BIENVENIDOS

aws Cloud Clubs

aws Student Community Day

Medellín

aws Women Colombia User Group

aws Cloud Clubs IUDIGITAL

aws Cloud Clubs UD

aws Cloud Clubs UDEA

aws Cloud Clubs EAN

Cloud Clubs
aws Student
Community Day
Medellín

MY FIRST AI IN AWS

Jimena Ospina Vergara

QA Engineer
Source Meridian

Mauricio Ospina Vergara

DevOps Cloud Engineer
Rockwell Automation

AGENDA

1 OBJETIVO

2 GEN IA Y DEEPSEEK

3 ¿TECNICAS DE ENTRENAMIENTO?

4 ¿POR QUÉ AWS?

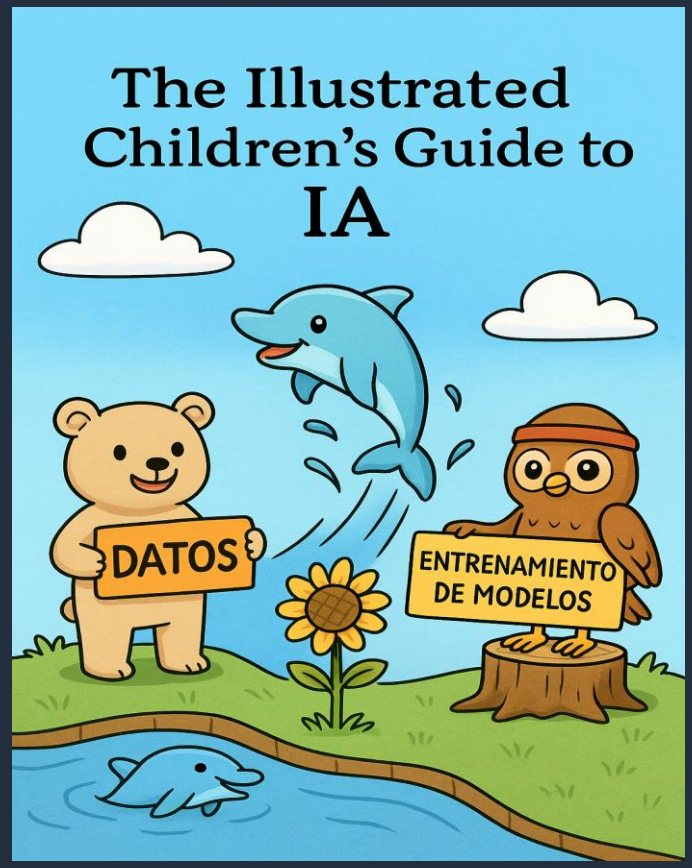
5 SAGE MAKER - DEMO

6 AMAZON BEDROCK - DEMO

aws Cloud Clubs aws Student Community Day Medellín

OBJETIVO

Que todos entiendan cómo usar **IA** en el contexto **empresarial**.



aws Cloud Clubs
aws Student Community Day
Medellín

¿GENAI?

Es un tipo de IA que puede crear nuevo contenido e ideas, incluyendo conversaciones, historias, imágenes, videos, etc.

¿FUNDATIONAL MODELS?

Son modelos de IA muy grandes y versátiles que sirven como base para múltiples aplicaciones.

TÉCNICAS DE ENTRENAMIENTO – FINE TUNING

Fine-tuning (IA Medicas, BioGPT)

Buscas **control total** sobre el comportamiento del modelo (sesgos, adaptación).

Necesitas un rendimiento altamente optimizado para **tareas específicas**.

Información **interna al modelo.**

IA CON TUS DATOS: FINE-TUNING

1. **Datos empresariales**
   Chats, correos, manuales, guias.
2. **Procesar y tokenizar**
   Se convierte la data en **secuencias** de **tokens** que el modelo puede entender y se guarda en un dataset.
3. **Entrenamiento**
   Con el dataset se hace **fine-tuning** al modelo base (DeepSeek por ejm). Allí se "**ajustan los pesos**".
4. **Custom Model**
   Se obtiene un **modelo personalizado** con las **respuestas heredadas** del dataset.
5. **Despliegue**
   Se despliega el **modelo** y está listo para usarse.

aws Cloud Clubs
aws Student Community Day
Medellín

TÉCNICAS DE ENTRENAMIENTO – RAG (GENERACIÓN AUMENTADA POR RECUPERACIÓN)

RAG (Copilot, NotebookLM)

Necesitas resultados rápidos y con bajo costo inicial.

