next EDUCACIÓN

GenAI Fundamentals

Docente: Isidre Munné-Bertran

De qué vamos a hablar

01

Introducción a la
IA, LLMs & **GenAI**

02

Sistemas **RAG**

03

Demo Chatbot

04

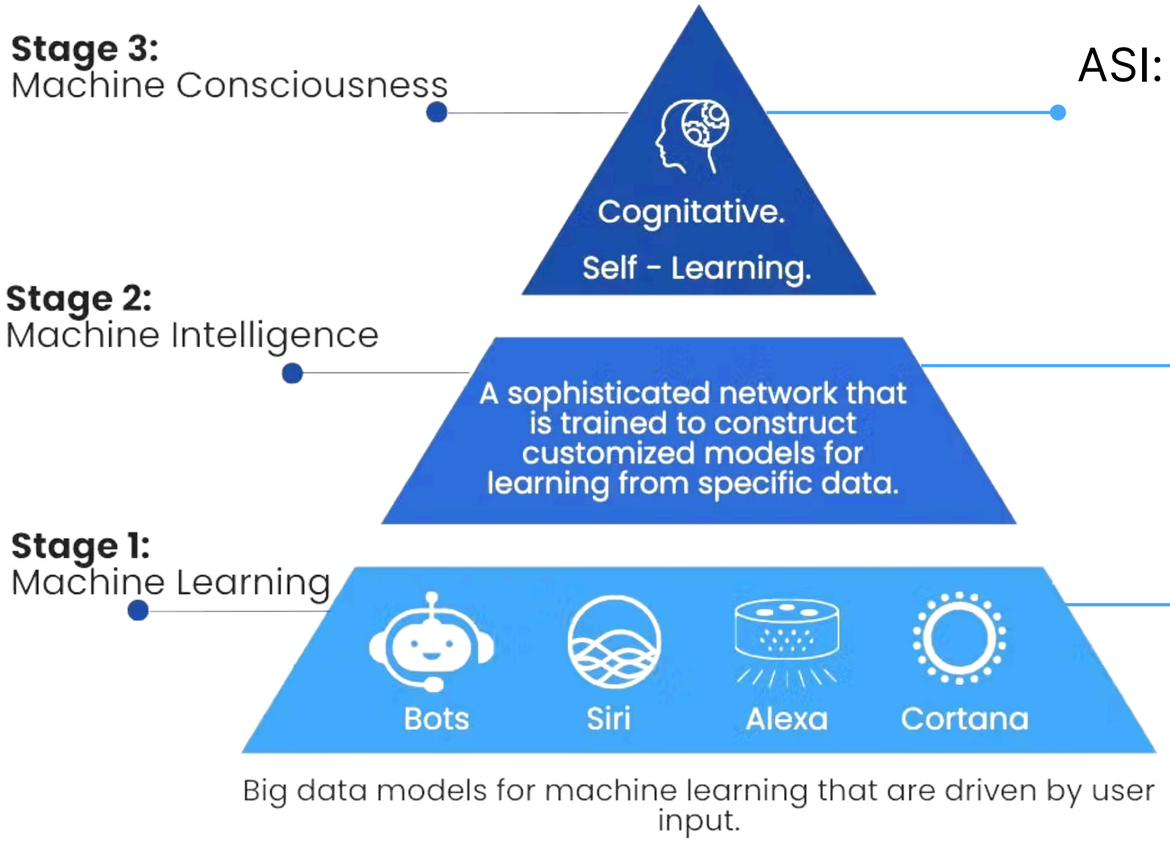
Requisitos

05

Hands-On

01 - Introducción a la **IA, LLMs** & **GenAI**

Tipos de IA

