PREPARACIÓN PARA LA CERTIFICACIÓN DE SOCIOS DE AWS

AI PRACTITIONER

Sesión de revisión de contenido 2

Dominio 2

Suscripción *opcional* a AWS Skill Builder

La suscripción a Skill Builder proporciona acceso a exámenes oficiales de práctica de certificación de AWS, contenido de capacitación digital a tu propio ritmo, incluidos desafíos abiertos, laboratorios a tu propio ritmo y aprendizaje basado en juegos. **Ten en cuenta que no se requiere la suscripción a Skill Builder para este programa Acelerador.**

Capacitación digital gratuita

ENLACE AQUÍ

Las características especiales incluyen:

- Más de 600 cursos digitales.
- Planes de aprendizaje.
- 10 sets de preguntas de práctica.
- *AWS Cloud Quest (Fundacional).*
- *AWS Simulearn*

Suscripción individual

ENLACE AQUÍ

Todo en la formación digital gratuita, además de:

- AWS Cloud Quest (Intermedio - Avanzado).
- Exámenes de práctica oficiales de certificación.
- Cursos de Preparación para Exámenes Mejorados.
- Acceso ilimitado a más de 1000 laboratorios prácticos.
- AWS Jam Journeys (desafíos basados en laboratorio).
- Aula digital de AWS (solo anual).

Acceso completo al plan de preparación para el examen.

AWS Cloud Quest es recomendado para tener experiencia práctica.

Las suscripciones individuales tienen un precio de $29 USD al mes (*Flexibilidad para cancelar en cualquier momento*) o $449 USD por año.

Resumen del programa

Sesión 1

Kickoff

Dominio 1:
Fundamentos de IA y ML

Sesión 2

Dominio 2:

Fundamentos de la IA Generativa

Sesión 3

Dominio 3:
Aplicaciones de Foundation Models

Sesión 4

Dominio 4: Directrices para la IA Responsable

Dominio 5: Seguridad, Cumplimiento y Gobernanza para Soluciones de IA

Plan de aprendizaje de AWS Artificial Intelligence Practitioner

Plan semanal de estudio digital

¿Qué recursos estudio esta semana? Haz todo lo posible para completar el contenido de los cursos de esta semana en AWS Skill Builder

Cursos del plan de aprendizaje
Fundamentals of Machine Learning and Artificial Intelligence
Exploring Artificial Intelligence Use Cases and Applications
Responsible Artificial Intelligence Practices

Plan completo de preparación para el examen (Optional)
Empezar – CloudQuest: Generative AI Practitioner; CloudQuest: Generative AI Architect*
Exam Prep Plan Overview
Domain 1 Review; Domain 1 Practice*
AWS SimuLearn: Document Handling Using Amazon Textract and Amazon Polly*
Domain 2 Review; Domain 2 Practice*

* Requiere suscripción a AWS Skill Builder

Resultados de aprendizaje de hoy

Durante esta sesión, cubriremos:

- Enunciado de tarea 2.1: Explicar los conceptos básicos de IA generativa.
- Enunciado de tarea 2.2: Comprender las capacidades y limitaciones de la IA generativa para resolver problemas comerciales.
- Enunciado de tarea 2.3: Describir la infraestructura y las tecnologías de AWS para crear aplicaciones de IA generativa.

PREPARACIÓN PARA LA CERTIFICACIÓN DE SOCIOS DE AWS

Dominio 2: Fundamentos de la IA Generativa

Enunciado de tarea 2.1: Explicar los conceptos básicos de IA generativa.

Modelo Fundacional (Foundation Model - FM)

1

Datos sin etiquetas

2

Pre-entrenar

3

Modelo Fundacional

4

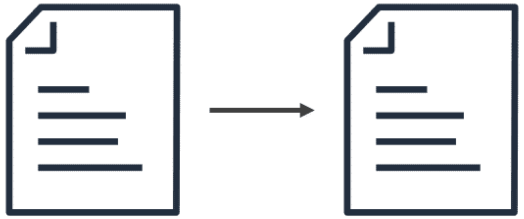
Adaptar

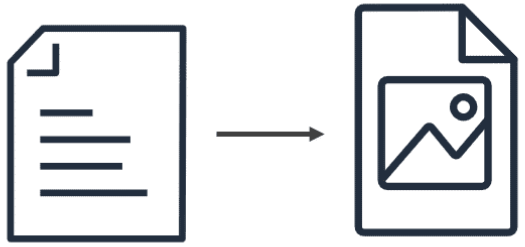
- **Resumen de contenido**
- **Generación de contenido**
- **Generación de código**
- **Preguntas y respuestas**

5

Amplio rango de tareas generales

Tipos de FMs





Modelos de texto a texto

- Los modelos de texto a texto son modelos de lenguaje grande (LLM) que están preentrenados para procesar grandes cantidades de datos textuales y lenguaje humano.
- Procesamiento del lenguaje natural (PNL).

