

# Generando imágenes ultrarealistas con IA



Un astronauta montando a caballo en una imagen fotorrealista

Jorge Gómez  
30/06/2022

## Inteligencia artificial

¿Qué es la inteligencia artificial?

¿En qué se diferencia de la inteligencia humana?

¿Es inteligente un ordenador actual?

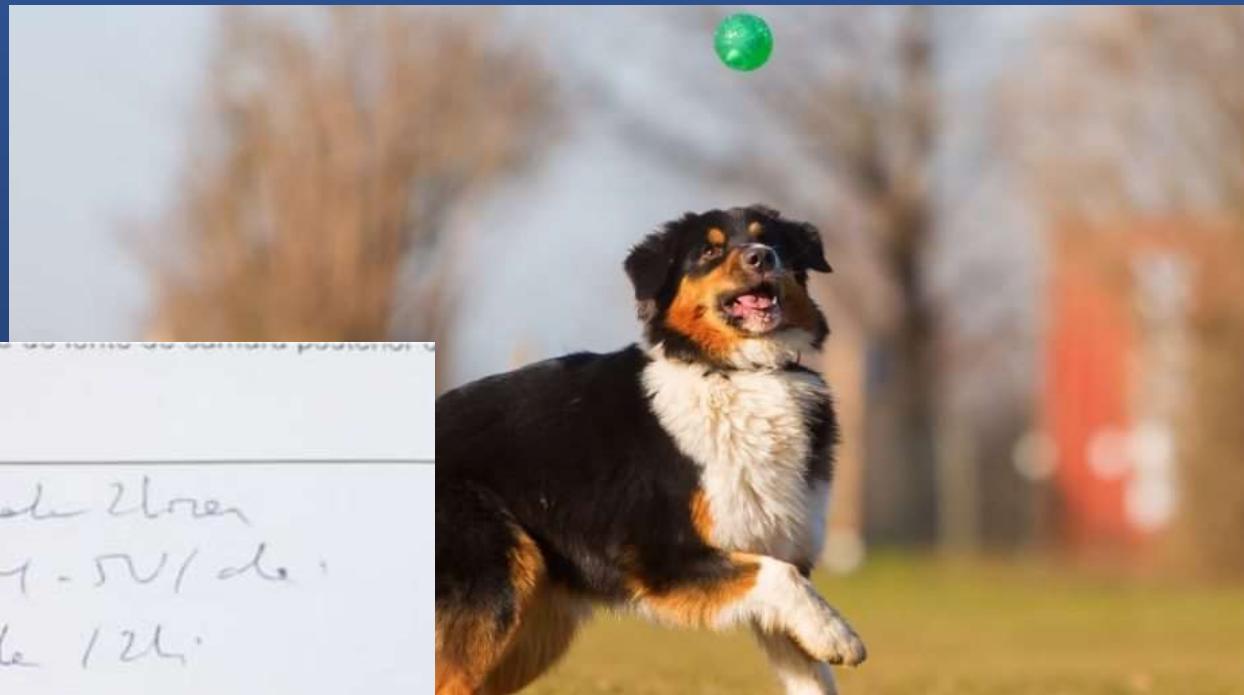
¿Estamos en un punto de ruptura?

¿De donde venimos en inteligencia artificial y hacia dónde nos dirigimos?

## Tipos de inteligencia (según Howard Gardner)



## Tipos de inteligencia (según Howard Gardner)

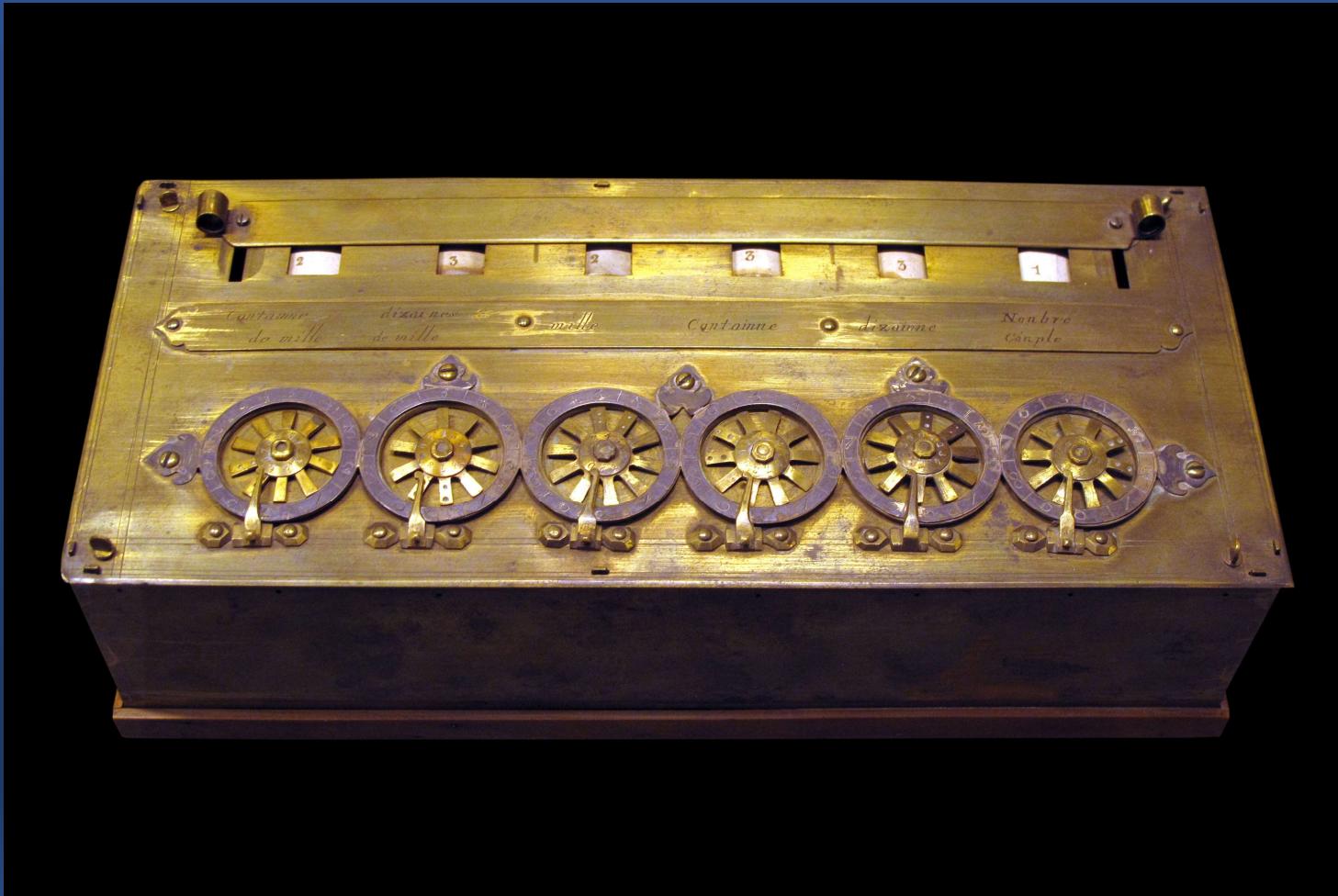


### TRATAMIENTO

- Colirio Tobradex 5 veces al día. *cada 2 horas*
- Crema de pel 4-5U/día
- Glu 1 libra cada 12h
- Edens cada hora 1/2

Revisión: Mañana en la 2<sup>a</sup> planta del Hospital Juan Cardona a las 10:00

# IA logico-matemática



Calculadora de Pascal (1652)

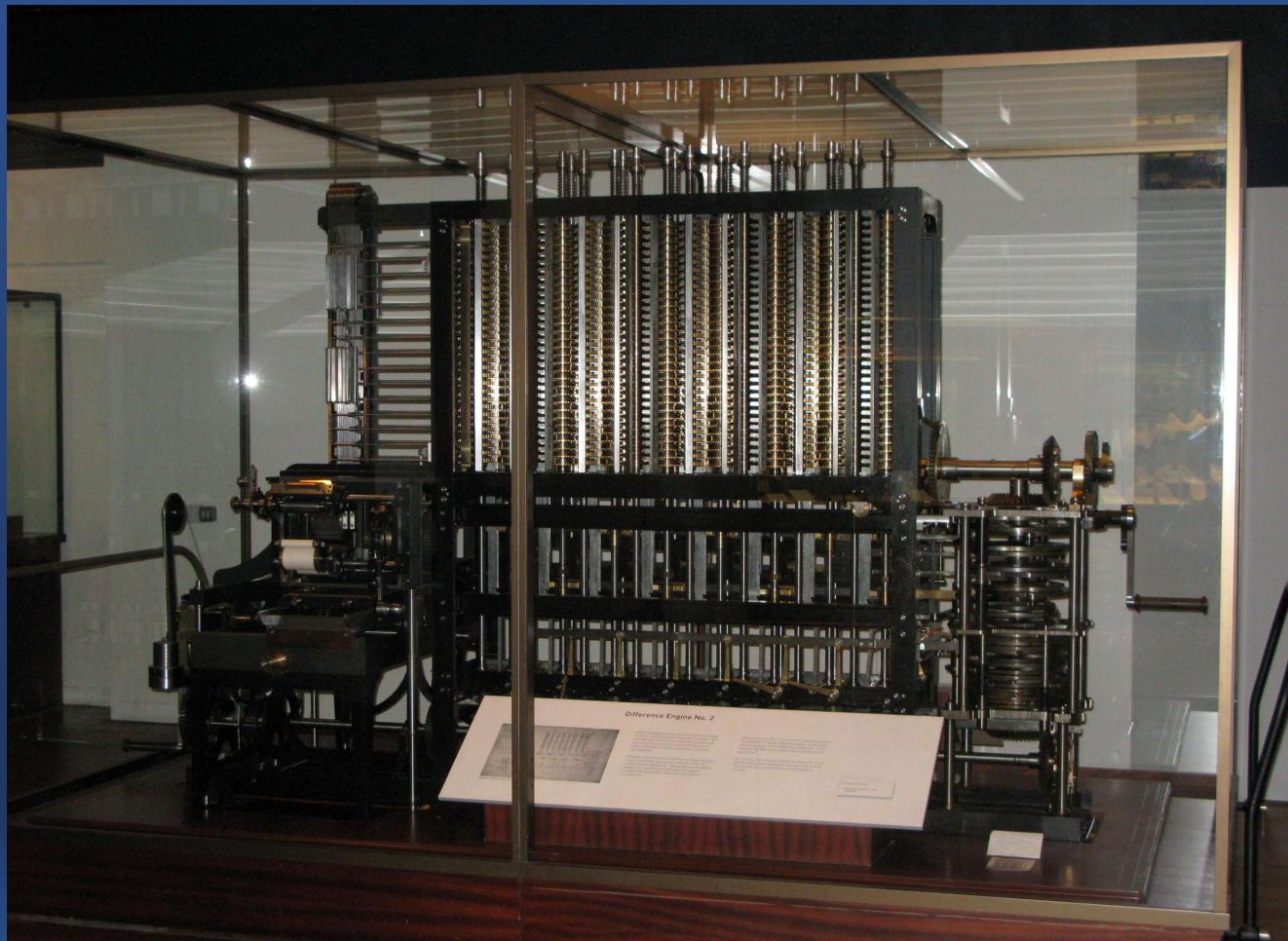
A large blue circle represents Pascal's calculator. Inside it, a smaller orange circle represents the central mechanism. Two orange arrows point from the text labels to the corresponding parts of the calculator.

Calculadora de Pascal (1652)

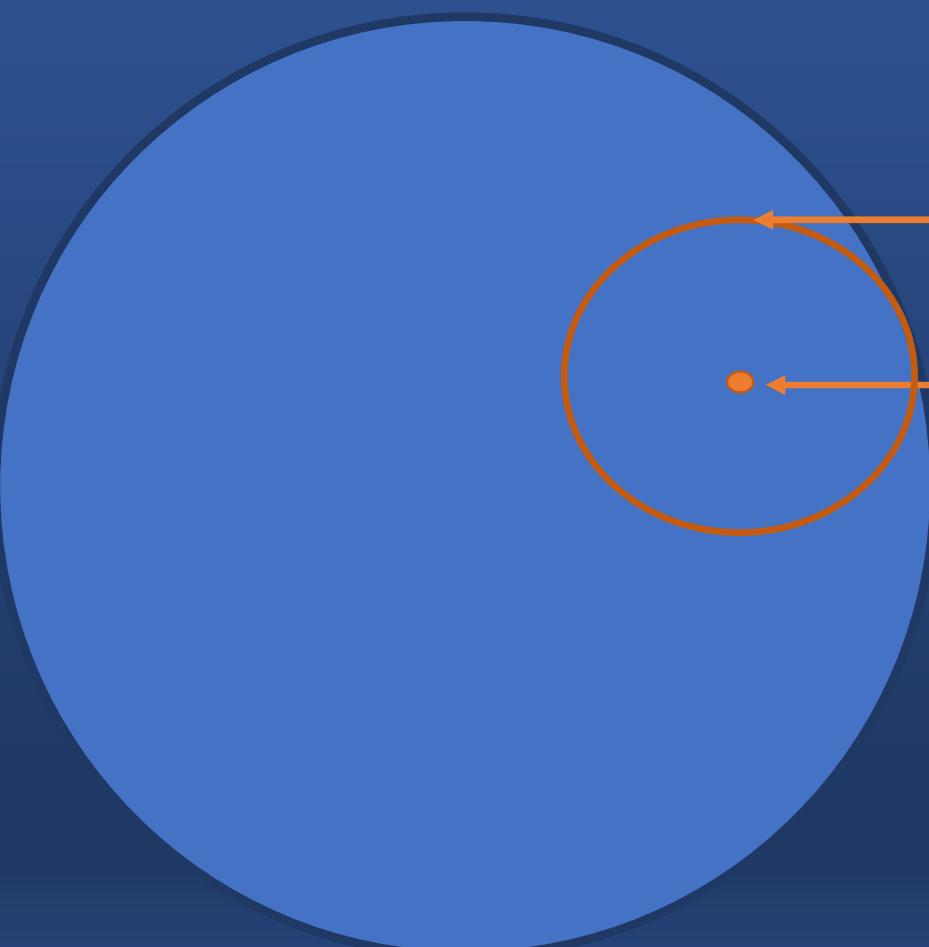
Inteligencia lógico-matemática

Capacidad de sumar y restar

# IA logico-matemática



Máquina diferencial (1822)



A diagram illustrating the components of Charles Babbage's Difference Engine No. 1. It features a large blue circle representing the machine itself, set against a dark blue background. Inside the blue circle is a smaller orange circle, which contains a small orange dot. Two orange arrows point from the text labels to these elements: one arrow points from the label "Inteligencia lógico-matemática" to the outer blue circle, and another arrow points from the label "Capacidad de sumar y restar, calcular logaritmos, funciones trigonométricas, polinomios, interpolar,..." to the inner orange circle.

Máquina diferencial (1822)

Inteligencia lógico-matemática

Capacidad de sumar y  
restar, calcular logaritmos,  
funciones trigonométricas,  
polinomios, interpolar,...

