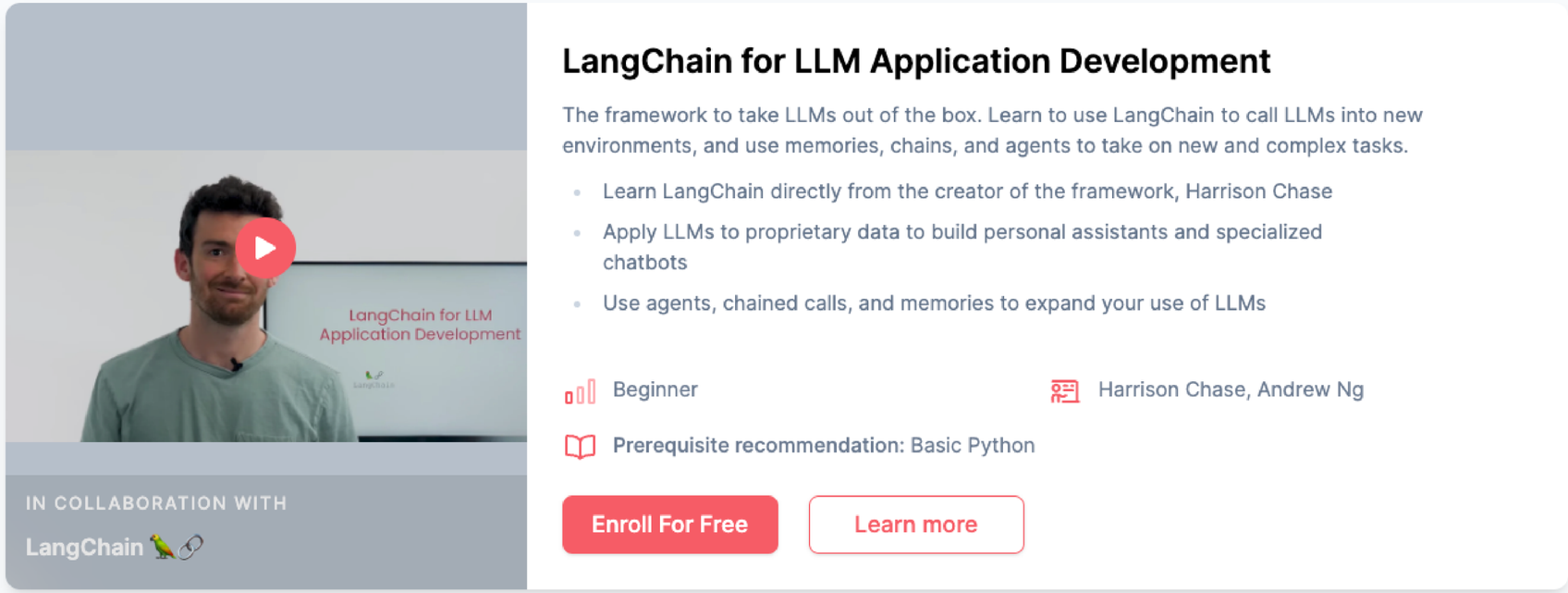
Azure Kubernetes Service(AKS)

Construyendo una solución de GenAI

Monitoreo:

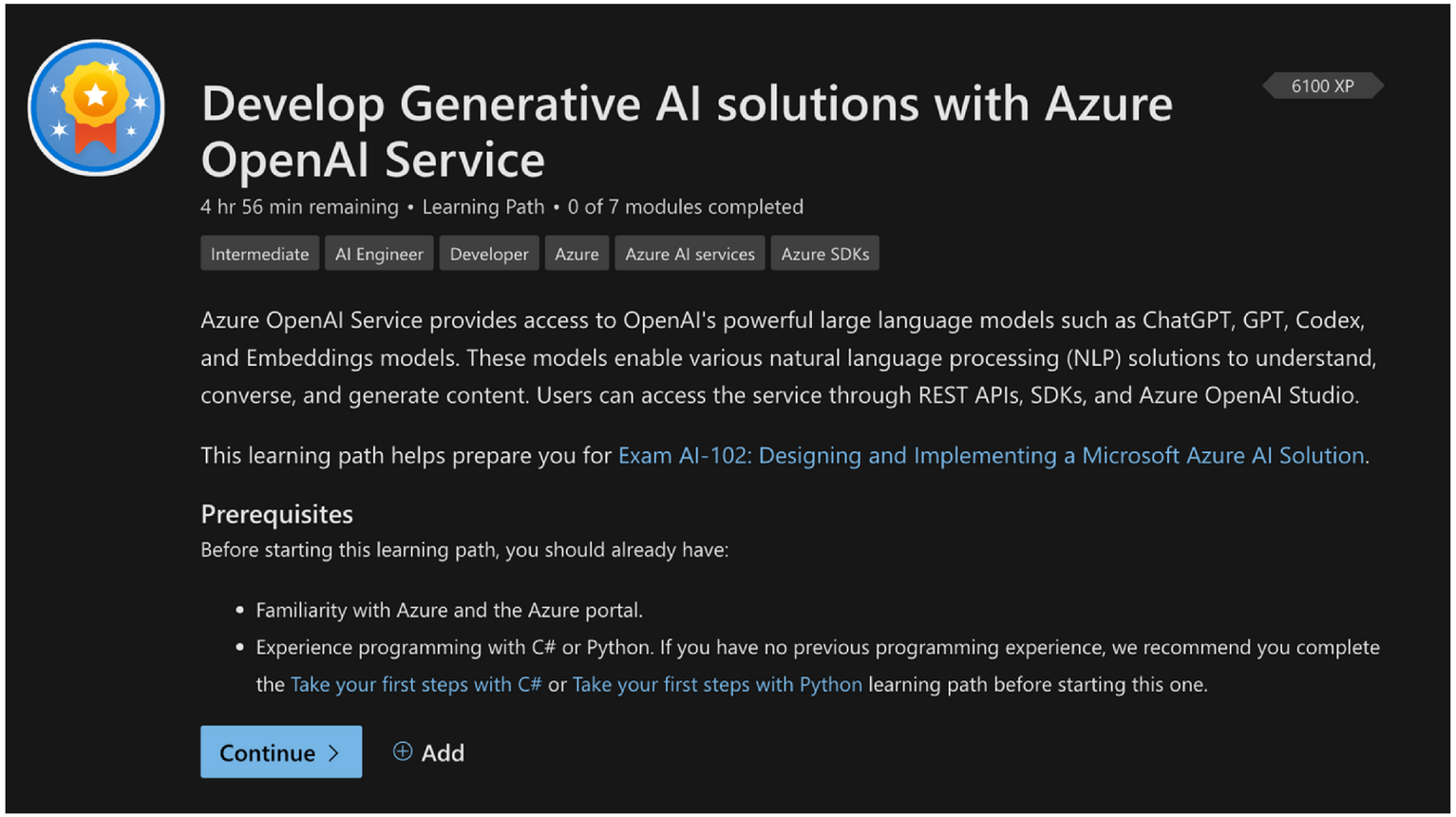


Recursos para aprender más sobre LLMs



Enlace: DeepLearningAI short courses

Recursos para aprender más sobre LLMs



Para obtener más información sobre cómo crear soluciones con el servicio Azure OpenAI.

Enlace: Microsoft Learn

OpenAI | Microsoft

Gracias!

Scanea el QR para acceder al app

Scan me!

¿Cómo contactarme?

- **LinkedIn**: Kevin Knights
  https://www.linkedin.com/in/knightsk/
- **Gmail**: kevin.k.knights@gmail.com