Modelos de texto a imagen

- Los modelos de texto a imagen toman entrada de lenguaje natural y producen una imagen de alta calidad que coincide con la descripción del texto de entrada.
- Arquitectura de difusión.

Componentes de FMs

Datos sin etiquetar

- Más fácil de obtener en comparación con los datos etiquetados.
- Los modelos de preentrenamiento toman en cuenta el contexto a partir de todos estos datos de entrenamiento.
  - Realiza un seguimiento de las relaciones en datos secuenciales.

Modelo grande

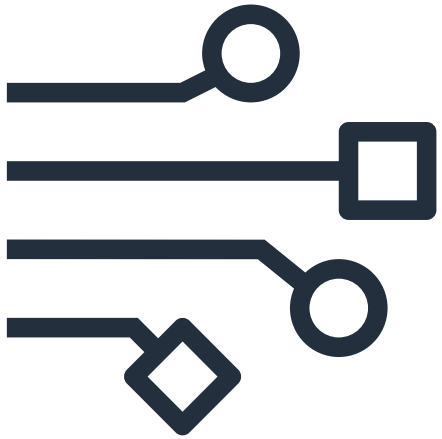
- Miles de millones de parámetros.
- Los modelos de preentrenamiento de este tamaño requieren acceso a:
  - Cantidad y calidad suficientes de los datos de entrenamiento
  - Infraestructura de entrenamiento a gran escala.

¿Cómo se procesan estos datos sin etiquetar?

Transformer

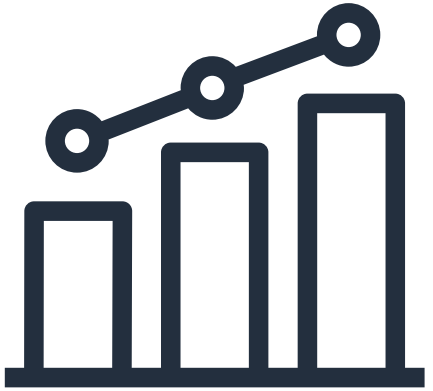
¿Qué es?	¿Cómo funciona?	¿Qué hace?
El Transformer es un tipo específico de red neuronal que potencia los modelos fundacionales mediante el procesamiento de secuencias de información	Analiza las relaciones entre palabras o imágenes para comprender el contexto y el significado	Transforman las entradas en salidas relevantes utilizando su comprensión de las relaciones entre palabras

Beneficios del Transformer



Procesamiento paralelo

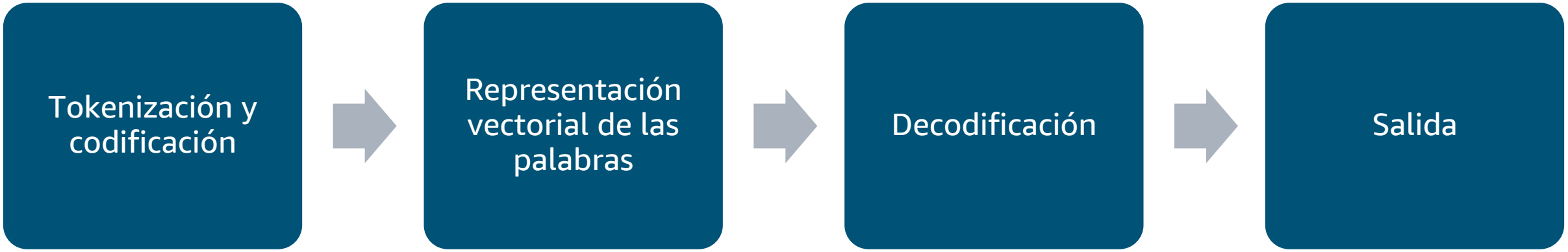
Mecanismo de atención



Flexibilidad y escalabilidad

Cachorro es a perro como pollito es a...

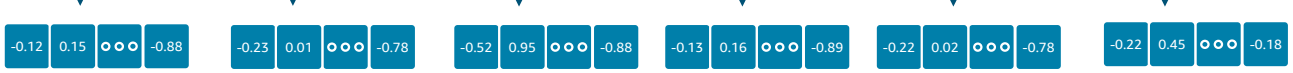
Cómo un modelo Transformer completa una oración



Tokenización y codificación

“Cachorro es a perro como pollito es a...”