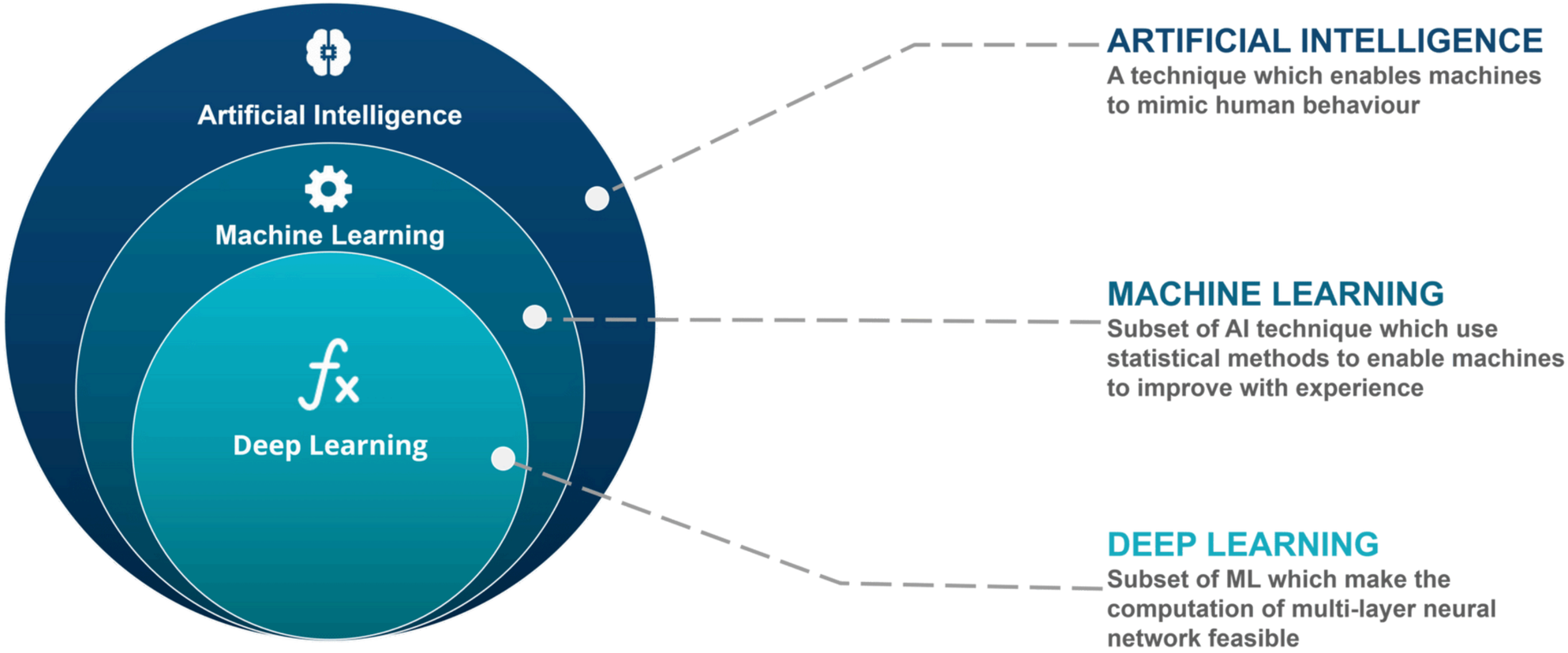


ASI: Artificial **Super** Intelligence

AGI: Artificial **General** Intelligence (**Strong AI**)

ANI: Artificial **Narrow** Intelligence (**Weak AI**)

¿Qué es la IA?

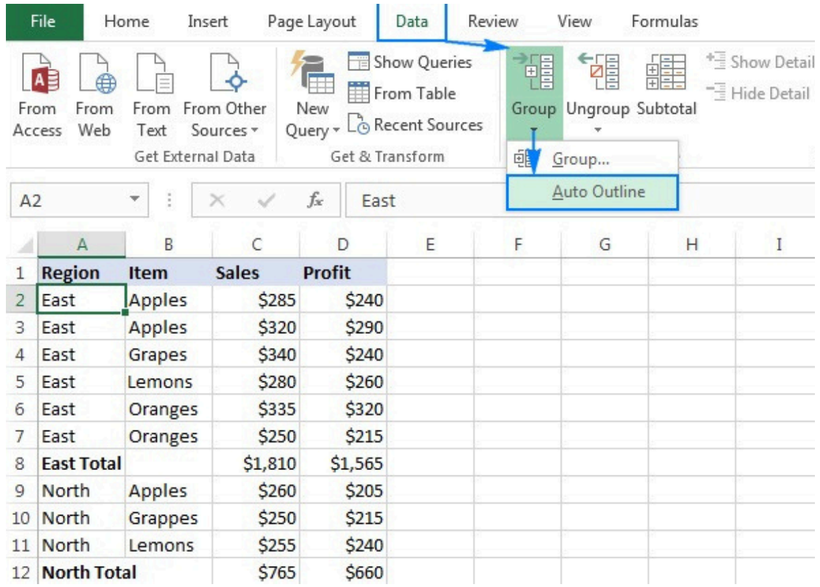


Aplicaciones: Robótica, GenAI, NLP, Computer Vision...

