

LLMs de IA y Distancia Vectorial

La IA no piensa, mide.

AI LLMs & Vector Distance
AI doesn't think, it measures.

Comencemos con un juego de distancia

Adivina en Qué Estoy Pensando

- Estoy pensando en un número del 1-100
- Obtienes pistas: "Más alto" o "Más bajo"

¡La IA hace lo mismo!

Todos usan: VECTORES Y DISTANCIA

- ChatGPT responde tus preguntas
- TikTok elige videos solo para ti
- Google Fotos encuentra tus fotos de perros
- 🔎 Búsqueda Google: Encuentra artículos similares
- 📽 Netflix: Recomienda shows como los que viste
- 🎵 Spotify: Encuentra canciones con la misma vibra
- 📱 Instagram: Muestra posts similares
- 🌎 Google Translate: Empareja palabras entre idiomas

¿Qué es un vector?

Un Vector es Solo una Lista de Números



¿Cuál es la distancia entre Colegio del Sol y Playa Bonita?

$$\text{Distancia} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \dots}$$

Playa Bonita: [0, 7.7]

Colegio del Sol: [1.5, 4.2]

$$x_1 = 0$$

$$y_1 = 7.7$$

$$x_2 = 1.5$$

$$y_2 = 4.2$$

$$\begin{aligned}\text{Distancia} &= \sqrt{(0-1.5)^2 + (7.7-4.2)^2} \\&= \sqrt{(-1.5)^2 + (3.5)^2} \\&= \sqrt{(-1.5 \times -1.5) + (3.5 \times 3.5)} \\&= \sqrt{2.25 + 12.25} = \sqrt{14.5} \\&= 3.8 \text{ kilometers}\end{aligned}$$

Crea un vector que te represente con números del 0 al 10.

1. ¿Tienes la altura? 0: la más baja de la clase; 10: la más alta.
3. ¿Tienes una familia numerosa? 1: no tienes hermanos; 10: tienes muchos hermanos, tíos y primos.
5. ¿Te gusta el sushi? 0: no; 5: más o menos; 10: sí.
7. ¿Te gusta jugar al fútbol? 0: nada; 10: muchísimo.

Kristin = [w, x, y, z] = [9, 3, 10, 6]

**Crea un grupo con 2 a 3 personas que
tengan un vector similar al tuyo.**

Calcula la distancia entre todos los miembros de tu grupo utilizando la ecuación de distancia::

$$D = \sqrt{[(w_1-w_2)^2 + (x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2]}$$

No es necesario calcular la raíz cuadrada, como $D = \sqrt{45}$

¿Quién es más similar?

Toda la IA/LLM Moderna Usa la Misma Idea

All Modern AI/LLM Uses the Same Idea

LLM = Large Language Model

- Convierte cosas en vectores → Mide distancia → Encuentra la coincidencia más cercana
Turn things into vectors → Measure distance → Find the closest match
- La Distancia Mide Similitud
Distance Measures Similarity
 - Distancia PEQUEÑA = vectores SIMILARES
 - Distancia GRANDE = vectores DIFERENTES
 - ¡Piensa en ello como medir qué tan lejos están dos cosas en el espacio!
Think of it like measuring how far apart two things are in space!