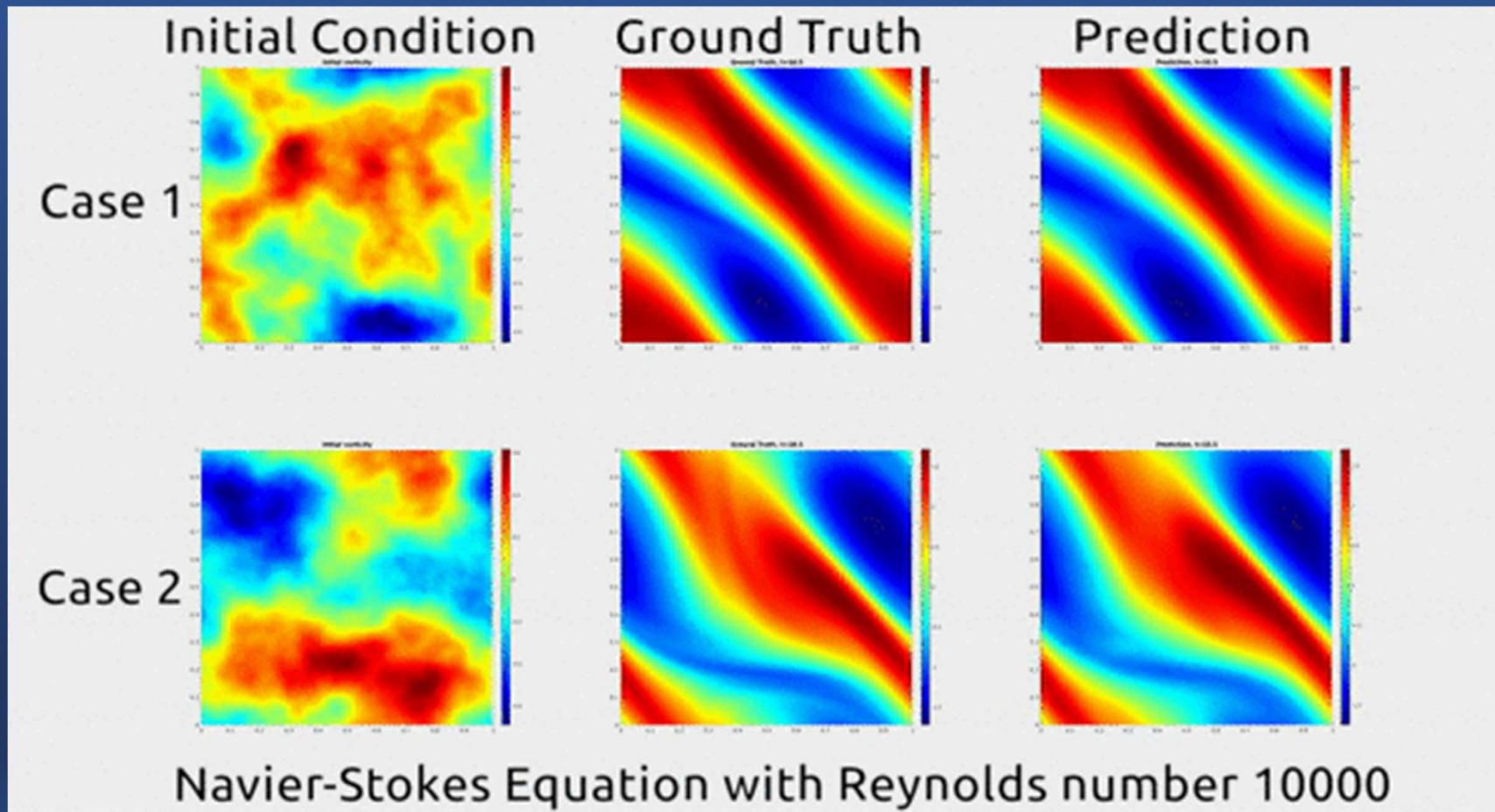
# IA logico-matemática



Computador actual



¿Y con deep learning?

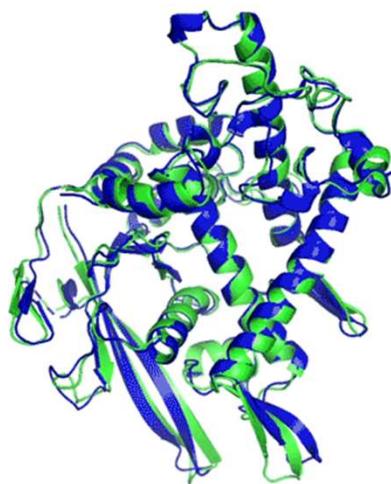


## ¿Y con deep learning?

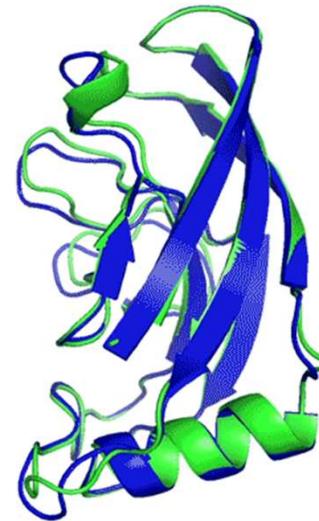
Equation	Solution
$y' = \frac{16x^3 - 42x^2 + 2x}{(-16x^8 + 112x^7 - 204x^6 + 28x^5 - x^4 + 1)^{1/2}}$	$y = \sin^{-1}(4x^4 - 14x^3 + x^2)$
$3xy \cos(x) - \sqrt{9x^2 \sin(x)^2 + 1}y' + 3y \sin(x) = 0$	$y = c \exp(\sinh^{-1}(3x \sin(x)))$
$4x^4yy'' - 8x^4y'^2 - 8x^3yy' - 3x^3y'' - 8x^2y^2 - 6x^2y' - 3x^2y'' - 9xy' - 3y = 0$	$y = \frac{c_1 + 3x + 3 \log(x)}{x(c_2 + 4x)}$

Table 4: Examples of problems that our model is able to solve, on which Mathematica and Matlab were not able to find a solution. For each equation, our model finds a valid solution with greedy decoding.

## ¿Y con deep learning?



**T1037 / 6vr4**  
90.7 GDT  
(RNA polymerase domain)

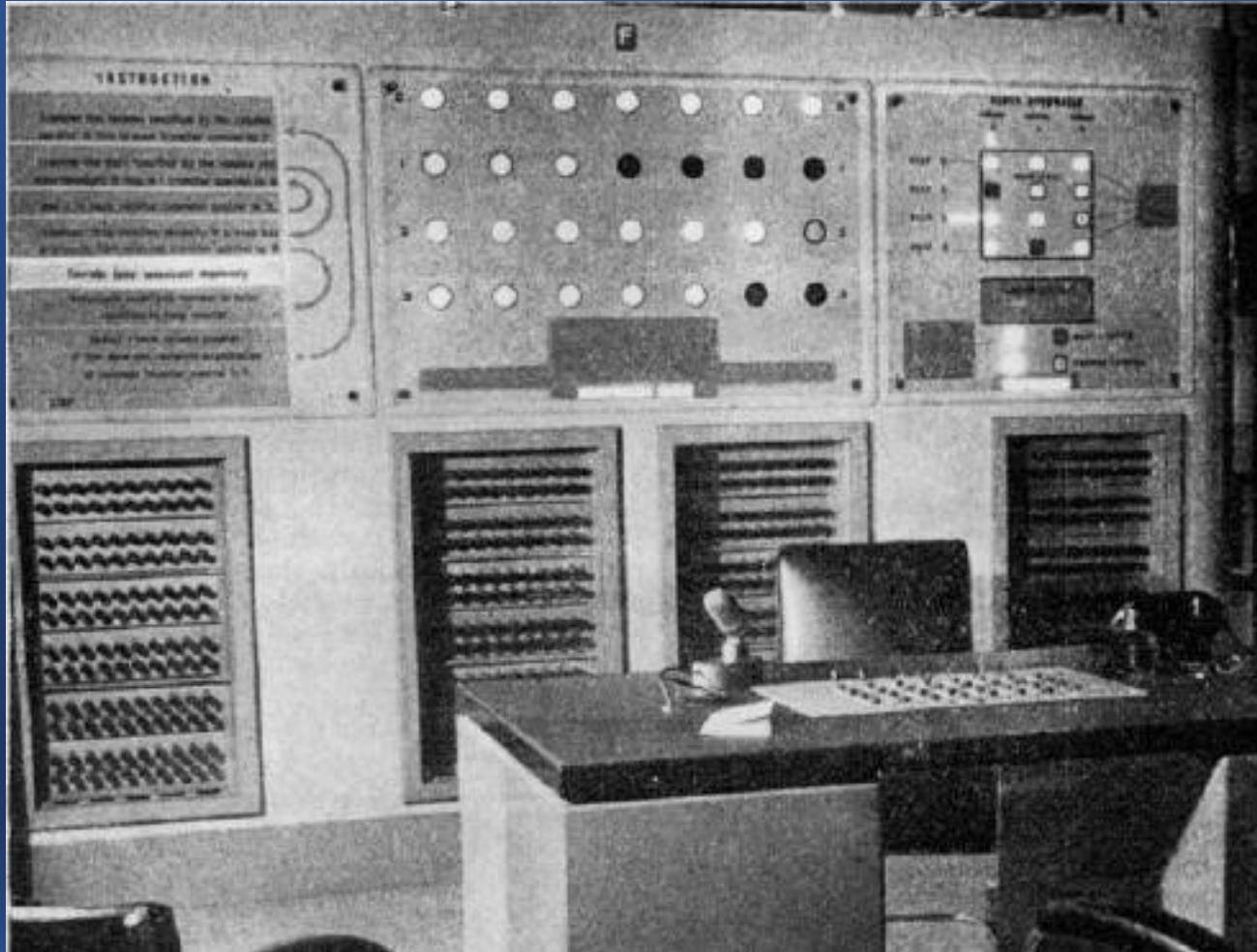


**T1049 / 6y4f**  
93.3 GDT  
(adhesin tip)

- Experimental result
- Computational prediction

AlphaFold (2021)

## IA logico-matemática



Nimrod (1951)



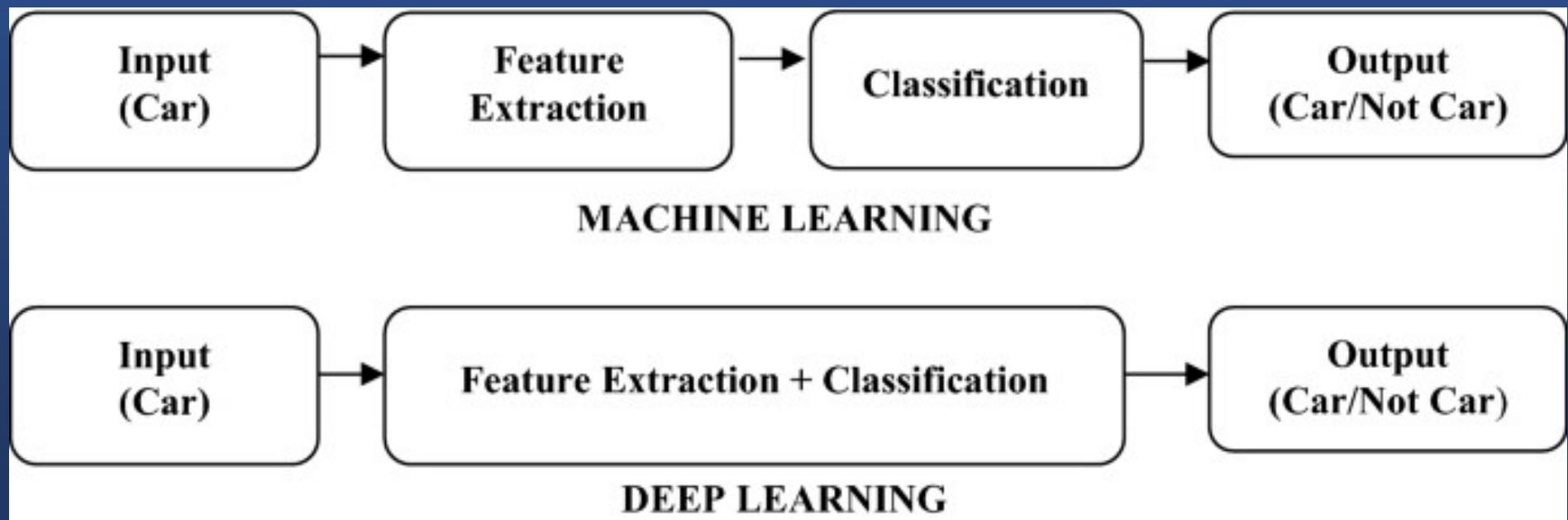
## IA logico-matemática



Deep Blue (1996)



## ¿Y con deep learning?

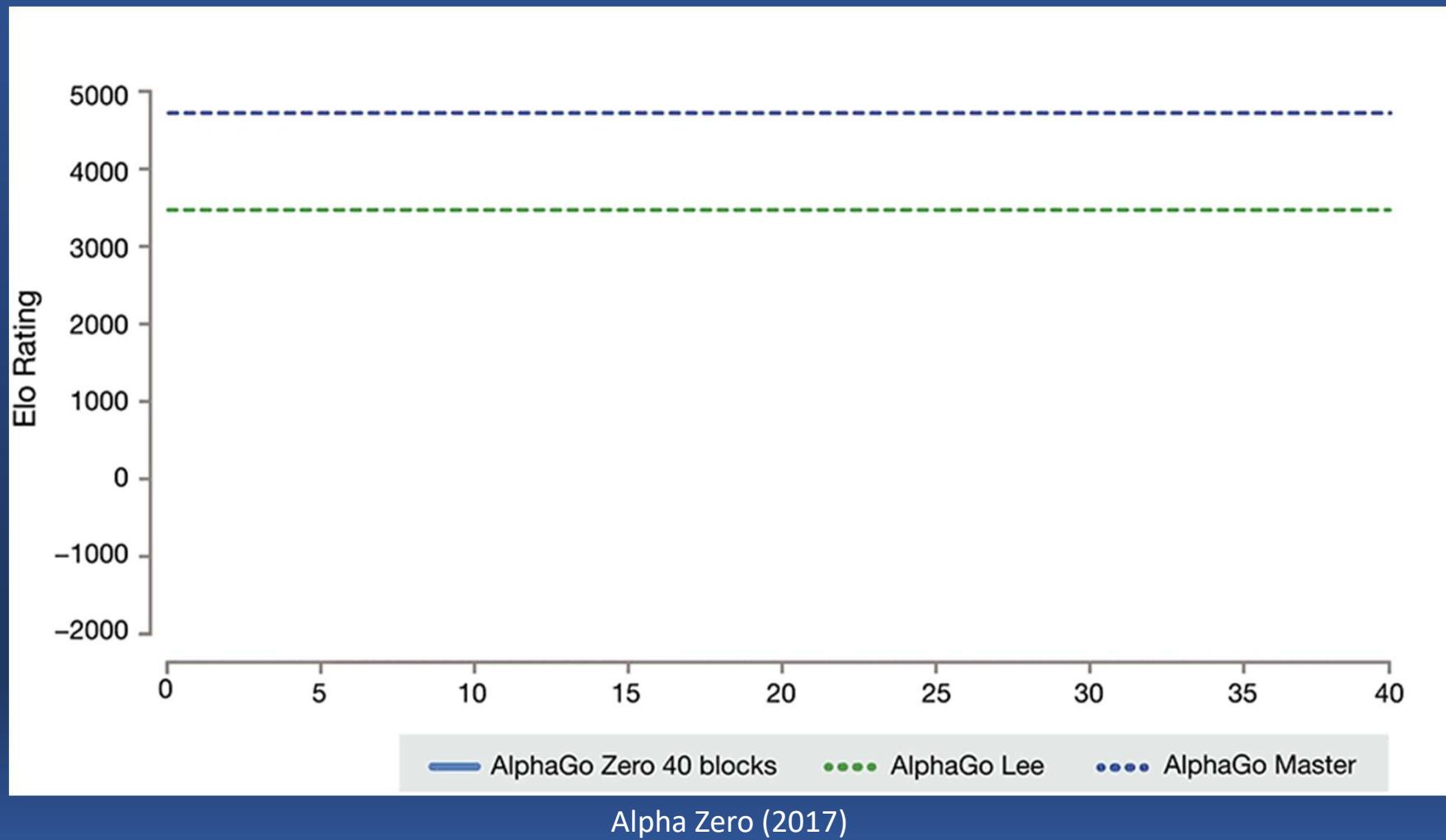


## ¿Y con deep learning?



Alpha Go (2016)