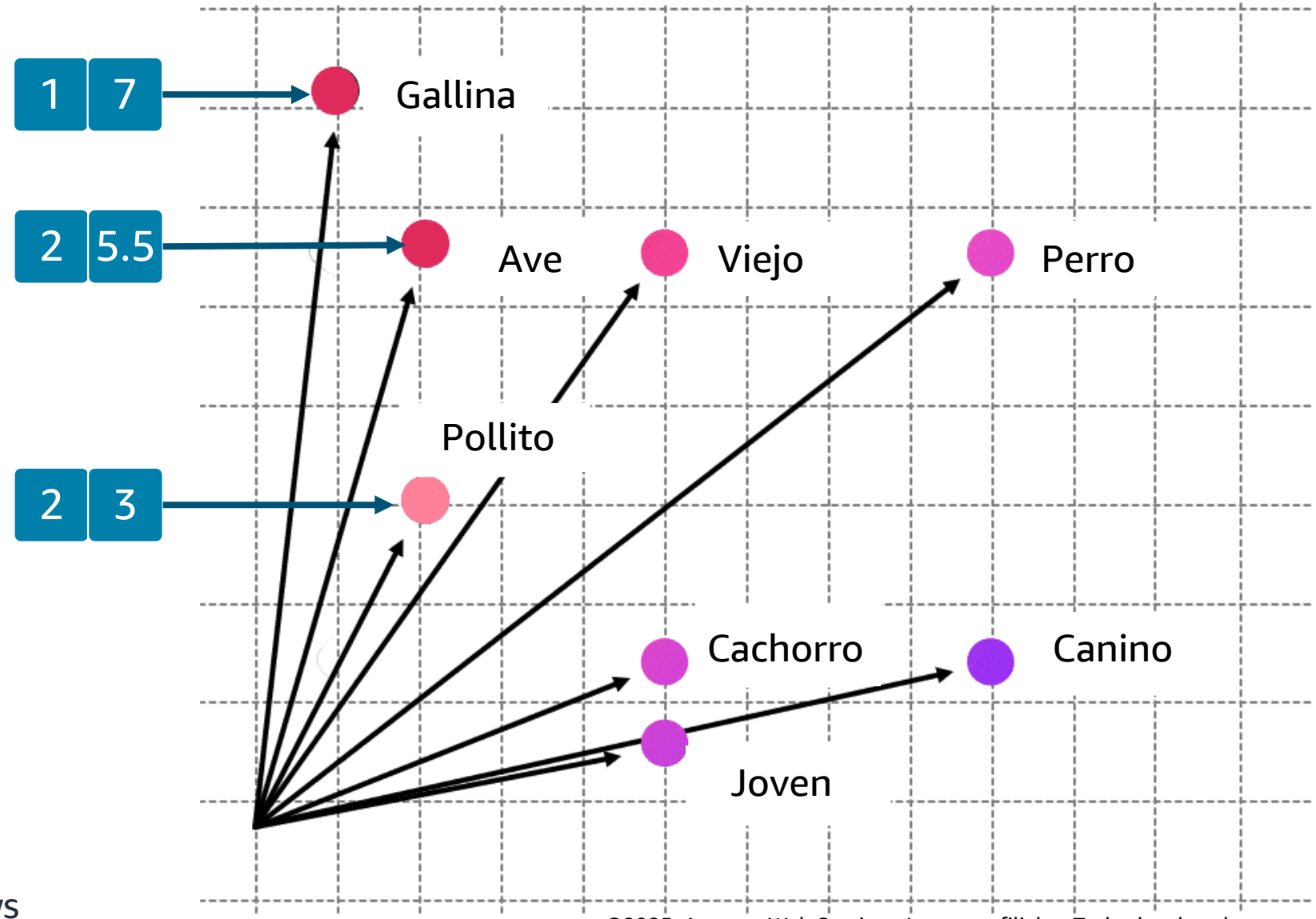
[CLS]	Cachorro	es	a	perro	como	gatito	es	a	[SEP]
101	17022	2003	2000	3899	2004	18401	2003	2000	102



Paso 1
Convertir palabra a token

Paso 2
Codificar (convertir) el significado de cada palabra / token en una **representación vectorial embebida (embeddings)**

Representación vectorial de las palabras (Word embedding)



Los vectores embebidos existen en este espacio multidimensional

Este proceso resulta en que las palabras con significados similares se ubiquen más cerca unas de otras en este espacio vectorial.

Nota: En esta diapositiva solo se muestran 2 dimensiones (X e Y). En la práctica, podría haber cientos o miles de dimensiones"

Decodificación

Cachorro es a perro como pollito es a _____.

gallina 0.30
gato 0.25
oso 0.20
humano 0.15

Salida:

Cachorro es a perro como pollito es a **gallina**.

PREPARACIÓN PARA LA CERTIFICACIÓN DE SOCIOS DE AWS

Dominio 2: Fundamentos de la IA Generativa

Enunciado de tarea 2.2: Comprender las capacidades y limitaciones de la IA generativa para resolver problemas comerciales.