Cómo funcionan realmente los LLM

How LLMs really work

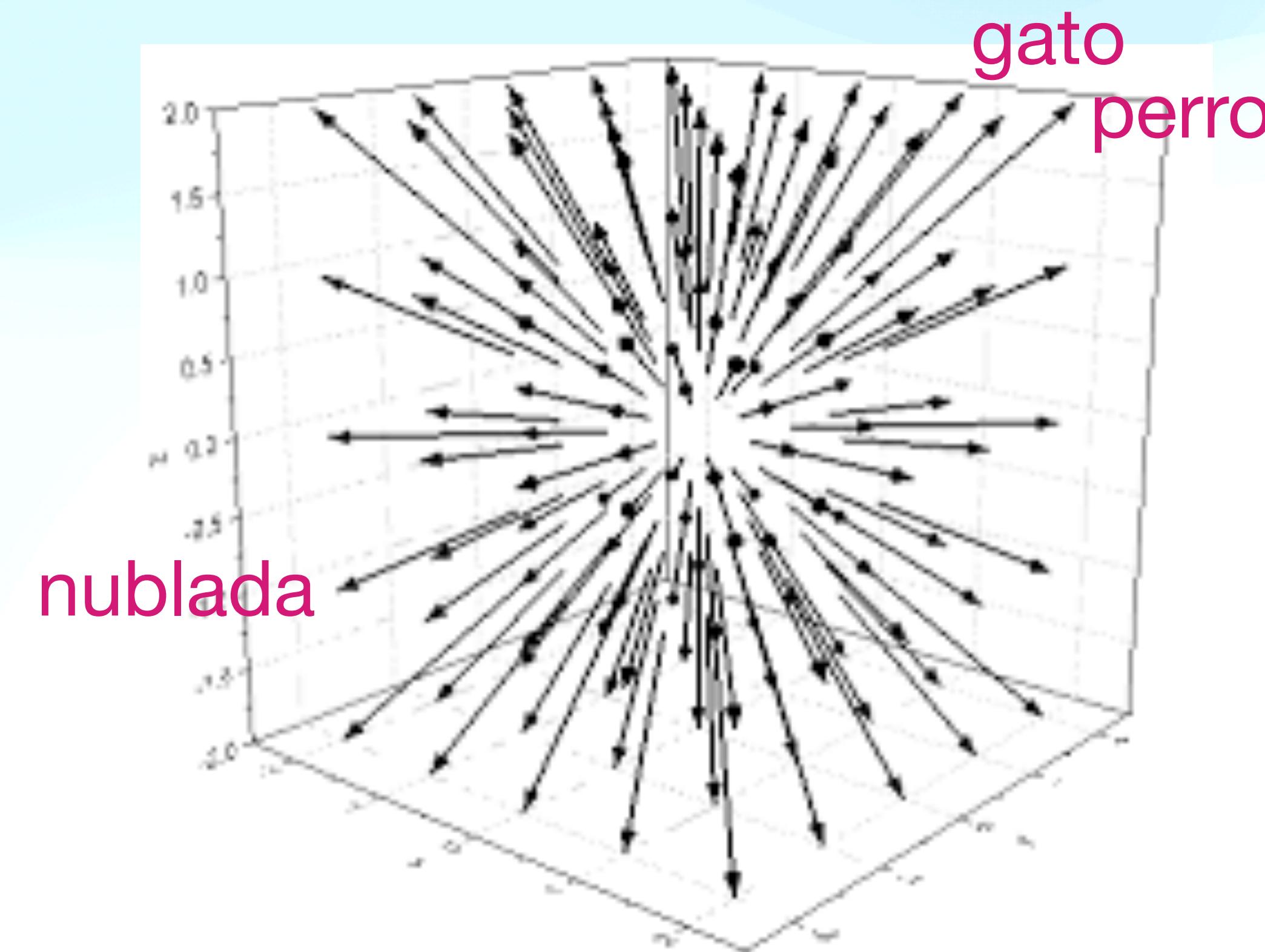
3 Pasos:

1. ENTRADA **input**: "¿Cuál es la capital de Francia?" → Convertir a vector
2. BÚSQUEDA **search**: Encontrar vectores más cercanos en memoria
3. SALIDA **output**: Generar "París"

La IA no "sabe" que París es la capital. ¡Solo encuentra el vector más SIMILAR!

¿Cuántos vectores hay en ChatGPT?

Probablemente **cientos de miles de millones de parámetros**, ya que contiene todo internet.



El problema del sesgo (bias)

Ejemplo Real

Los datos de entrenamiento contienen más fotos de hombres realizando cirugías.

The training data contains more photos of men performing surgeries.

Entonces el vector "cirujano" se acerca más al vector "hombre".

Resultado: La IA muestra principalmente hombres para puestos de cirujano.

Result: The AI primarily shows men performing surgery.

¡Esto es injusto y debe corregirse!

This is unfair and must be corrected!



El problema de las alucinaciones (hallucinating)

Ejemplo Real

Alucinar significa que la IA da una respuesta que suena correcta pero que en realidad es inventada.

Hallucinating means the AI gives an answer that sounds correct but is actually made up.

Preguntas: "¿Quién inventó la pizza arcoíris en 1850?" Pero no existía tal cosa como una "pizza arcoíris" en 1850.

You ask: "Who invented the rainbow pizza in 1850?" But, there is no such thing as a "rainbow pizza" in 1850.

Pero la IA está entrenada para dar siempre una respuesta, incluso cuando la pregunta no tiene una respuesta real. Así, la IA podría decir: «La pizza arcoíris fue inventada en 1850 por Giovanni Bellini». Pero eso no es cierto; la IA está rellenando los huecos.

But the AI is trained to always give an answer, even when the question has no real answer. So the AI might say: "The rainbow pizza was invented in 1850 by Giovanni Bellini." But that's not true — the AI is filling in blanks

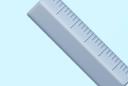
¡Puedes Mejorar la IA! You can improve AI!

-  Aprende matemáticas y programación
-  Haz preguntas difíciles sobre sesgos
-  Piensa en IA ética
-  Learn math and programming
-  Ask tough questions about bias
-  Think about ethical AI

Creciste con IA. Entiendes cómo afecta tu mundo. ¡Usa ese poder responsablemente!

You are growing up with AI. You understand how it affects your world. Use that power responsibly!

Lo que aprendiste hoy

-  Vectores = listas de números que describen cosas
-  Distancia = qué tan similares son dos vectores
-  Los LLMs (IA) encuentran el vector más cercano para generar respuestas
-  Esto impulsa ChatGPT, Spotify, TikTok, Google
-  La IA puede ser sesgada si los datos de entrenamiento lo son
-  Vectors = lists of numbers that describe things
-  Distance = how similar two vectors are
-  AI LLMs find the closest vector to generate answers
-  This powers ChatGPT, Spotify, TikTok, Google, Instagram
-  AI hallucinates and can be biased if the training data is biased