## ¿Y con deep learning?

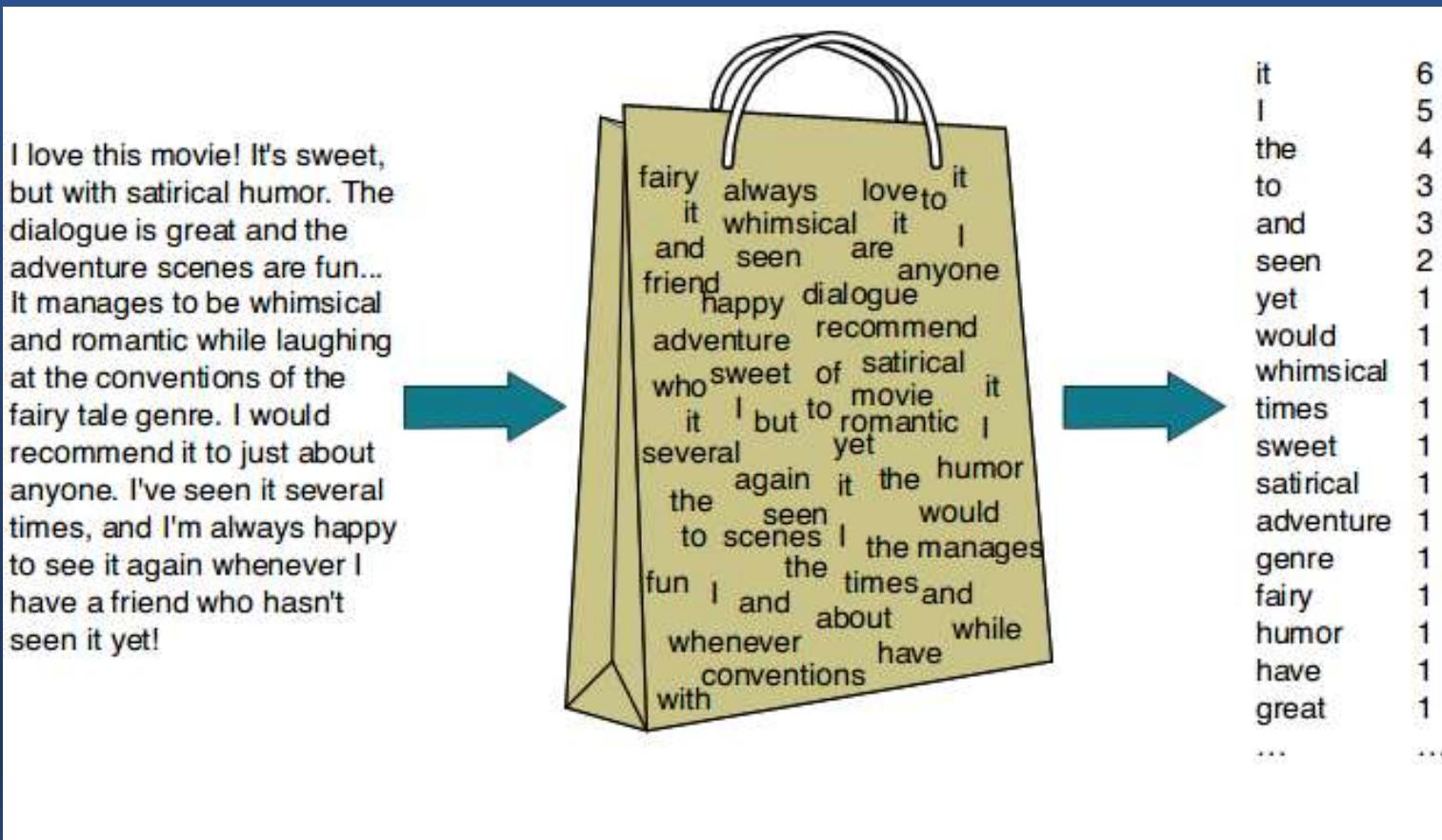




Inteligencia lógico-matemática

Jugar a cualquier  
juego de estrategia  
mejor que un humano

## IA lingüística



Bag of words (1954)

# IA lingüísitca

Welcome to

EEEEEE	LL	IIII	ZZZZZZ	AAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LLLLLL	IIII	ZZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.

The original program was described by Joseph Weizenbaum in 1966.  
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?

YOU: Men are all alike.

ELIZA: What is the connection, do you suppose ?

YOU: They're always bugging us about something or other.

ELIZA: Can you think of a specific example ?

YOU: Well, my boyfriend made me come here.

ELIZA: Is it important to you that your boyfriend made you come here ?

YOU: He says I'm depressed much of the time.

ELIZA: I am sorry to hear that you are depressed.

YOU: It's true. I am unhappy.

ELIZA: Can you explain what made you unhappy ?

YOU:

Eliza (1966)

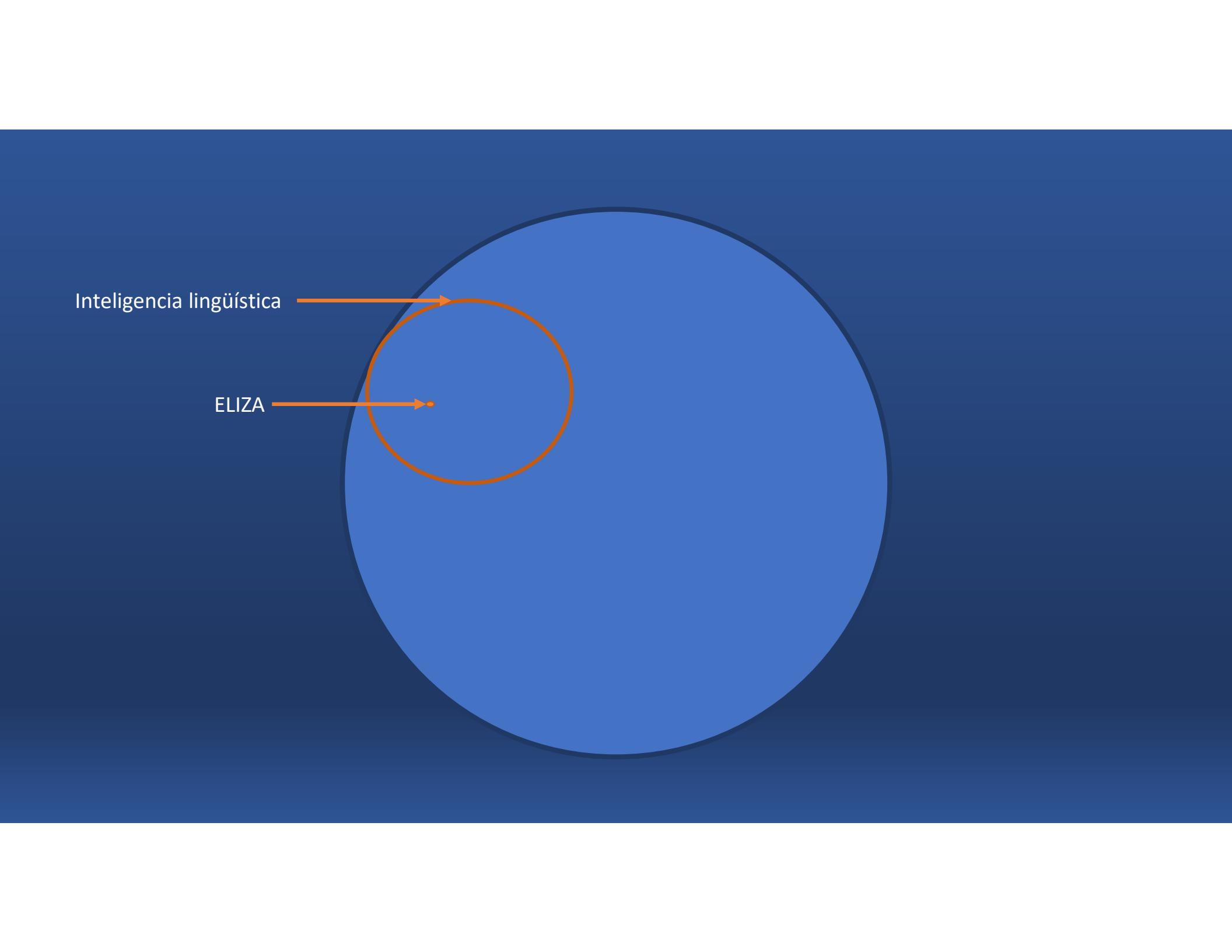
## IA lingüística

Efecto ELIZA: tendencia a asumir, inconscientemente, que los comportamientos informáticos son análogos a los comportamientos humanos

Ejemplo: si compramos unas entradas de cine en una máquina automática y la máquina escribe en la pantalla “gracias por su compra” no está mostrando gratitud

En el caso de ELIZA, los usuarios atribuían comprensión y motivación en las respuestas del programa

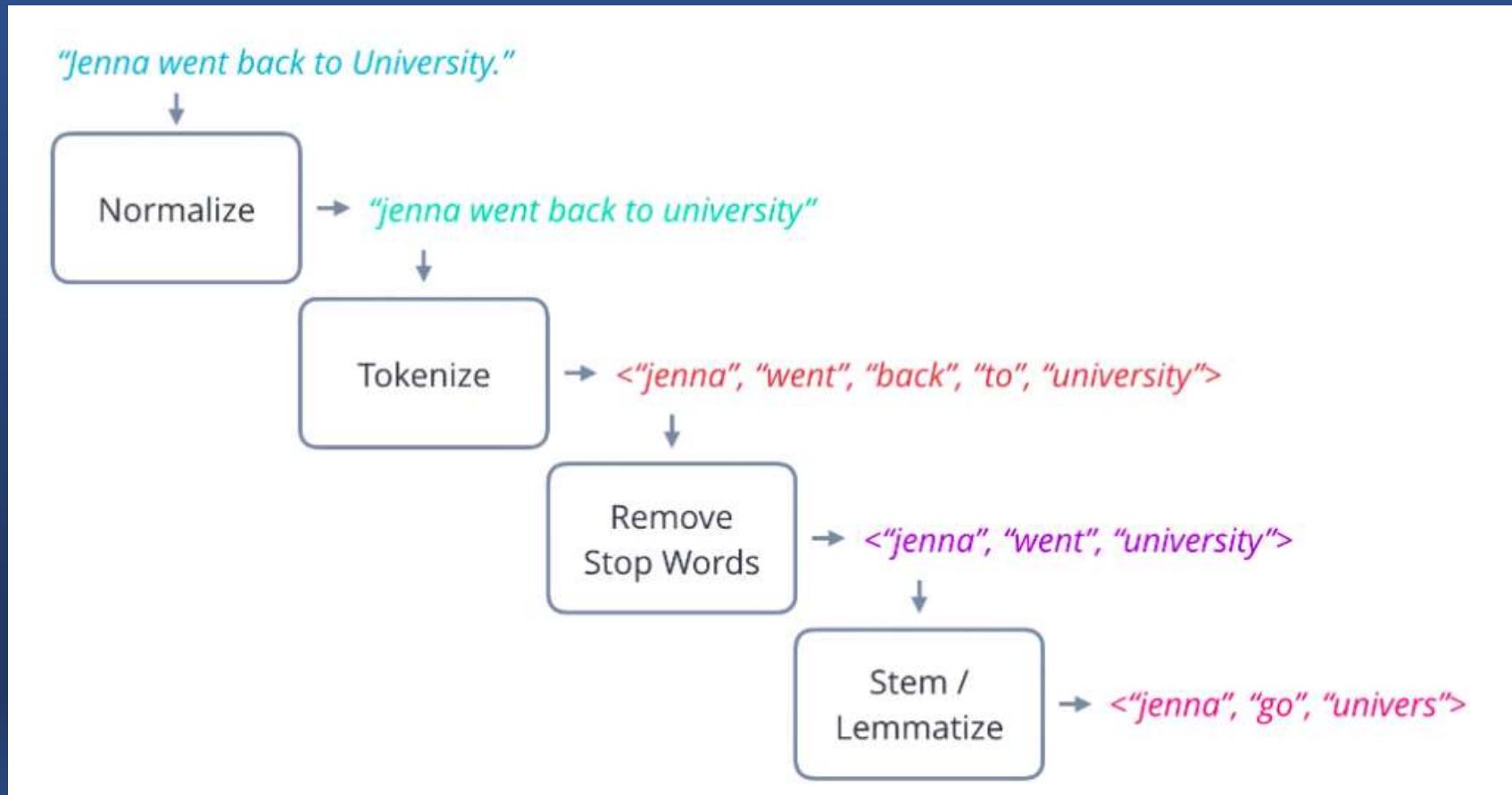
Weizenbaum escribió: "No me había dado cuenta... de que exposiciones extremadamente cortas a un programa de ordenador relativamente simple podrían inducir un poderoso pensamiento delirante en personas bastante normales"