Parámetros de inferencia FM - Aleatoriedad y diversidad

Filtrado y limitación de tokens	Control de Creatividad
• Top k • Top p	• Temperatura

Top k

Selecciona los k tokens más probables de todo el vocabulario

Top k = 1

gallina 0.30

Top k = 2

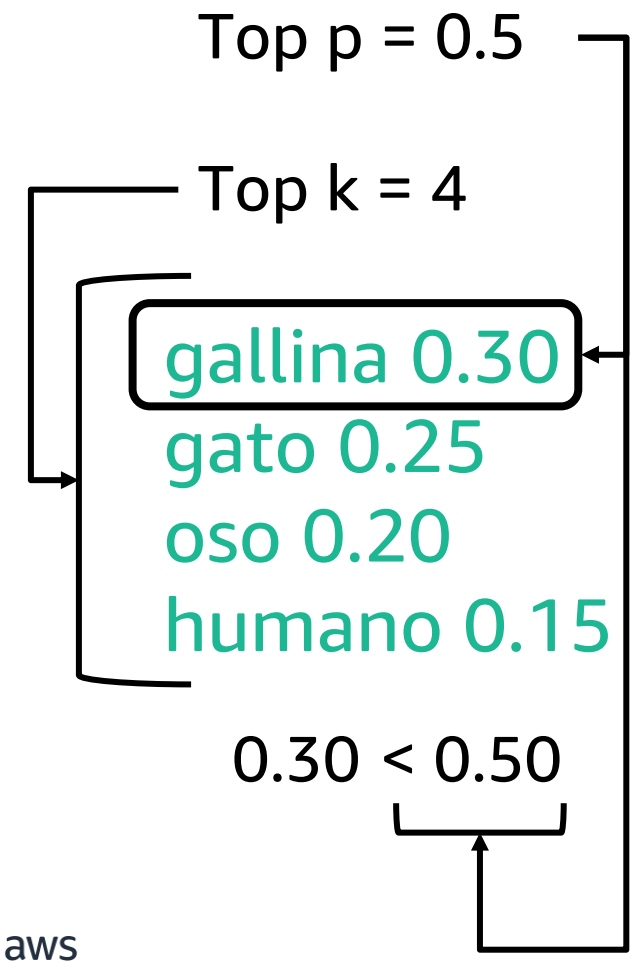
gallina 0.30
gato 0.25

Top k = 4

gallina 0.30
gato 0.25
oso 0.20
humano 0.15

Top p

Selecciona tokens en función de la probabilidad acumulativa hasta que se alcance el umbral.



Top p = 0.60

Top k = 4

gallina 0.30
gato 0.25
oso 0.20
humano 0.15

0.30 + 0.25 < 0.60

Top p = 0.85

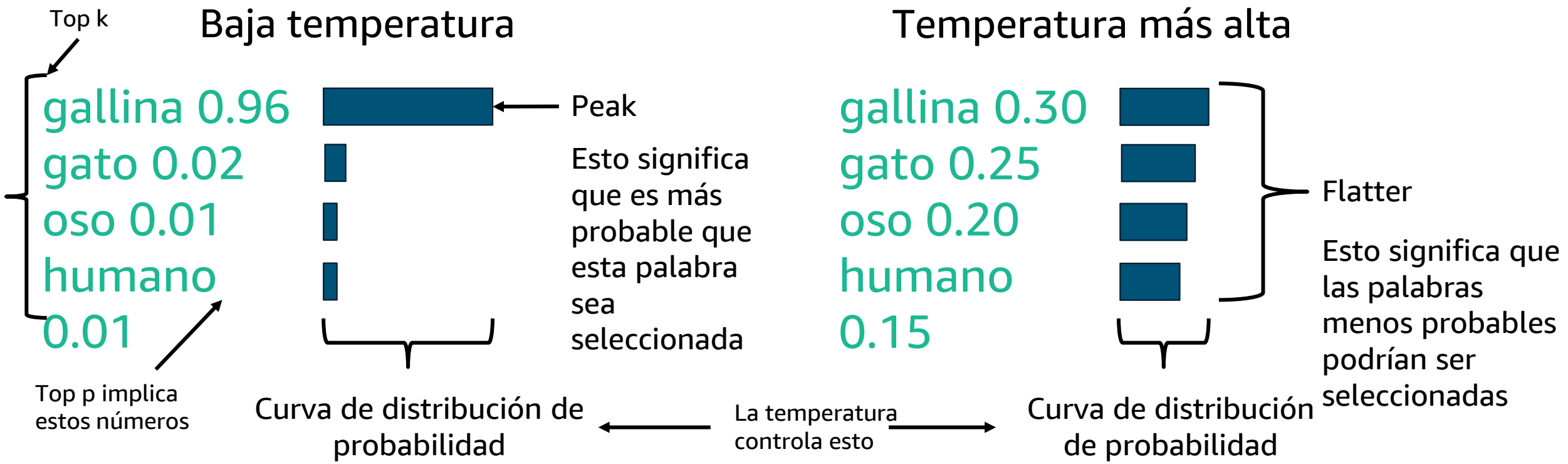
Top k = 4

gallina 0.30
gato 0.25
oso 0.20
humano 0.15

0.30 + 0.25 + 0.20 < 0.85

Temperatura

Controla la aleatoriedad y la creatividad modificando la forma de la distribución de probabilidad del siguiente token. Baja temperatura significa menor aleatoriedad y creatividad. La temperatura más alta da como resultado más aleatoriedad y creatividad.