Se les da un contexto y la **información es externa al modelo.**

IA CON TUS DATOS: RAG

1. **Datos empresariales**
   Documentos, PDF's, manuales.

2. **Embeddings**
   Dividir documentos en **chunks**, generar **embeddings** y **guardar** en bases de datos vectoriales.

3. **Conversion de input**
   El **usuario** realiza una **pregunta** que se convierte en **embedding**, esta se **busca** en la DB y trae los **documentos** mas **similares.**

4. **LLM**
   Se toma el texto de esos **documentos relevantes** y se **envia** como contexto al **modelo**.

5. **Respuesta**
   El modelo está listo para **responder**.

aws Cloud Clubs aws Student Community Day Medellín

¿POR QUÉ AWS?

AWS diferencia su propuesta GenAI con **arquitectura serverless multi-modelo** que ofrece acceso **plug-and-play** a modelos líderes (OpenAI, Anthropic, Meta, Amazon).

AWS proporciona **vendor-agnostic flexibility** con deployment en *15 minutos*, con pricing.

¿QUE PUEDO HACER EN AWS?

Entrenar **modelos**, hacer inferencias, automatizar procesos y personalizar soluciones.

https://aws.amazon.com/machine-learning/customers/

Cloud Clubs
aws Student
Community Day
Medellín

¿SAGE MAKER?

Es la **plataforma integral** de AWS para desarrollar y gestionar modelos de inteligencia artificial **personalizados**.

COSTOS SAGE MAKER

Sage Maker

Instance: ml.g5.12xlarge(4vCPU, 192RM, 4GPU)
Costo por hora: Aproximadamente $7.91 USD
720 horas × $7.91 USD/hora = $5,695.20 USD

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

aws Cloud Clubs aws Student Community Day Medellín

DESPLIEGUE EN SAGEMAKER CON HUGGING FACE

Configurar sesión AWS

Permisos y entorno para **acceder** a SageMaker

Definir modelo y endpoint

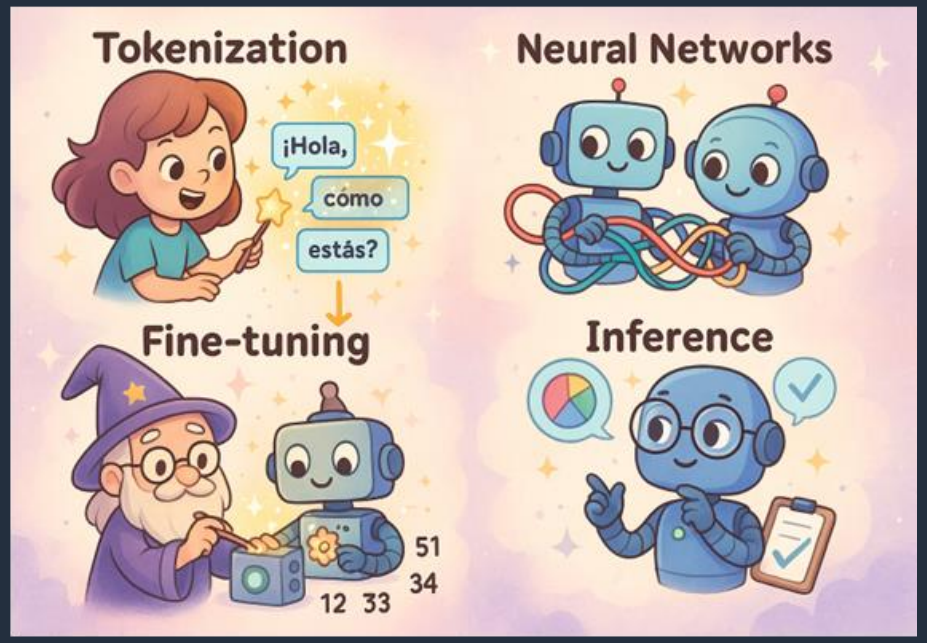
Usar instancia GPU para **despliegue** eficiente