Tipos de Machine Learning

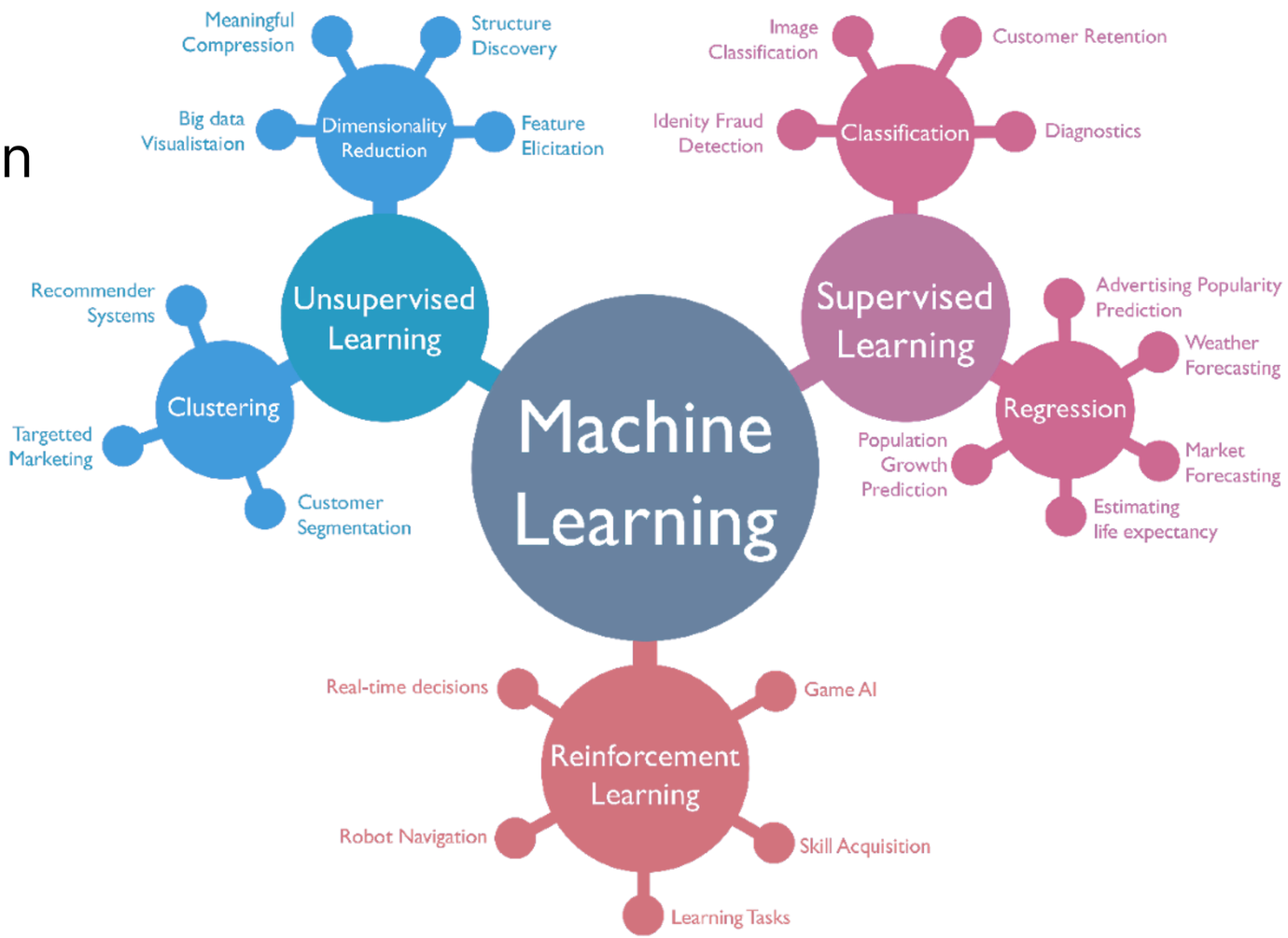
Aplicaciones:

- Detección de fraude
- Sistemas de recomendación
- Segmentación de clientes
- ...