Parámetros de inferencia de los FM – Longitud

Longitud de la respuesta

- Un valor exacto para especificar el número mínimo o máximo de tokens que se devolverán en la respuesta generada.
- Ayuda a gestionar los recursos computacionales y los costos

Penalizaciones

- Especificar el grado en que se penalizarán las salidas en una respuesta.

Ejemplos:

- La longitud de la respuesta.
- Tokens repetidos en una respuesta.
- Frecuencia de tokens en una respuesta.
- Tipos de tokens en una respuesta.

Secuencias de detención

- Especificar secuencias de caracteres que detienen al modelo para que no genere más tokens.
- Si el modelo genera una secuencia de parada que usted especifica, dejará de generar después de esa secuencia.

Contexto

El contexto es un intercambio privado entre el usuario y el modelo.

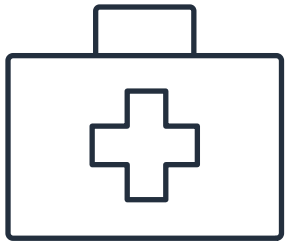
- No persiste.
- Hay un límite superior en el número de tokens.
- Se puede perder la información inicial que está utilizando el modelo.

Ejemplo

El Transformer tiene que averiguar a qué se refiere esto en el contexto de la conversación actual.

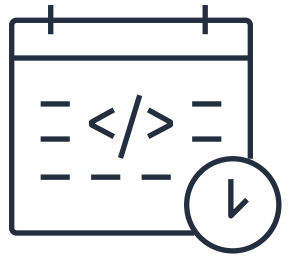
Prompt (Interacción 1)
Hola, ¿cuál es el mejor lugar en Seattle para visitar?
Respuesta:
El Columbia Center ofrece impresionantes vistas del horizonte de la ciudad [...].
Preguntar (Interacción 2, continuar interacción):
¿Será esto divertido para los niños?
Respuesta:
No. Hay lugares mucho mejores para niños en Seattle.

Preocupaciones de IA generativa



Toxicidad

Alucinaciones



Propiedad intelectual

Plagio y trampa

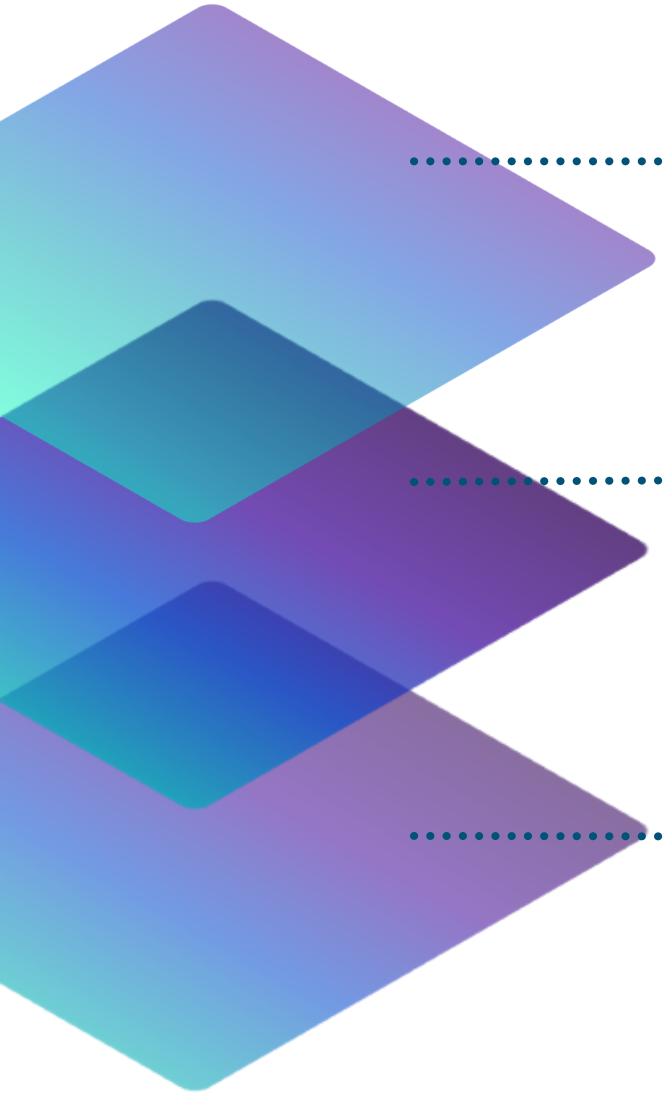
Disrupción de la naturaleza del trabajo

PREPARACIÓN PARA LA CERTIFICACIÓN DE SOCIOS DE AWS

Dominio 2: Fundamentos de la IA Generativa

Enunciado de tarea 2.3: Describir la infraestructura y las tecnologías de AWS para crear aplicaciones generativas de IA.