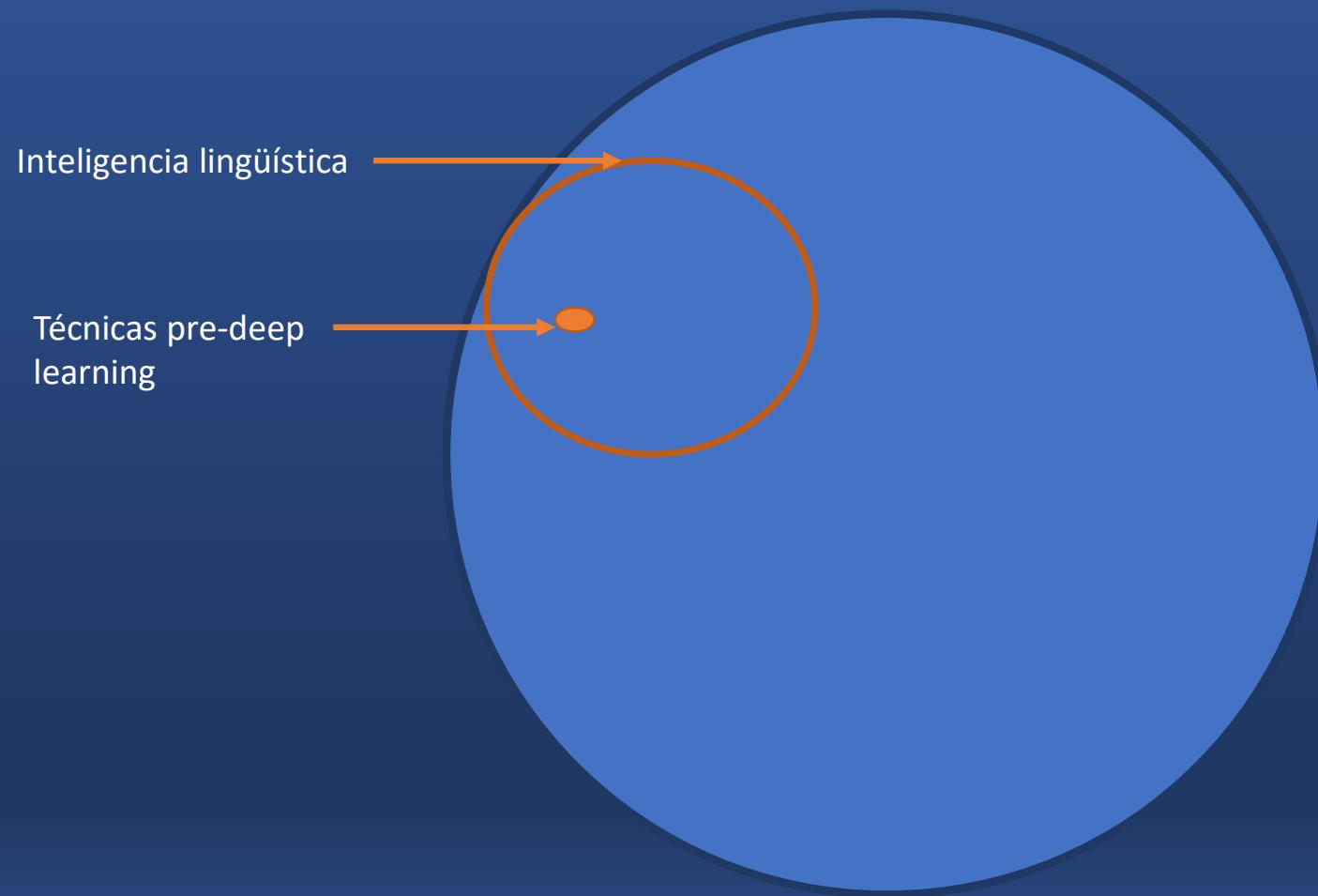


Inteligencia lingüística

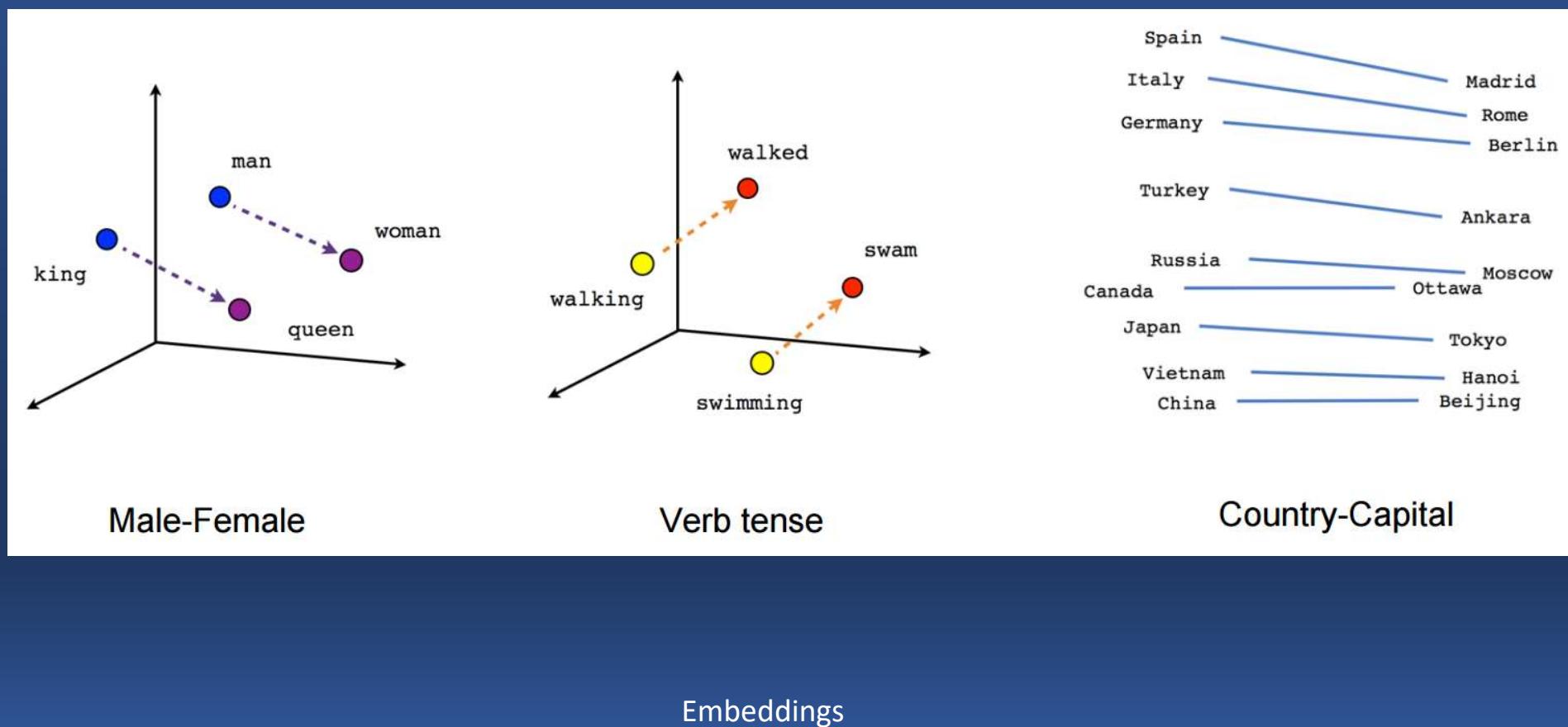
ELIZA

## IA lingüísitca

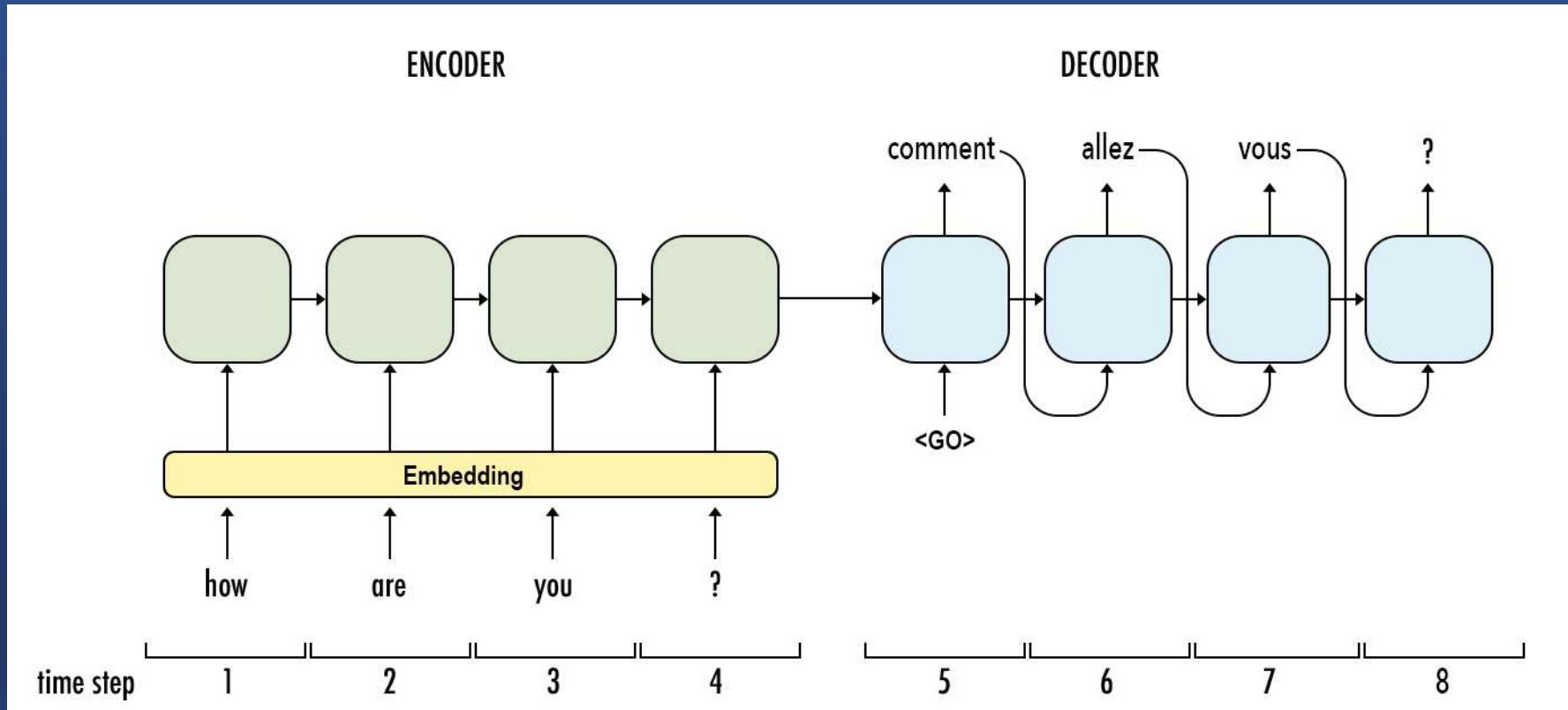




# ¿Y con deep learning?

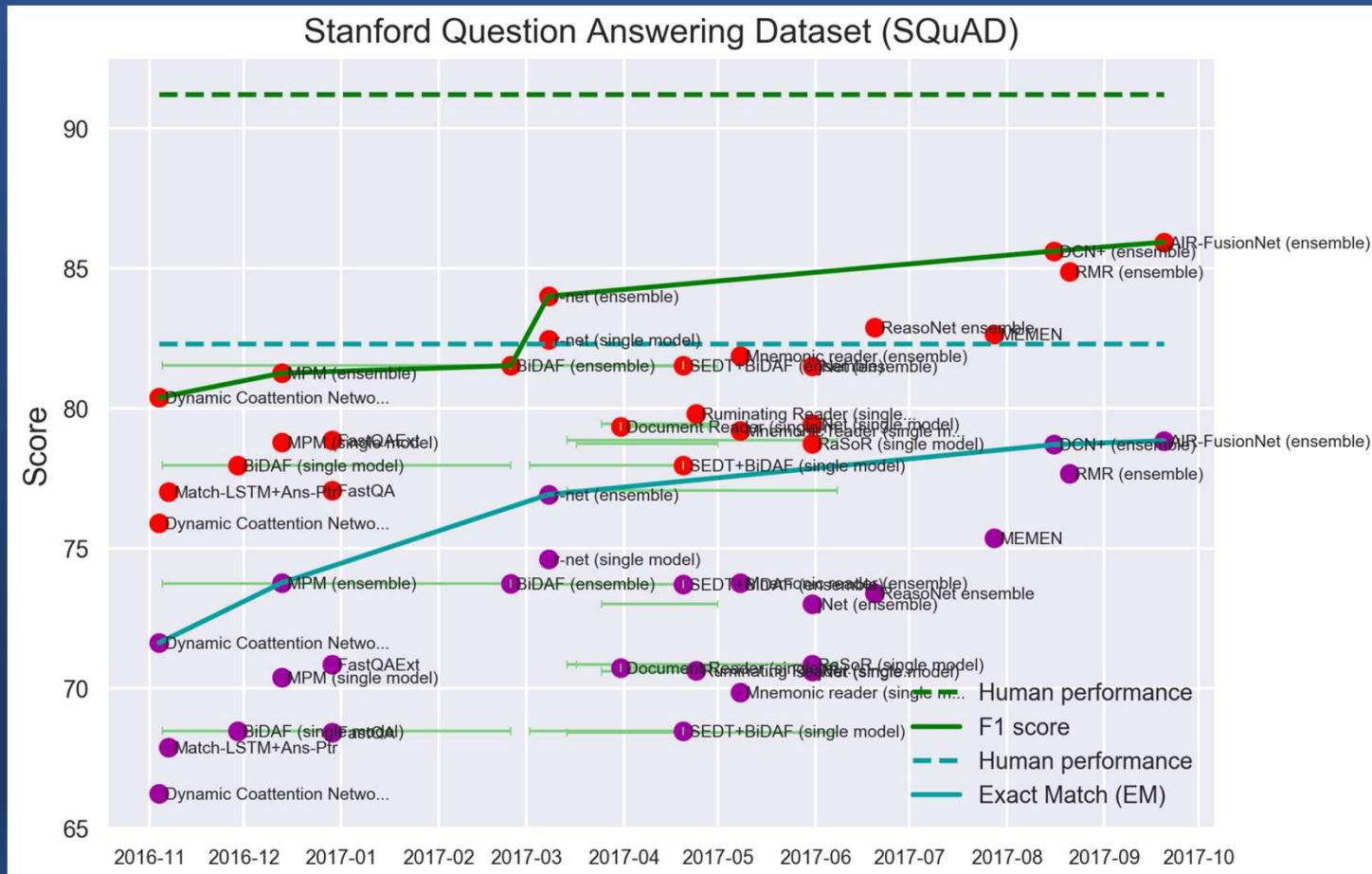


# ¿Y con deep learning?



Redes recurrentes (LSTM, 1997)

## ¿Y con deep learning?

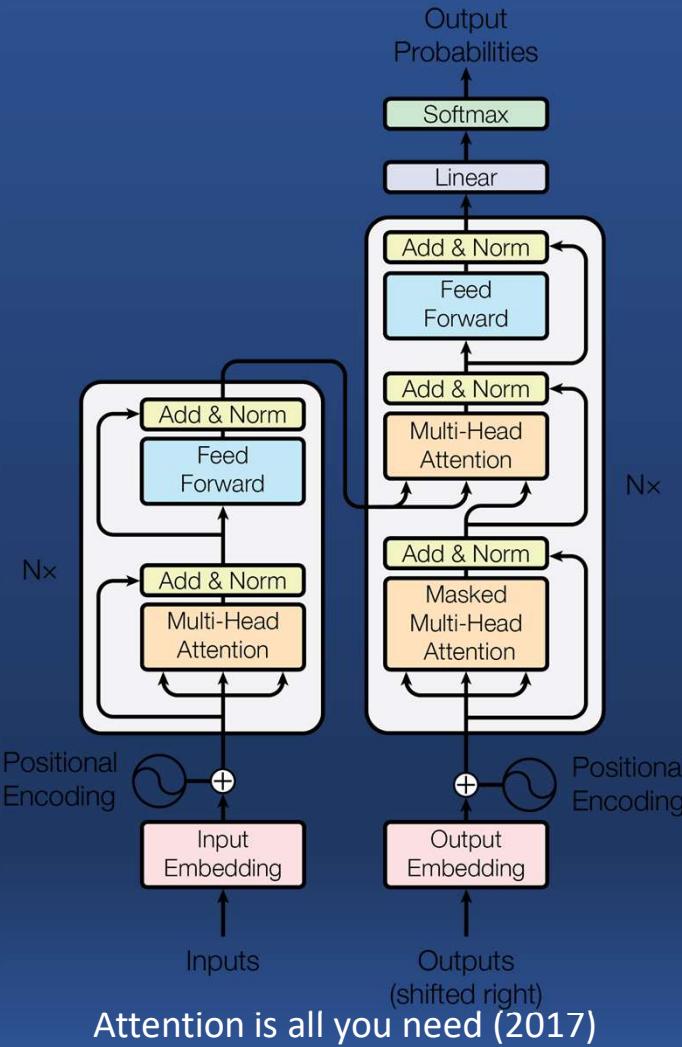


¿Y con deep learning?