	A	B	C	D
1	Region	Item	Sales	Profit
2	East	Apples	$285	$240
3	East	Apples	$320	$290
4	East	Grapes	$340	$240
5	East	Lemons	$280	$260
6	East	Oranges	$335	$320
7	East	Oranges	$250	$215
8	East Total		$1,810	$1,565
9	North	Apples	$260	$205
10	North	Grappes	$250	$215
11	North	Lemons	$255	$240
12	North Total		$765	$660

.csv, .json, .xlx, .xlsx, .xlms...

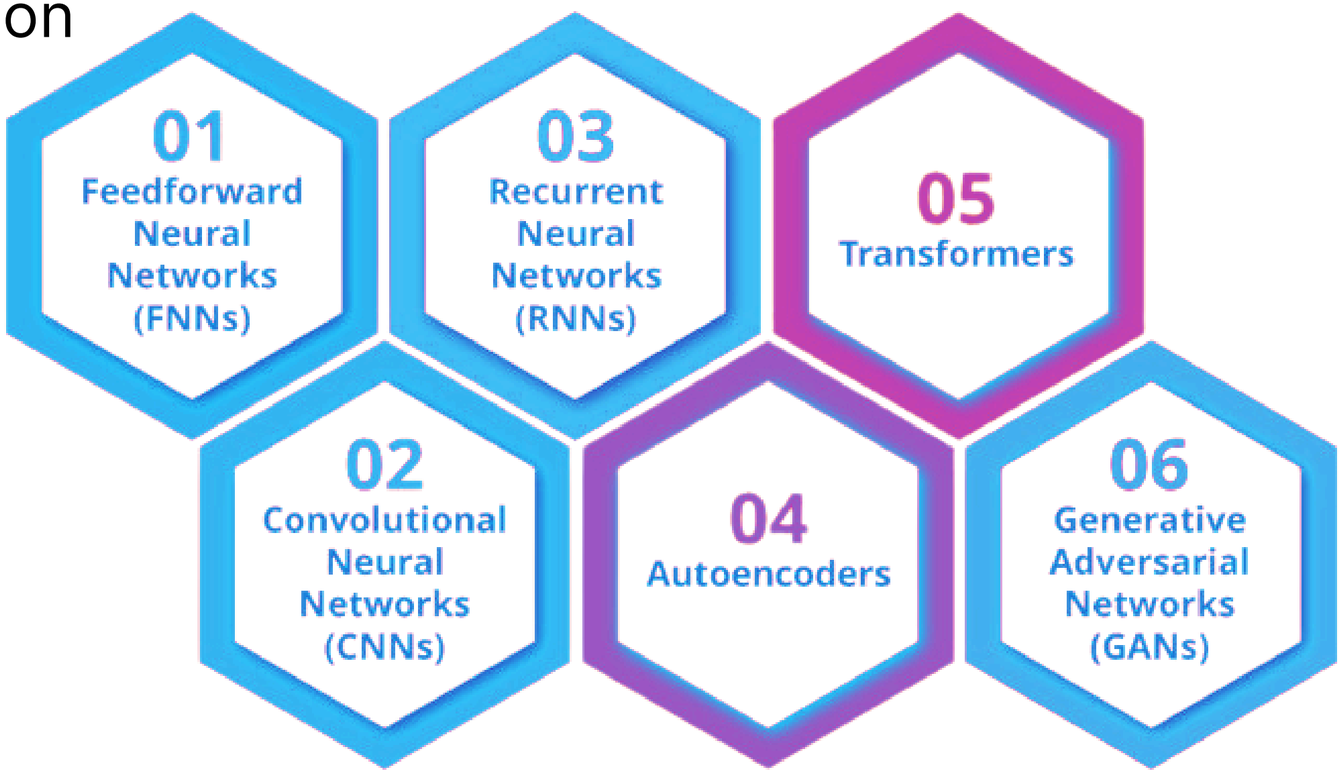


Tipos de Deep Learning

.jpg, .png, .tiff, .dicom, .svs, .wav, .mp3...