Recursos de IA generativa de AWS



APLICACIONES QUE APALANCAN LLM Y OTRAS FMs

Amazon Q Business | Amazon Q Developer | Amazon Q en QuickSight | Amazon Q en Connect

PERSONAS

Negocio

HERRAMIENTAS PARA CONSTRUIR CON LLM Y OTROS FMs

Amazon Bedrock

Normativas| Agentes | Capacidades de personalización | Importación de modelos personalizados

Creador de aplicaciones

INFRAESTRUCTURA PARA ENTRENAMIENTO E INFERENCIA DE FM

GPUs | Trainium | Inferentia | SageMaker

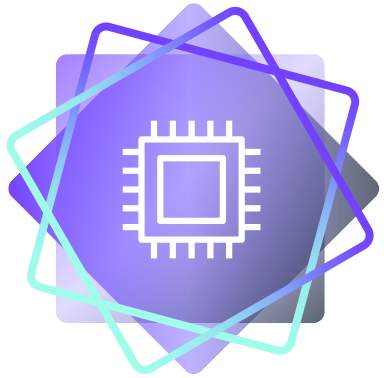
UltraCLUsters | EFA | EC2 Capacity Blocks | Nitro | Neuron

Científico de datos

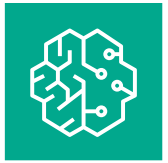
aws

Innovando a nivel de silicio

AWS Trainium



AWS Inferentia



Amazon SageMaker AI

Crea, entrena e implementa modelos de ML a escala, incluidos los FMs.

Acceda a los últimos FMs disponibles públicamente a través de SageMaker Jumpstart.

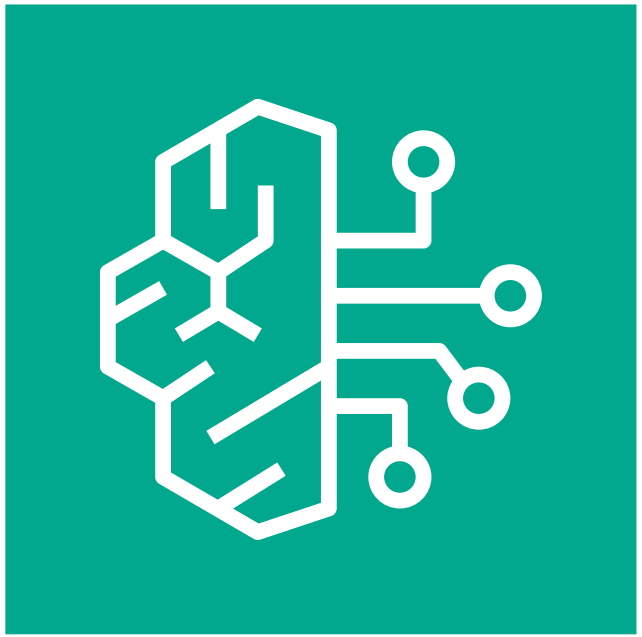
Construye FMs desde cero.

Personaliza FMs.

Ejecuta inferencia.

Implementa FMOPs y gobernanza.

Amazon Bedrock



Elección de FMs líderes a través de una única API.

Personalización del modelo.

Generación Aumentada de Recuperación (RAG).

Agentes que ejecutan tareas de varios pasos.

Seguridad, privacidad y gobierno.