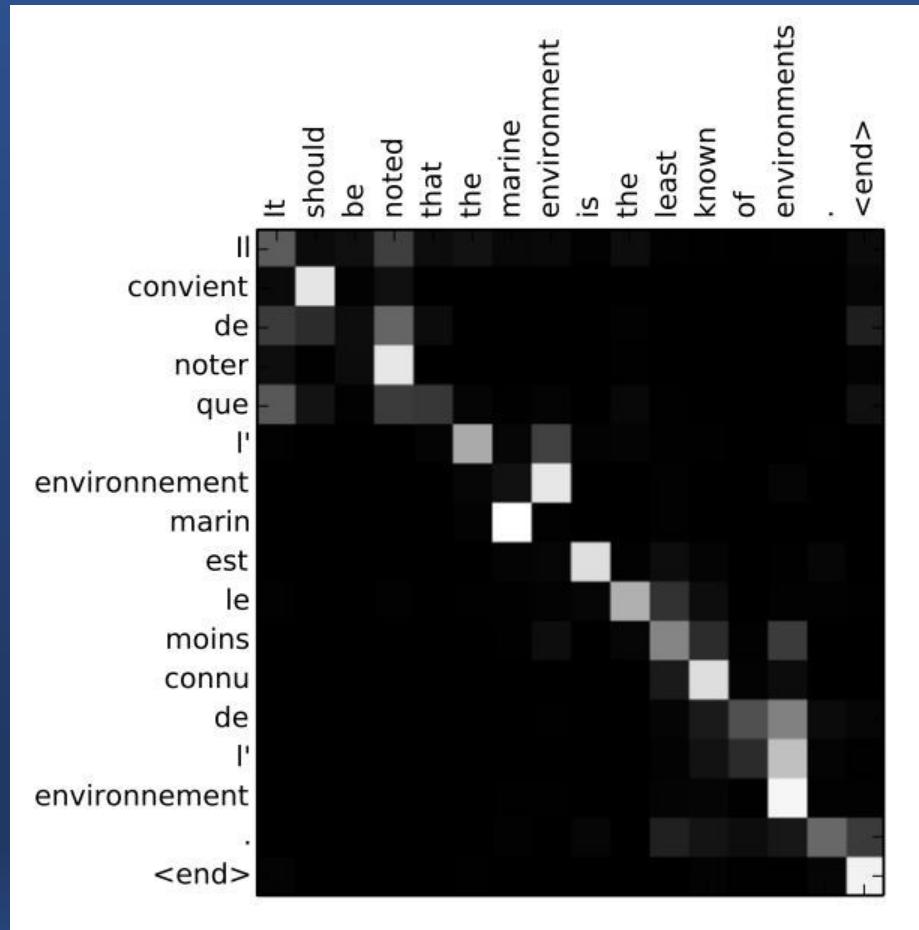


Transformers

# ¿Y con deep learning?

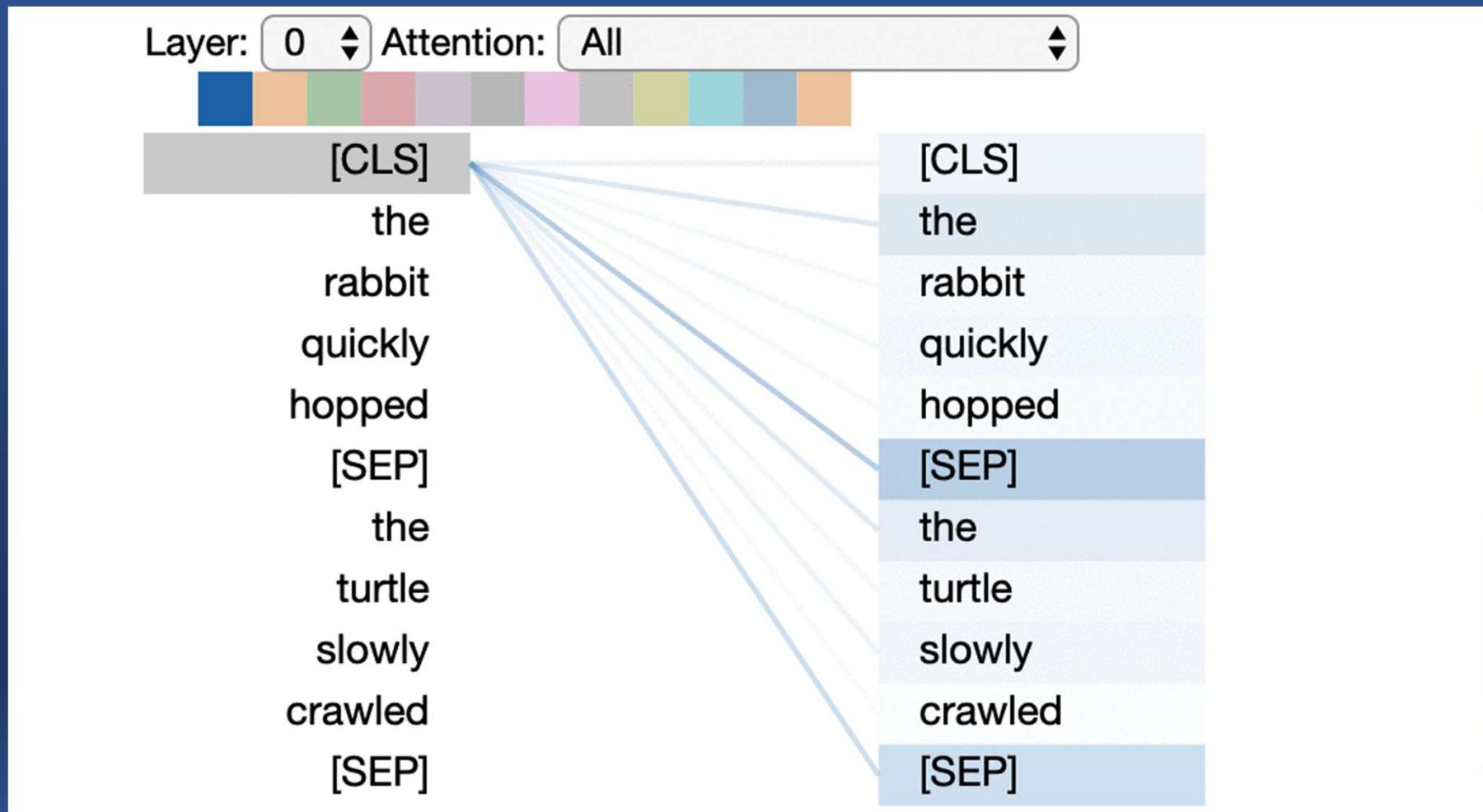


## ¿Y con deep learning?



Traducción con mecanismos de atención

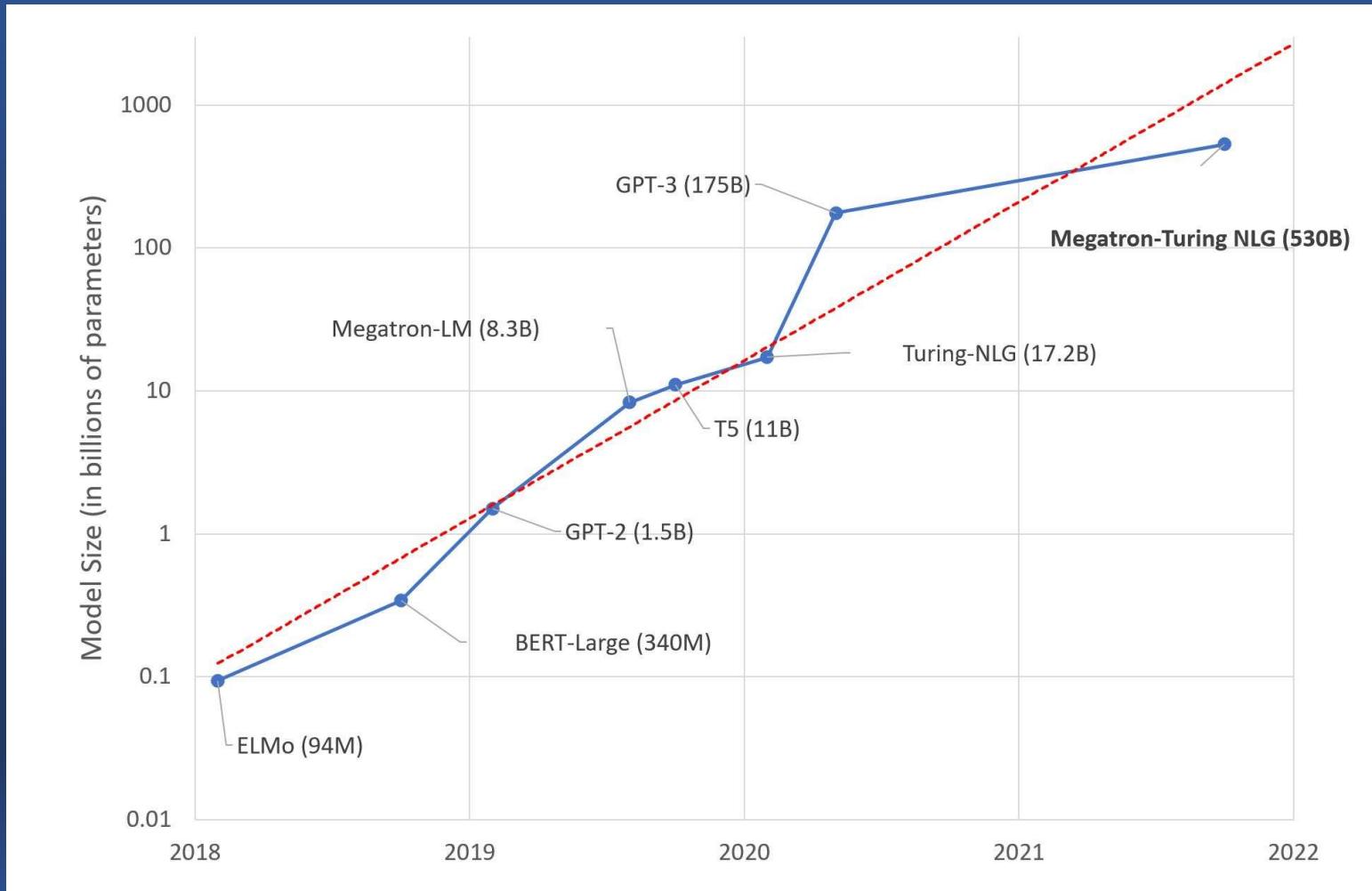
## ¿Y con deep learning?



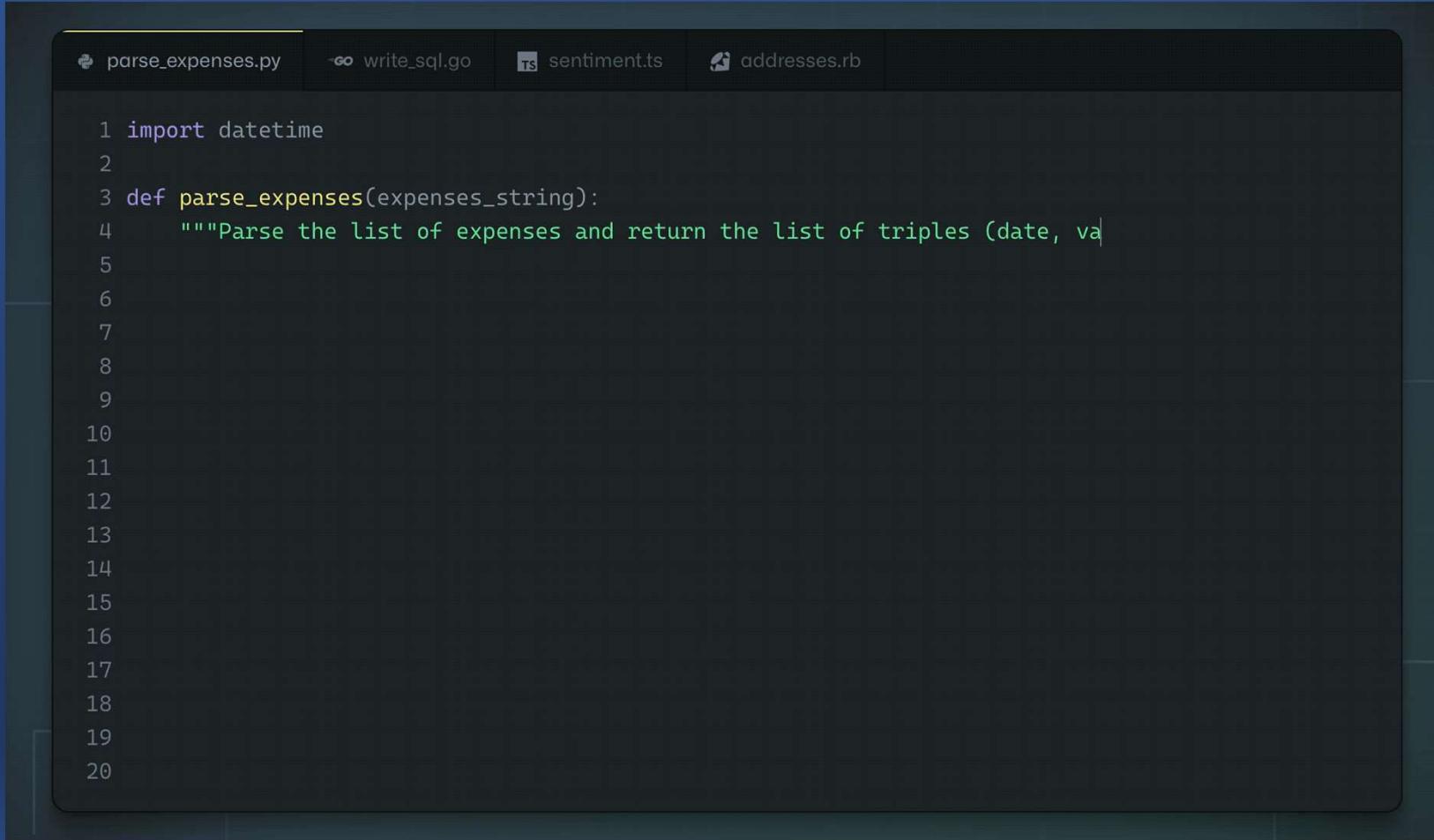
¿Y con deep learning?