Aplicaciones:

- Computer Vision
- GenAI
- NLP
- ...

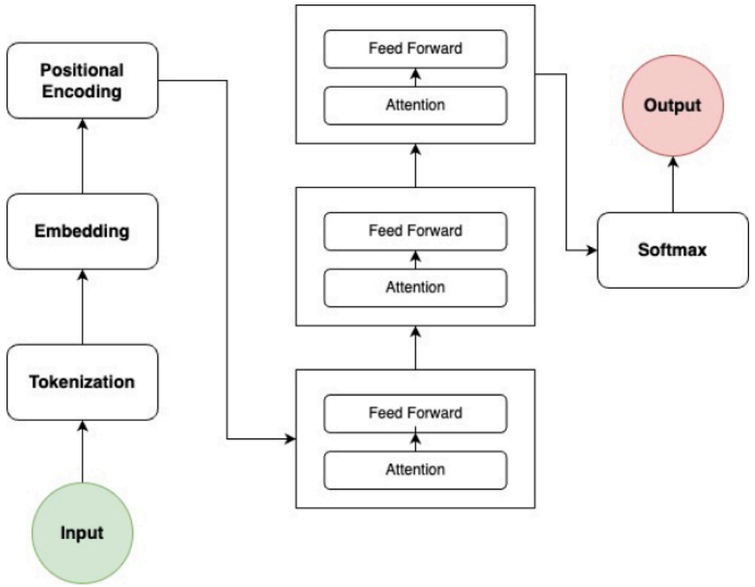


¿Qué son los LLM?

- Modelos de lenguaje **a gran escala** que generan texto basado en patrones de entrenamiento.
- Pertenecen a la **IA Débil** (Narrow/Weak AI) porqué **no tienen comprensión real** -> Son modelos probabilisticos.
- Se entrenan con enormes volúmenes de datos usando **redes neuronales profundas (Deep Learning)**
- **Ejemplos conocidos:** GPT-4, Claude, DeepSeek, Gemini, LLaMA

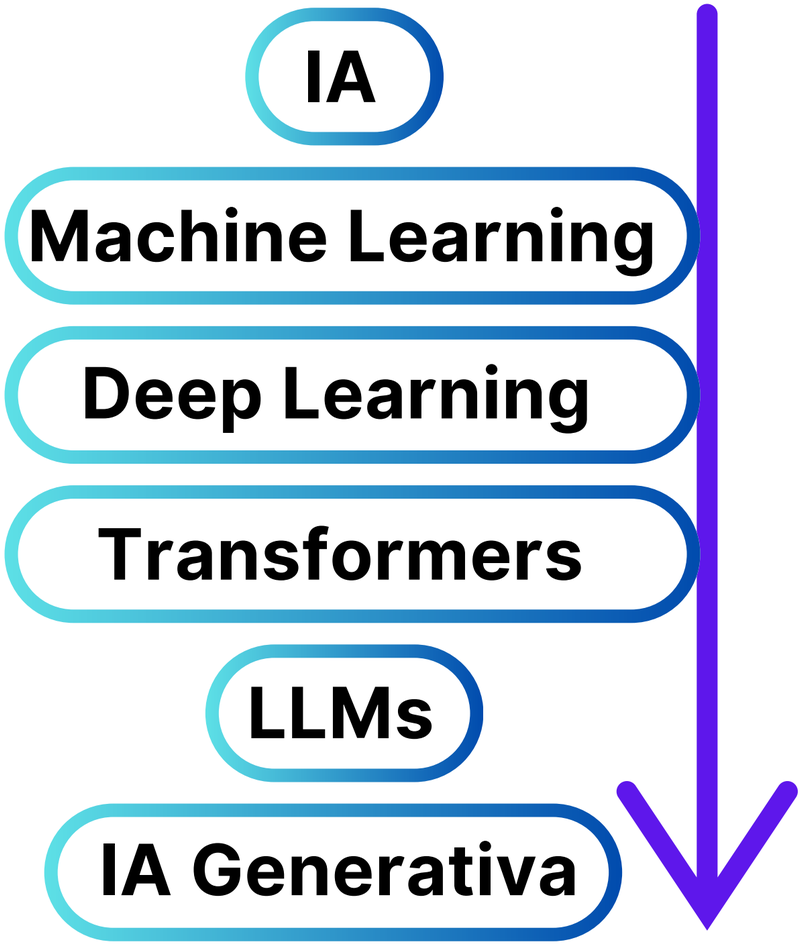
¿Qué son los Transformers?