Amazon Q

NEGOCIO

Amazon Q Negocios

BÚSQUEDA DE CONOCIMIENTO

RESUMEN

CREACIÓN DE CONTENIDO

EXTRAER CONOCIMIENTOS

Investigación y análisis

Amazon Q en QuickSight

ENTENDER LOS DATOS

CONSTRUYE Y REFINE VISUALES

CÁLCULOS DE CONSTRUCCIÓN

RESÚMENES EJECUTIVOS

CREAR HISTORIAS DE DATOS

DESARROLLADORES

Amazon Q Developer

APLICACIÓN DEL PLAN

GENERACIÓN DE CÓDIGO

PRUEBAS UNITARIAS

ESCANEO DE SEGURIDAD

REMEDIACIÓN DE CÓDIGO

MIGRACIÓN DE CÓDIGO

SOLUCIÓN DE PROBLEMAS

CONOCIMIENTO DEL DESARROLLADOR

USUARIOS ESPECIALIZADOS

Amazon Q in Connect

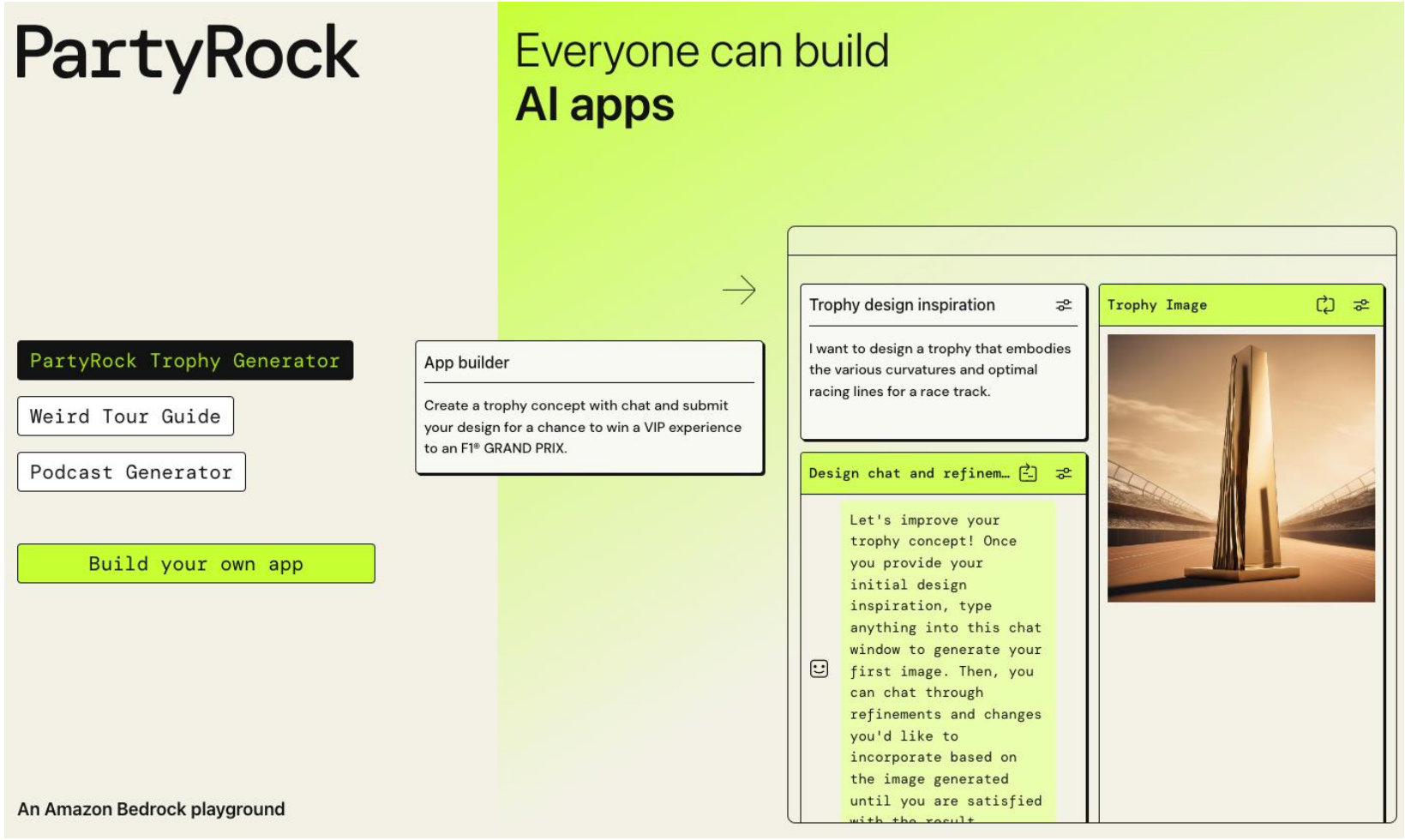
AGENTE DE ASISTENCIA

Amazon Q in AWS Supply Chain

CADENA DE SUMINISTRO

PartyRock

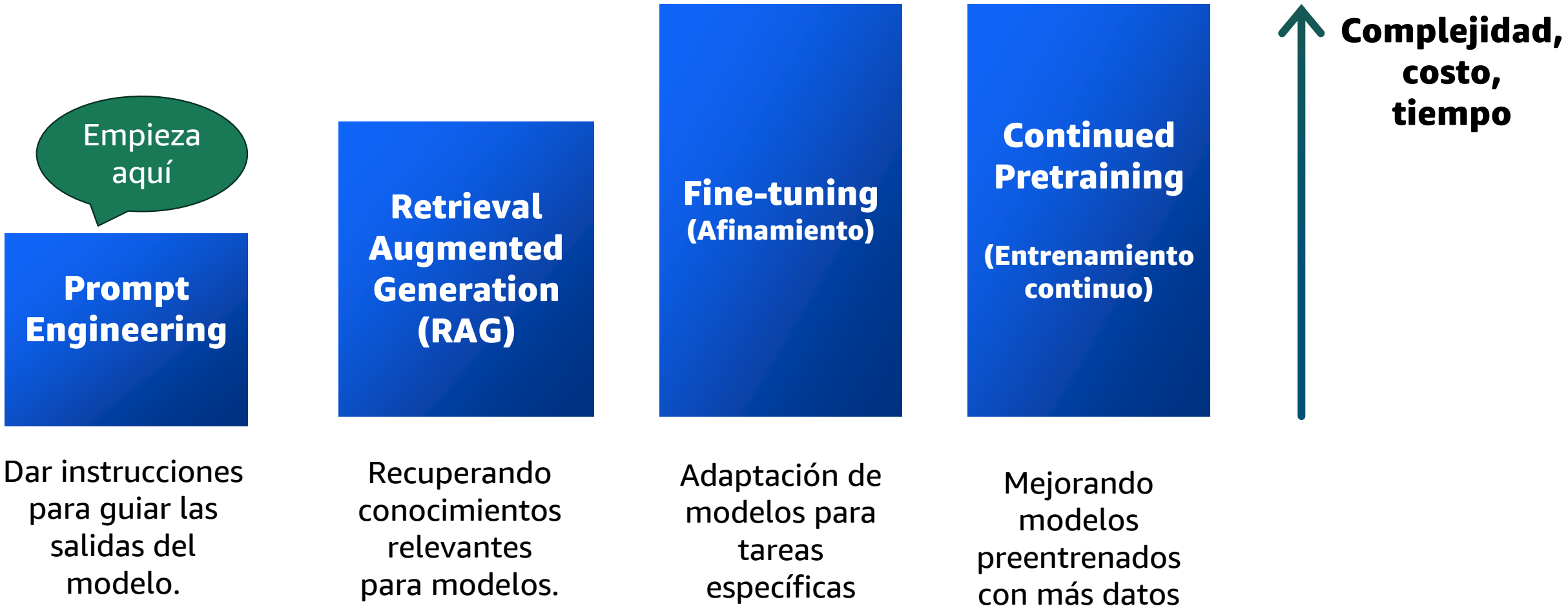
- Crea aplicaciones generadas por IA impulsadas por Amazon Bedrock.
- Permite a los usuarios crear, compartir y remezclar aplicaciones de IA para diversas tareas divertidas.
- Enseña habilidades fundamentales de IA generativa a través de la experimentación práctica.



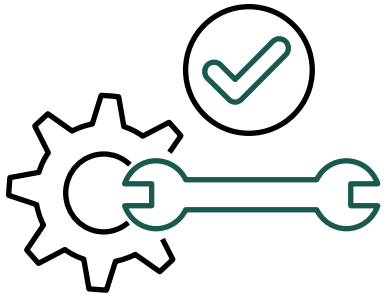
PREPARACIÓN PARA LA CERTIFICACIÓN DE SOCIOS DE AWS

Entrenamiento de modelos y afinación

Enfoques comunes para personalizar FMs



Personalización de las respuestas del modelo para su negocio



Afinamiento

PROPÓSITO

Maximizar la precisión para tareas específicas.

NECESIDAD DE DATOS

Pequeño número de ejemplos etiquetados.