“Griezmann’s announcement comes as a bit of a shock. After enduring the drama surrounding his potential last summer, many thought he was committed to Atletico for more than a year, but the Frenchman seems to have changed his mind.”

## ¿Y con deep learning?



# ¿Y con deep learning?



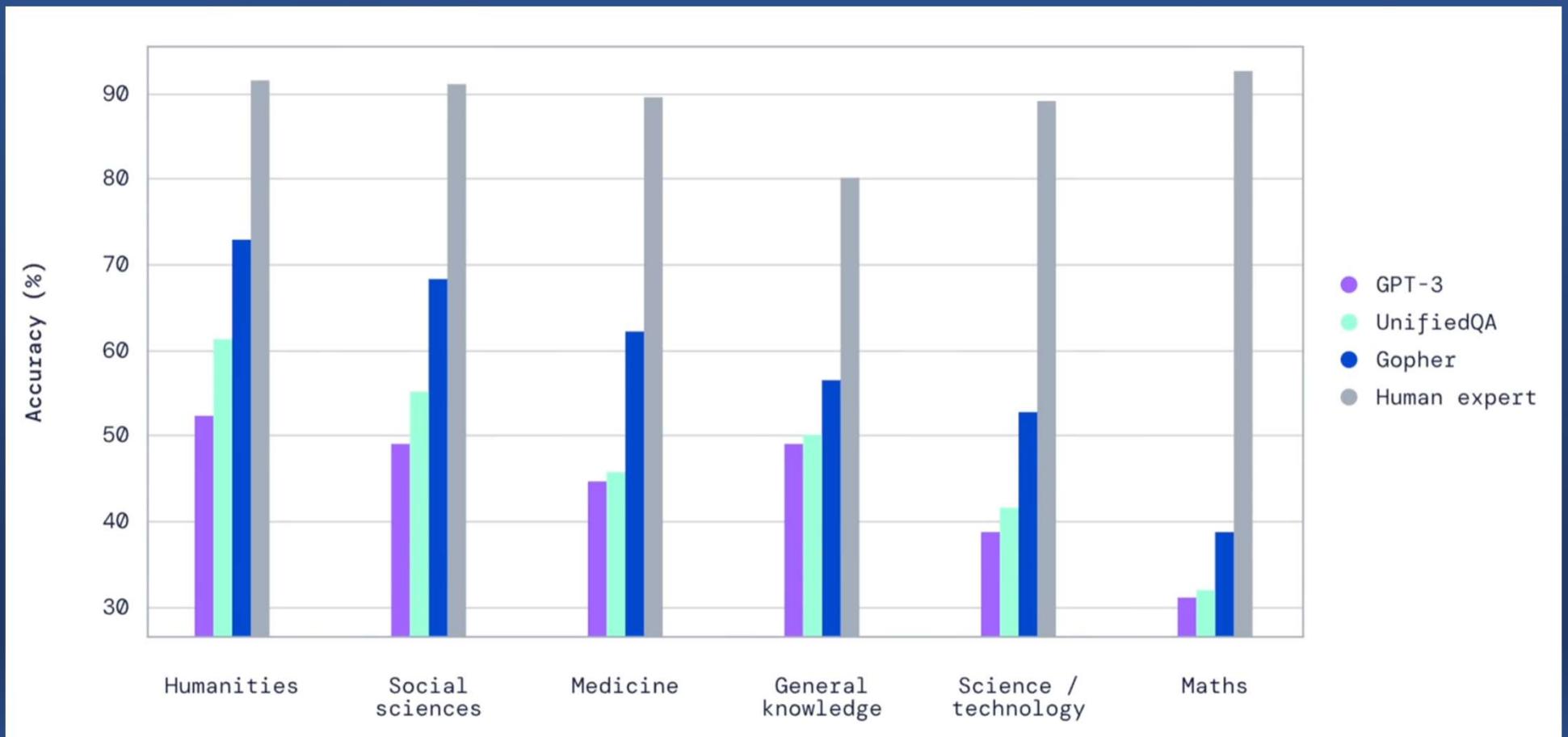
A screenshot of a terminal window with a dark background. At the top, there are four tabs: 'parse\_expenses.py' (highlighted in blue), 'write\_sql.go', 'sentiment.ts', and 'addresses.rb'. The main area of the terminal shows the following Python code:

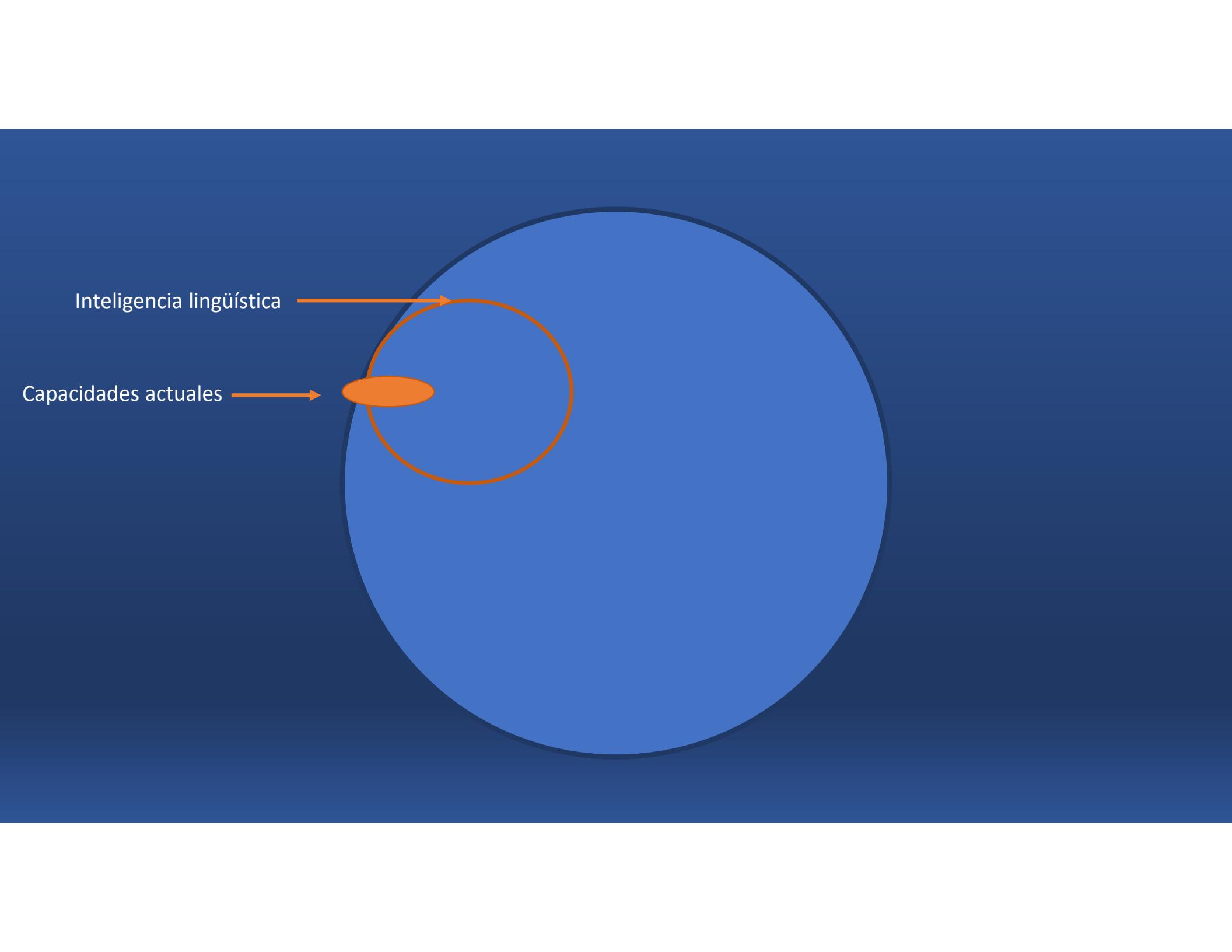
```
1 import datetime
2
3 def parse_expenses(expenses_string):
4     """Parse the list of expenses and return the list of triples (date, va
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
```

The code starts with an import statement for 'datetime'. It then defines a function 'parse\_expenses' that takes a string parameter 'expenses\_string'. The function includes a multi-line docstring starting with '"""Parse the list of expenses and return the list of triples (date, va'. The code is numbered from 1 to 20 on the left side.

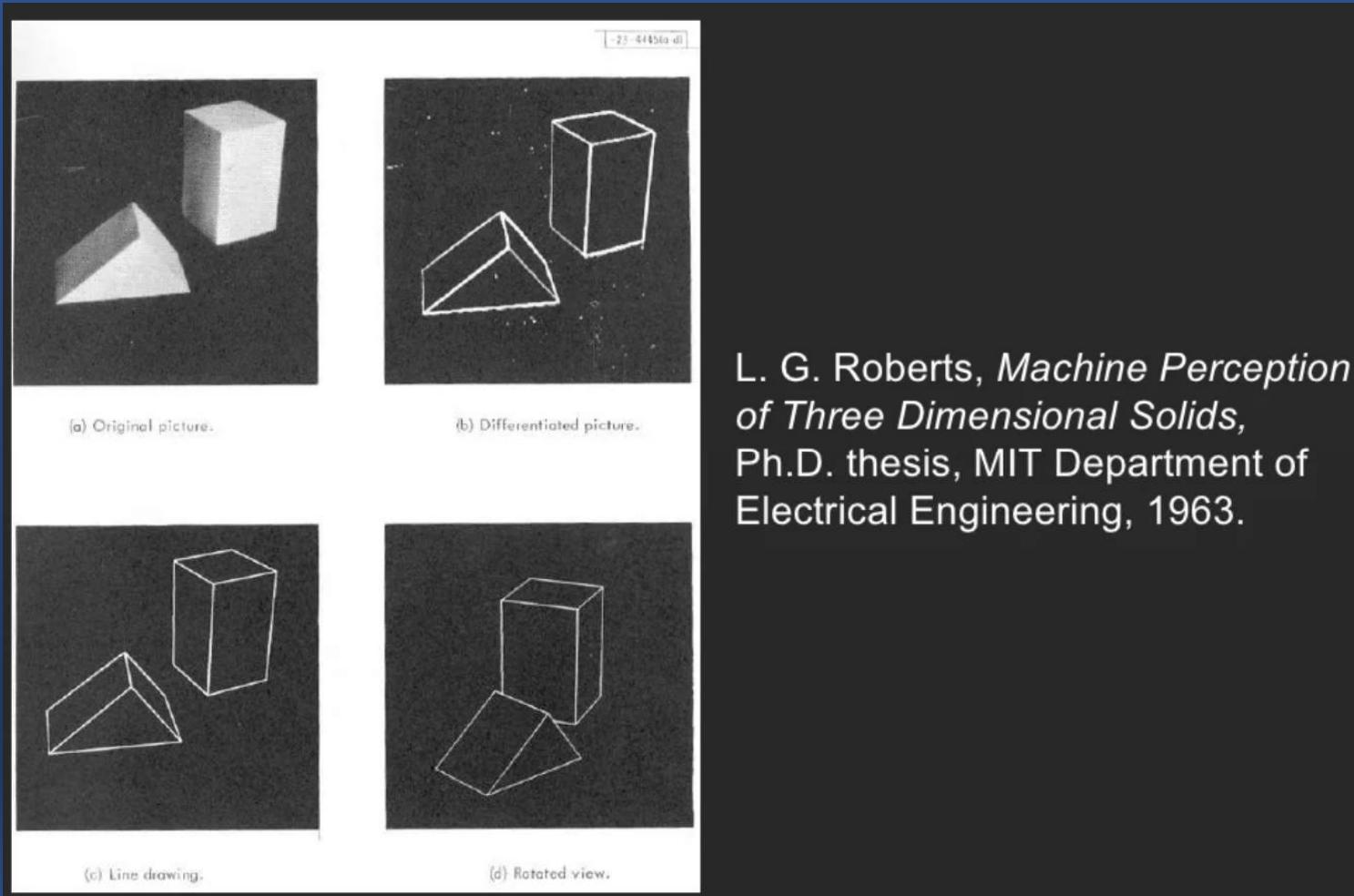
GPT-3 (2020)

## ¿Y con deep learning?



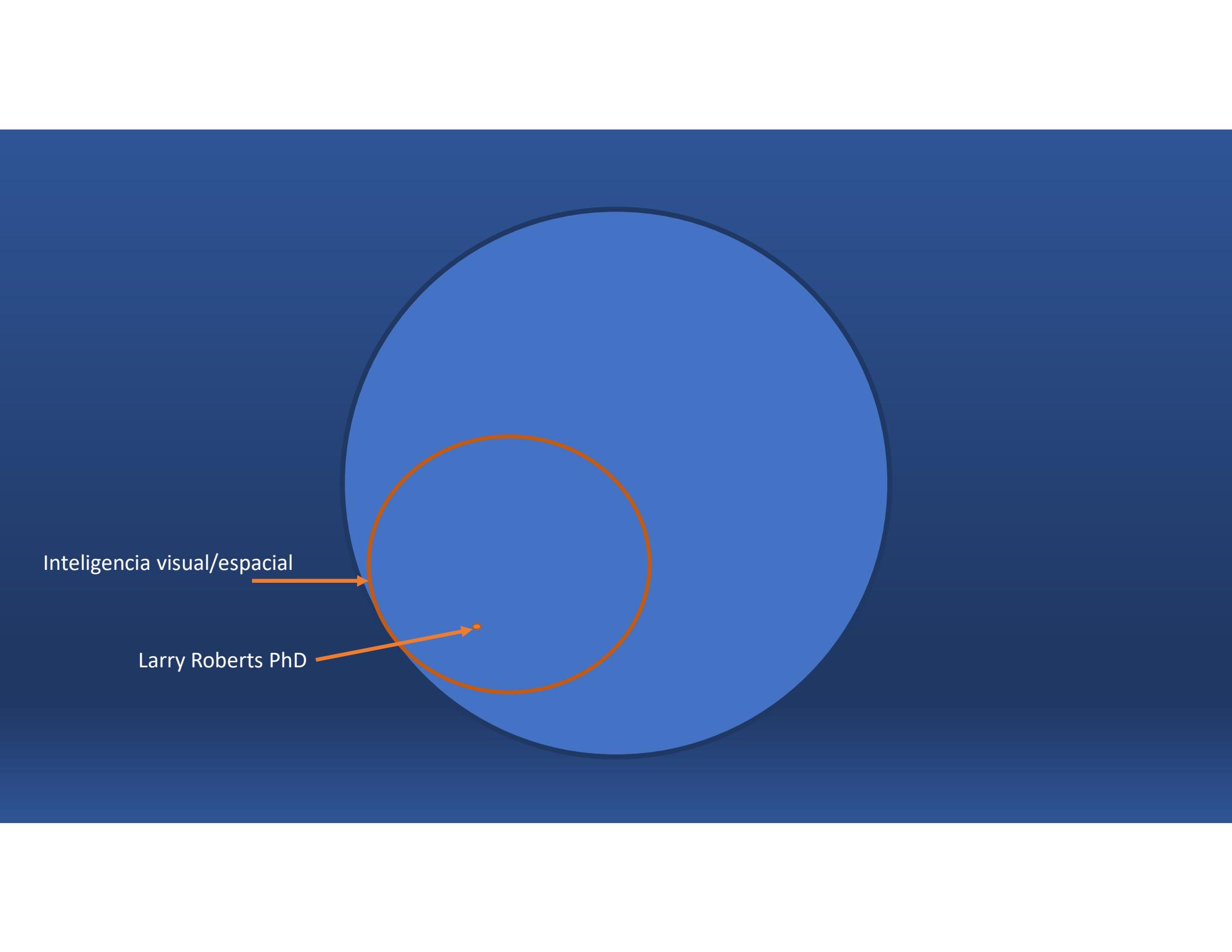


## Avances en inteligencia visual



L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Larry Roberts PhD (1963)