- Una arquitectura de **red neural** que revolucionó NLP en 2017 =)
  - “Attention Is All You Need” (Google) -> https://research.google/pubs/attention-is-all-you-need/
- Utilizan **mecanismos de atención** para **priorizar** palabras en un texto.
- Procesamiento de texto en **paralelo** en comparación con los RNN (**secuencias**)
- **G**enerative... **P**re-trained...**T**ransformer



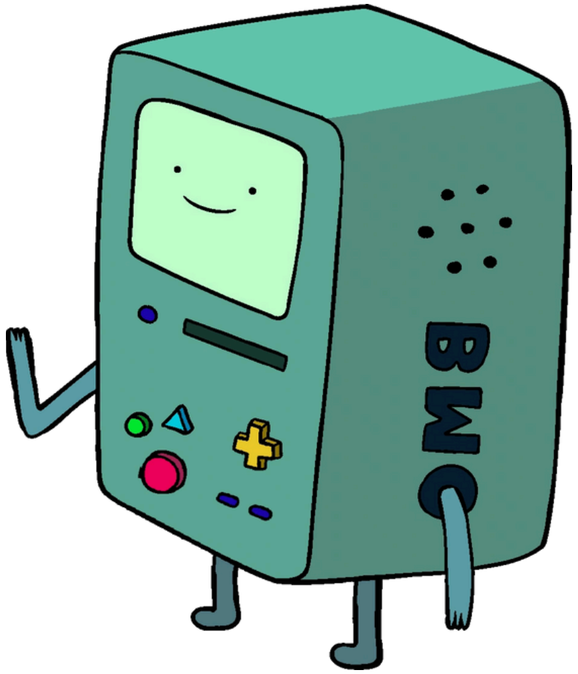
¿Qué es la IA Generativa (*GenAI*)?

- La **IA Generativa** no sólo analiza datos
  - Crea contenido con los patrones aprendidos.
- No está programada con respuestas fijas
  - Generaliza para generar **texto, imágenes, música, código, etc.**
- **Ejemplos:**
  - **Texto:** GPT-4, DeepSeek...
  - **Imágenes:** DALL·E, Midjourney...
  - **Código:** GitHub Copilot, Code Llama
  - **Música:** OpenAI Jukebox, AIVA...



AI+

- No es “solo” un LLM generando texto.
- Mejora los **LLM**/**GenAI** con técnicas avanzadas:
  - **RAG** (Retrieval-Augmented Generation)
  - Fine-Tuning
  - Memoria a largo plazo
  - Integración con bases de datos a tiempo real



Hoy: Chatbot con RAG/AI+

02 - Sistemas **RAG**

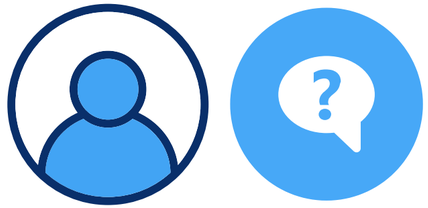
¿Qué es un sistema RAG?

- RAG **=** GenAI + Recuperación de información
- A diferencia de un LLM standard, **RAG** buscará información en fuentes externas para **mejorar** precisión y **reducir** **alucinaciones.**

- Estas **fuentes** pueden ser **documentos**, **búsquedas en Google, bases de datos**, etc.,

¿Cómo funciona?

Front-end: User query



Back-end

1- Document Retrieval

- Extraer fragmentos de una base de datos
- **Fuentes**: PDFs, bases de datos, APIs...

2- Embeddings & Vector Search

- Texto a **vectores númericos**
- **Bases de datos vectoriales:** ChromaDB, FAISS, Pinecone...

Front-end: Output

3- Generative AI

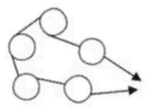
- Se inyecta la información recuperada en el **prompt** del LLM
- La IA genera una respuesta precisa y documentada.

¿Qué son los Embeddings?