Desplegar modelo

Preparar para recibir **consultas**

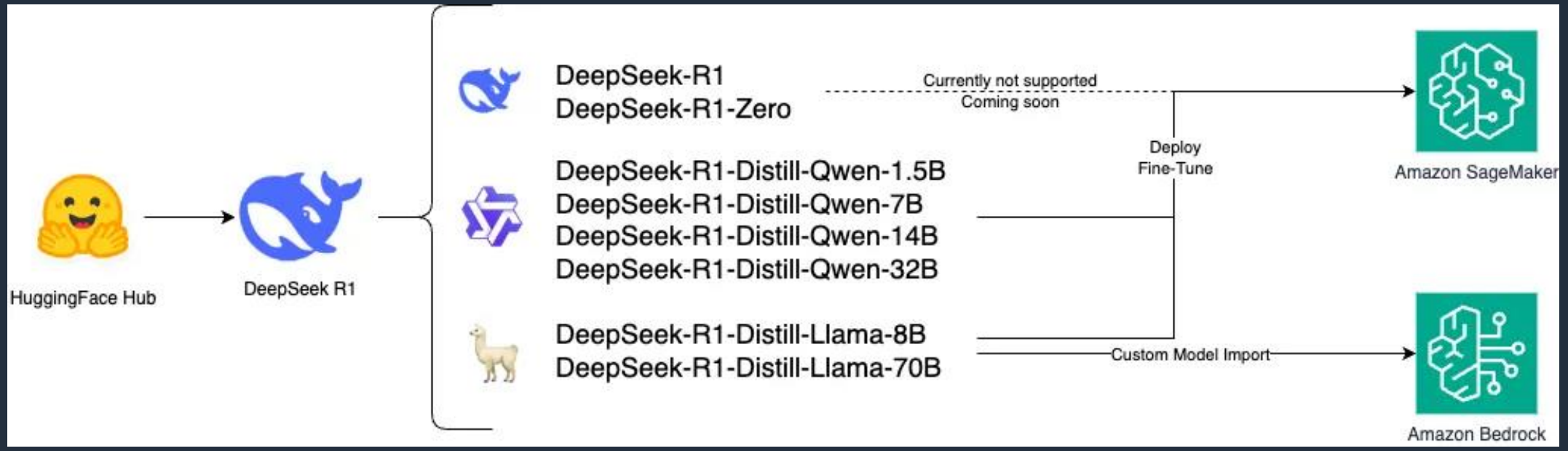
Realizar consulta

Obtener **respuestas** a preguntas reales



aws Cloud Clubs
aws Student Community Day
Medellín

HERRAMIENTAS AWS PARA IA



aws Cloud Clubs
aws Student Community Day
Medellín

¿AMAZON BEDROCK?

Es el servicio de AWS que te permite usar modelos de IA generativa **ya entrenados** de forma rápida y segura, sin necesidad de entrenar tus propios modelos.

COSTOS AMAZON BEDROCK

Amazon Bedrock

Modelo: DeepSeek-R1 1000 tokens de
1000 tokens de **Entrada**: $0.00135
1000 tokens de **Salida**: $0.0054

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

aws Cloud Clubs
aws Student Community Day
Medellín

KNOWLEDGE BASES EN AMAZON BEDROCK

- **Extracción automática** de datos desde S3 (PDFs, etc.).
- **Embeddings** con Titan.
- **Vector store** (OpenSearch o Aurora con pgvector).
- **Conexión RAG** con un modelo LLM (**DeepSeek**).
- Interfaz de preguntas lista para usar (consola o **API**).

Amazon Titan

MODELOS EN BEDROCK

aws Cloud Clubs
aws Student Community Day
Medellín

Thank you!