Inteligencia visual/espacial

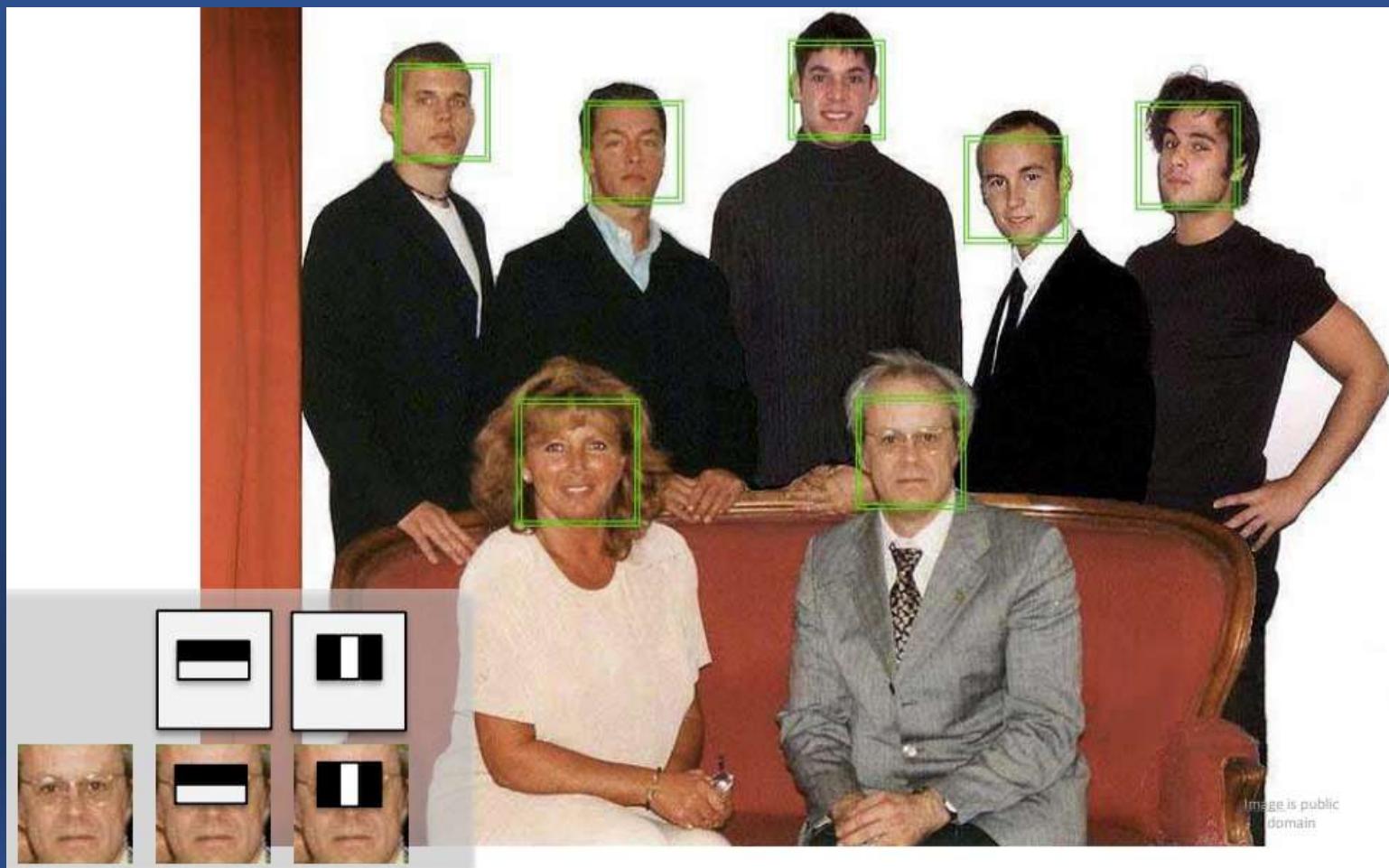
Larry Roberts PhD

## IA visual

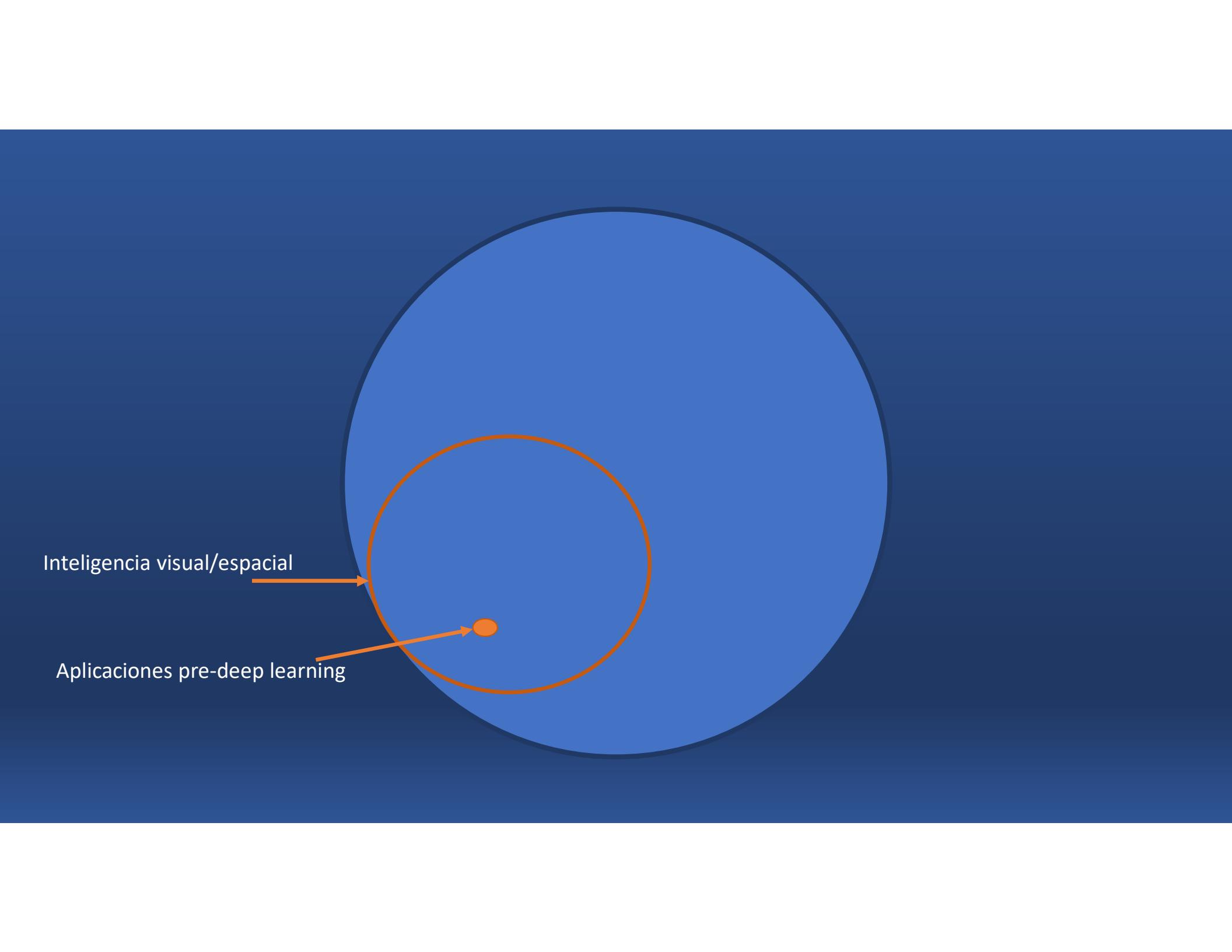


Object Recognition, SIFT (1999)

## IA visual



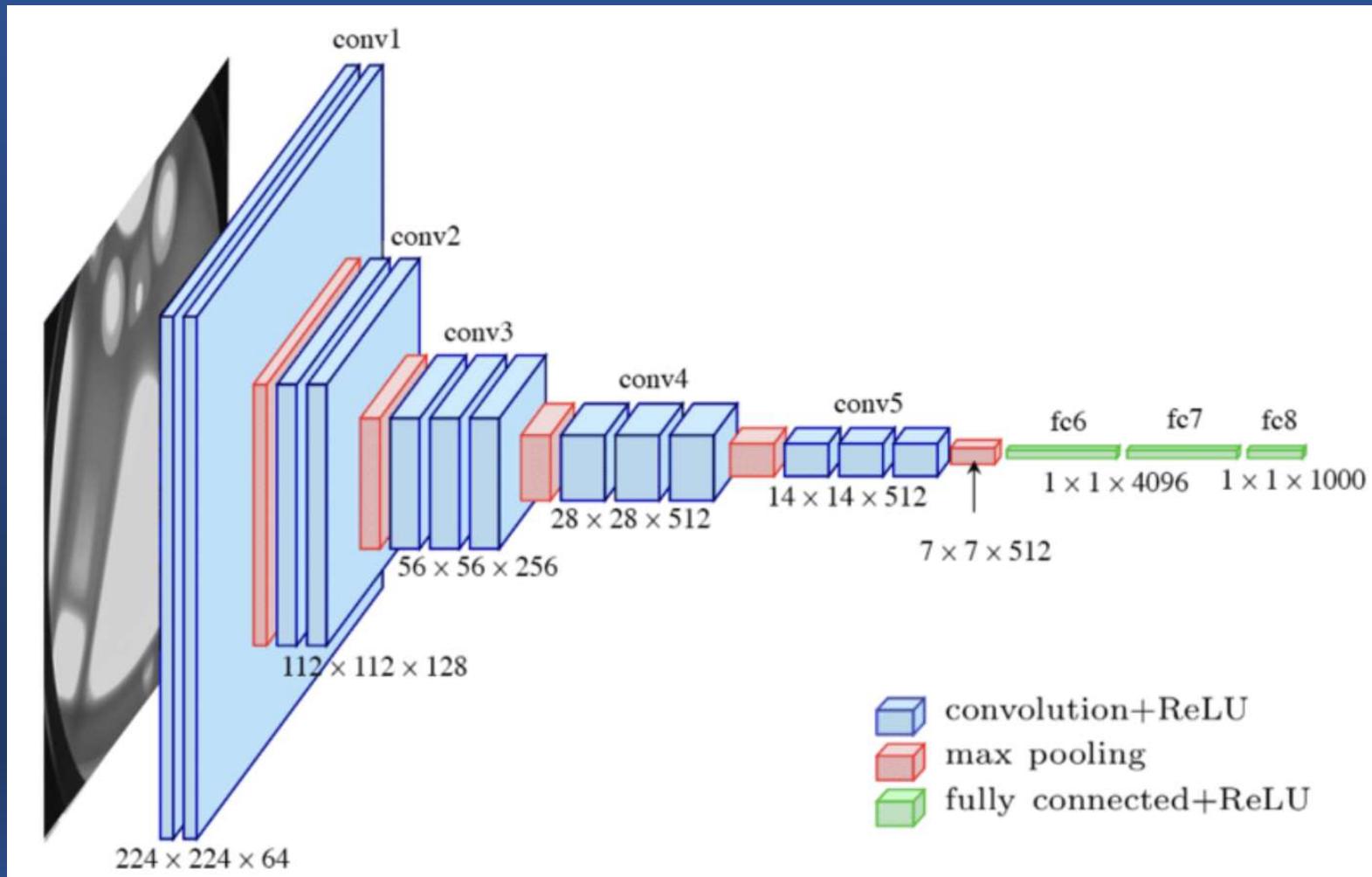
Face detection, Viola & Jones (2001)