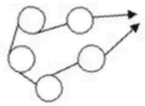
- Simplemente, convertir **palabras** o **frases** a **números.**
- Cuando el usuario haga una pregunta, se convertirá en un **embedding** y se comparará con los **embeddings** de la base de datos.
- Estos **números** se almacenarán como vectores (**listas de números**), representando las palabras en un espacio **multidimensional.**
- Se pueden **generar** con **APIs** como GPT, DeepSeek, Hugging Face...

Word	alive	mammal	rodent	bipedel	gender	plural
bunny	0.9	0.7	-0.1	-0.1	0.2	-0.3
rabbit	0.9	0.8	-0.1	-0.1	0.4	-0.1
hamster	0.9	0.6	0.7	-0.3	-0.4	-0.4
hutches	-0.9	-0.3	-0.1	-0.9	0.0	0.9
mother	0.9	0.1	0.1	0.2	0.8	-0.7
father	0.9	0.2	0.1	0.2	-0.8	-0.7
woman	0.9	0.8	-0.9	0.9	0.8	-0.8
man	0.9	0.8	-0.7	0.9	-0.8	-0.8

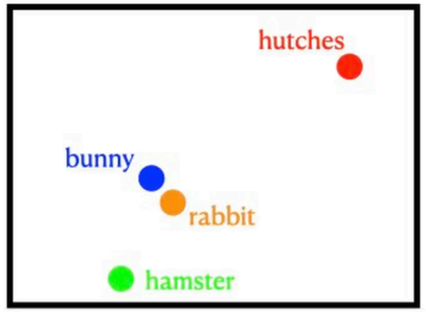
Vector Embedding

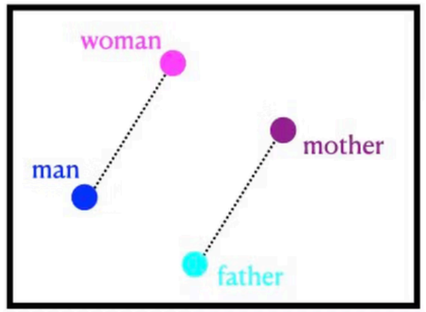


Dimensionality Reduction



Dimensionality





2D Visualization

¿Qué es ChromaDB?

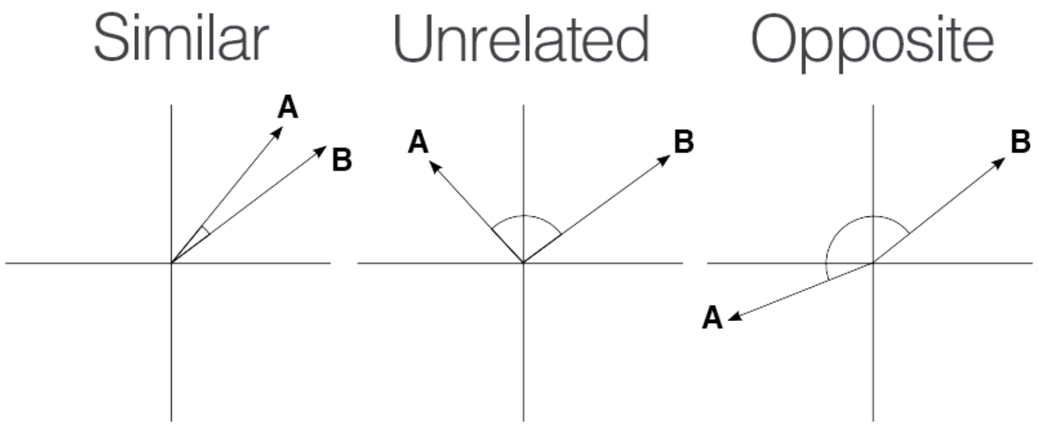
Una **base de datos** especialziada en búsqueda de texto mediante **IA**.