Pre-entrenamiento continuo

PROPÓSITO

Mantener la precisión del modelo para su dominio.

NECESIDAD DE DATOS

Gran cantidad de conjuntos de datos sin etiquetar.

Recursos adicionales

Documentación útil

Plan de estudios de formación digital semanal

¿Qué empezar a hacer antes de la próxima sesión?

Cursos del plan de aprendizaje de AWS Skill Builder
Developing Machine Learning Solutions
Developing Generative Artificial Intelligence Solutions
Optimizing Foundation Models

Plan de preparación para el examen (opcional)
Continuar – CloudQuest: Generative AI Practitioner; CloudQuest: Generative AI Architect*
Domain 3 Review; Domain 3 Practice*

* Requiere suscripción a AWS Skill Builder

Recursos

https://aws.amazon.com/what-is/foundation-models/

https://aws.amazon.com/what-is/transformers-in-artificial-intelligence/

https://aws.amazon.com/what-is/vector-databases/

https://docs.aws.amazon.com/bedrock/latest/userguide/inference-parameters.html

https://aws.amazon.com/ai/generative-ai/

https://docs.aws.amazon.com/sagemaker/latest/dg/jumpstart-foundation-models-customize.html

https://docs.aws.amazon.com/bedrock/latest/userguide/general-guidelines-for-bedrock-users.html#use-inference-parameters

https://aws.amazon.com/sagemaker/jumpstart/

https://aws.amazon.com/bedrock/pricing/

¡Gracias por asistir a esta sesión!