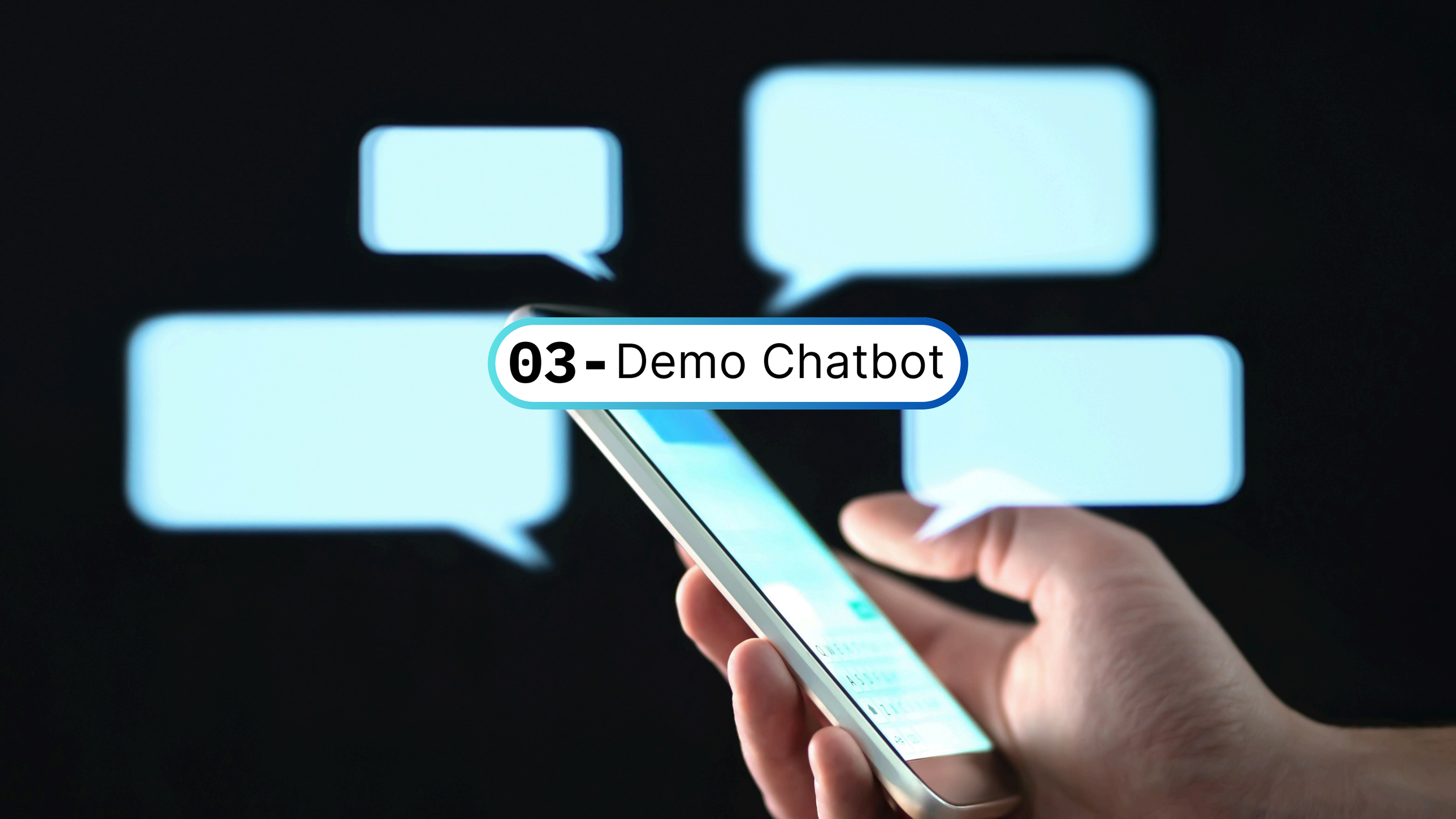
- Nos permitirá almacenar localmente (o en cloud) los **embeddings** de los PDF.
- Utilizando **similitud de cosenos** se comparará la pregunta del usuario con ChromaDB y seleccionará los fragmentos más relevantes.

¿Cómo funciona la similitud de cosenos?

- Cosine similarity **mide** la similitud entre **dos frases** o **palabras,** comparando los **ángulos** entre sus vectores en un espacio vectorial multidimensional euclidiano .

- **Simplificado**:
  - Si las flechas apuntan a la misma dirección -> son similares
  - Si las flechas están en ángulos diferentes -> son menos similares
  - Si son direcciones opuestas -> son completamente diferentes





03 - Demo Chatbot

APP

04 - Requisitos

Requisitos

- ¡Atreverse!
- **IDE**:
- **Lenguage**:
- **Librerias**

05 - Hands-On

next EDUCACIÓN