Inteligencia visual/espacial

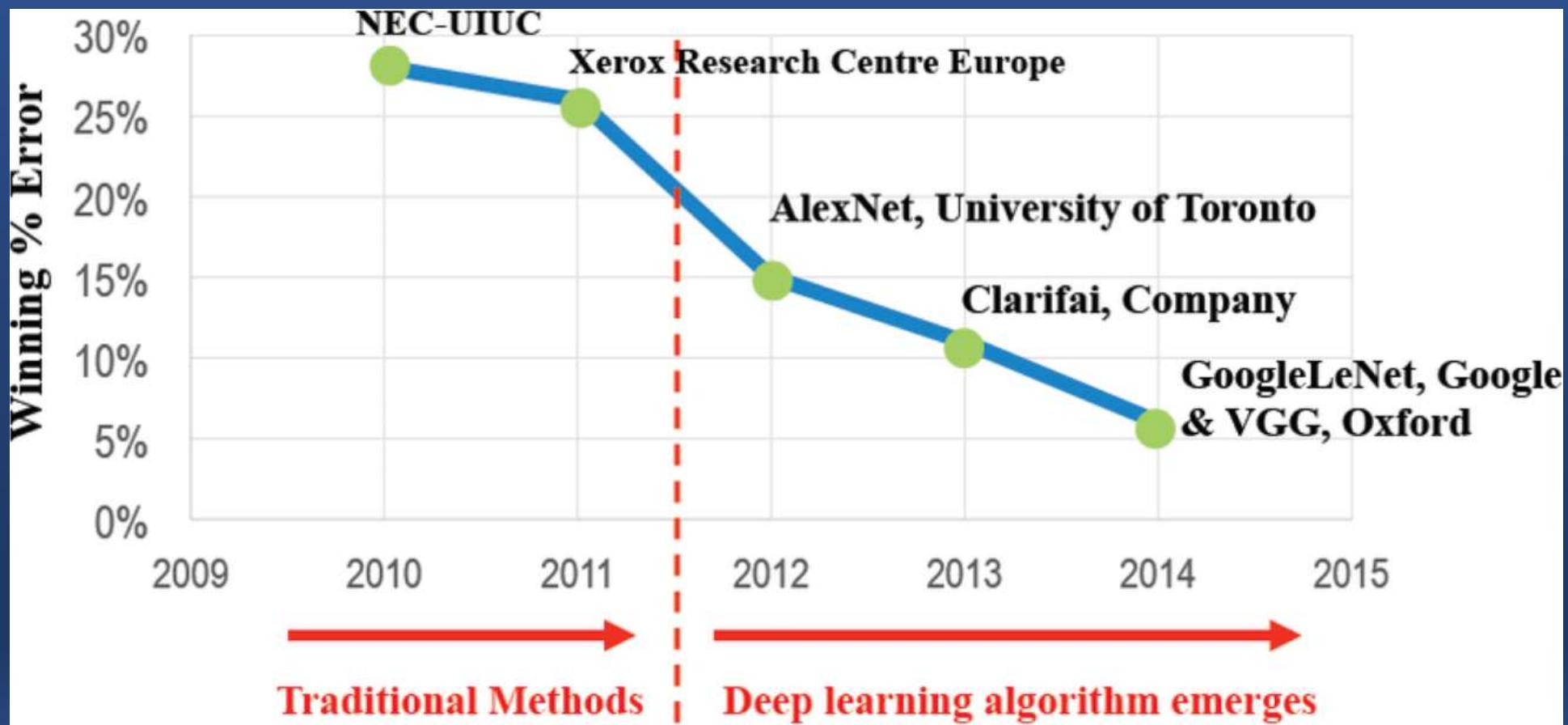
Aplicaciones pre-deep learning

## ¿Y con deep learning?



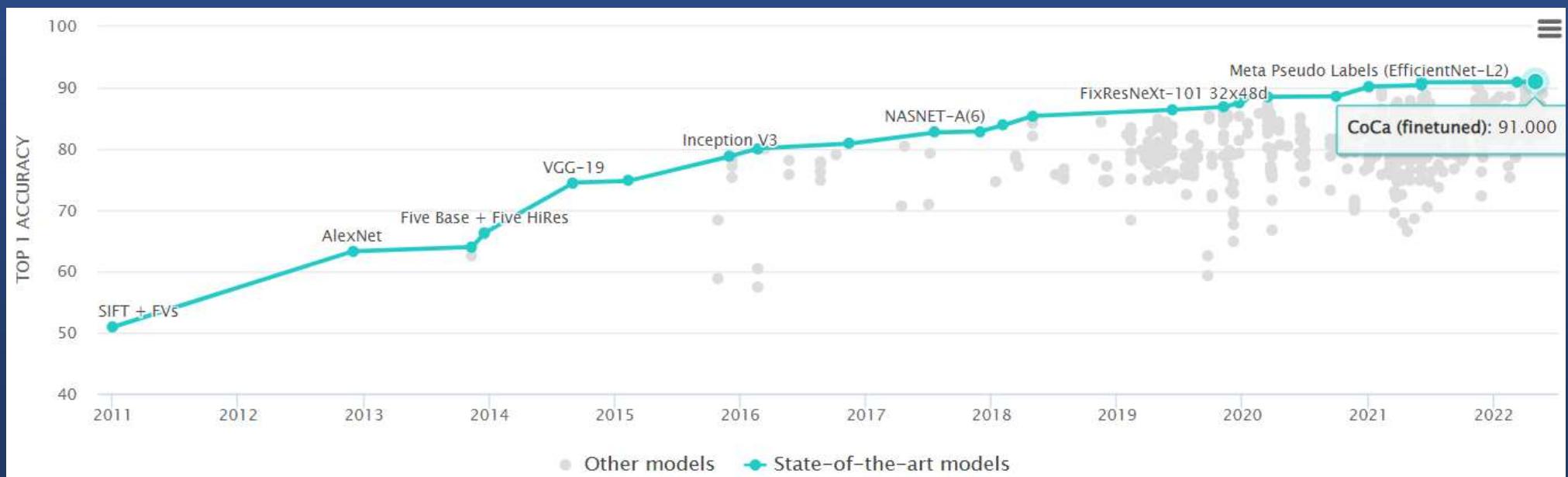
VGG16 (2014)

## ¿Y con deep learning?



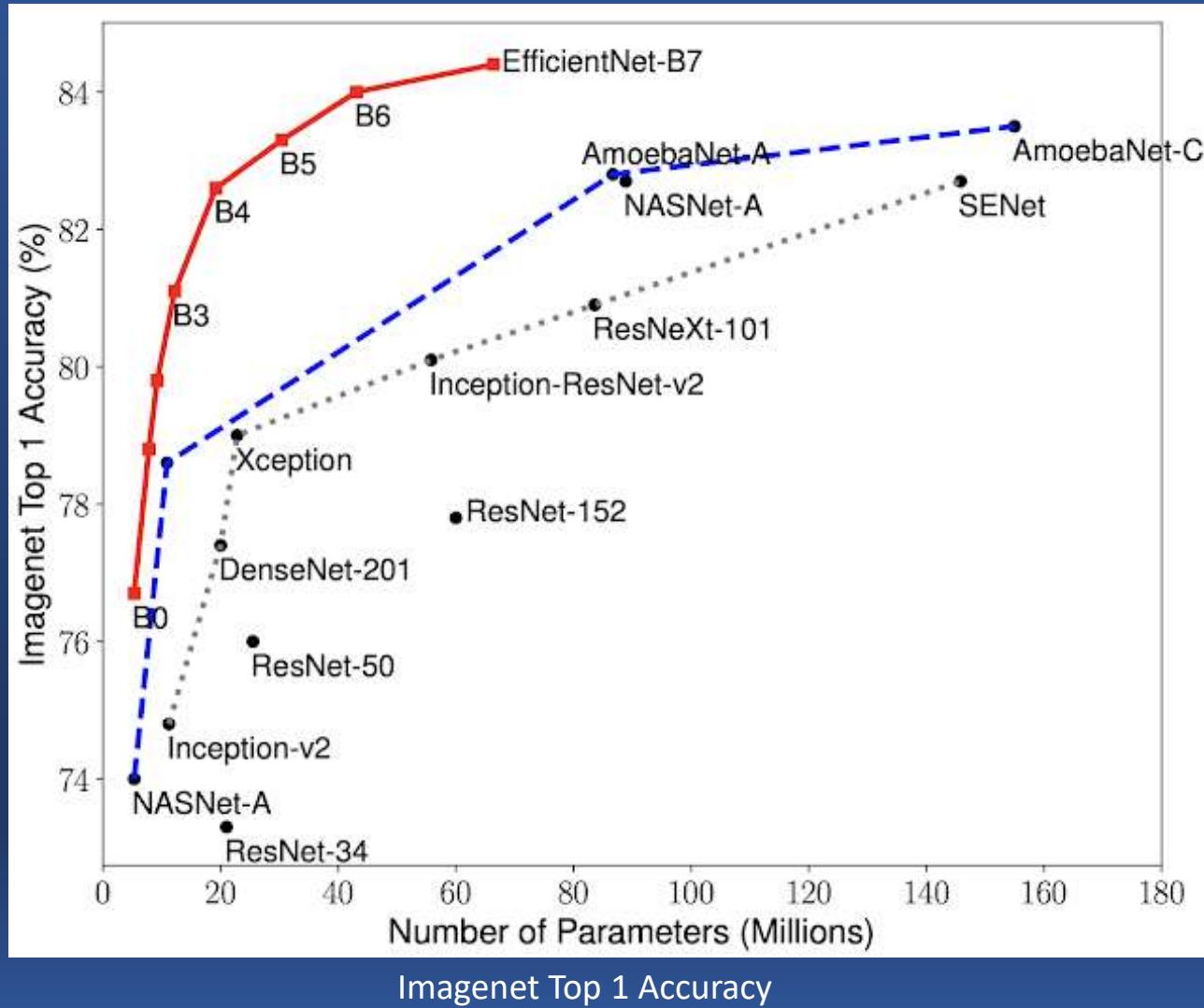
VGG16 (2014)

## ¿Y con deep learning?

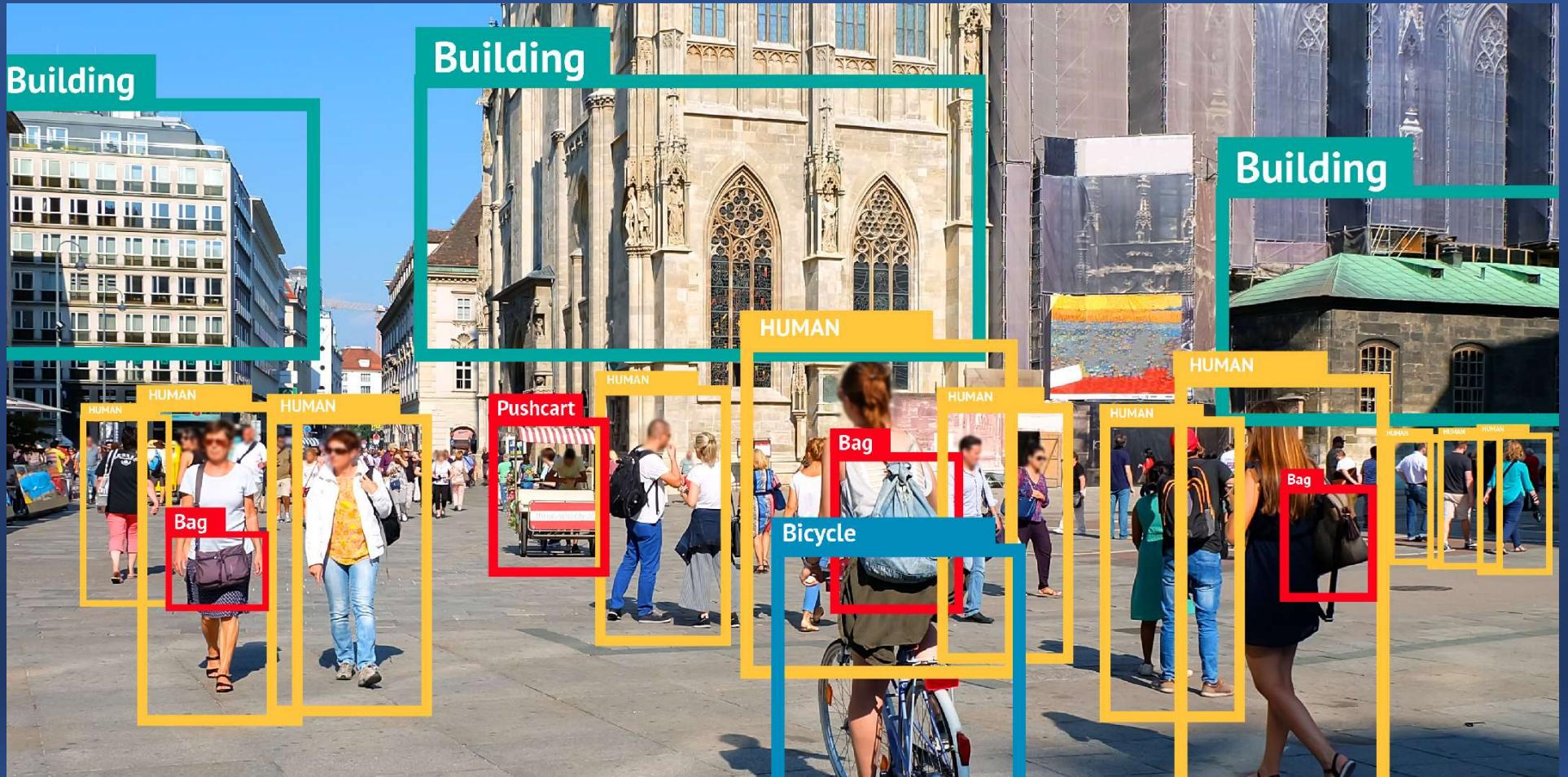


Imagenet Top 1 Accuracy

## ¿Y con deep learning?



## ¿Y con deep learning?



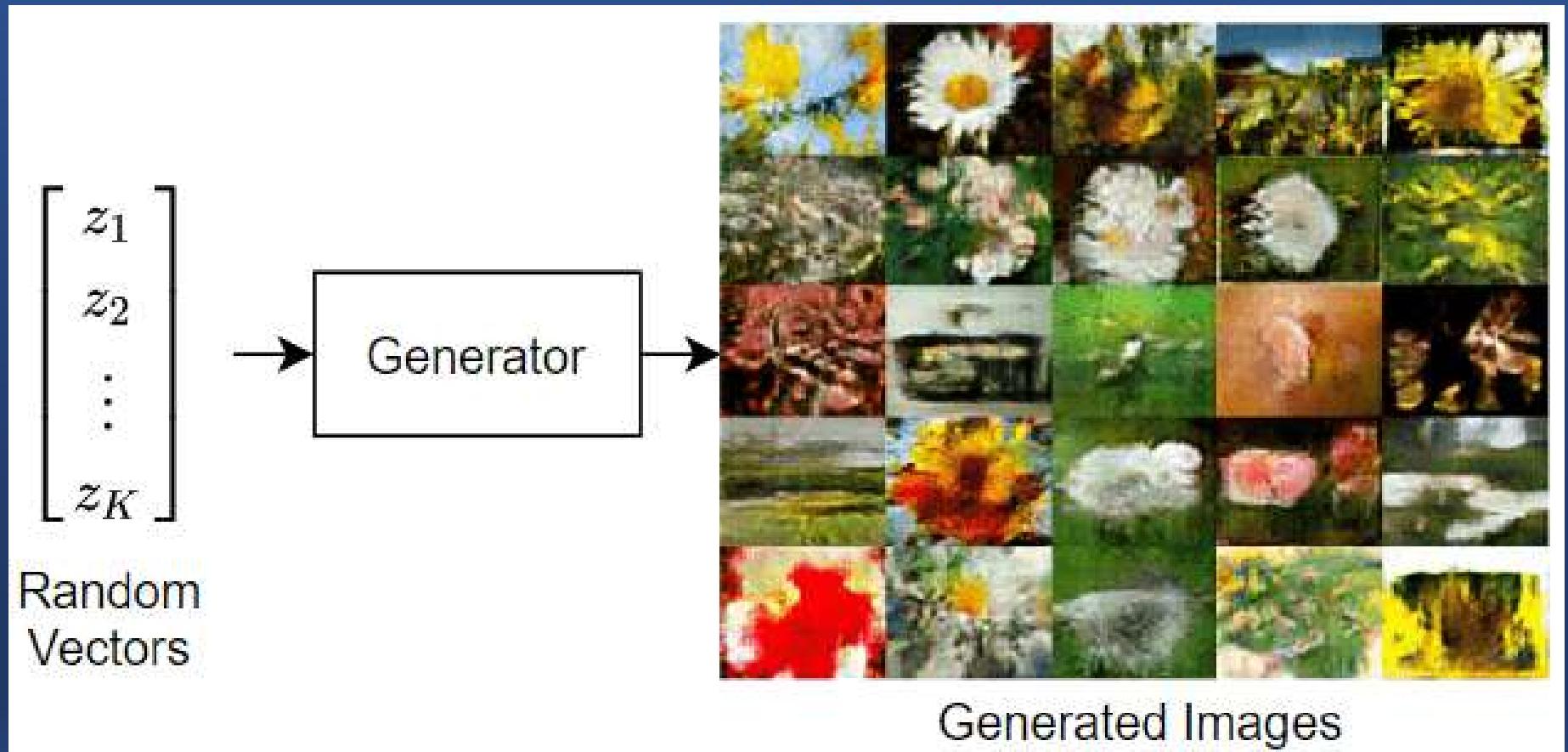
Mask R-CNN (2017)

¿Y con deep learning?



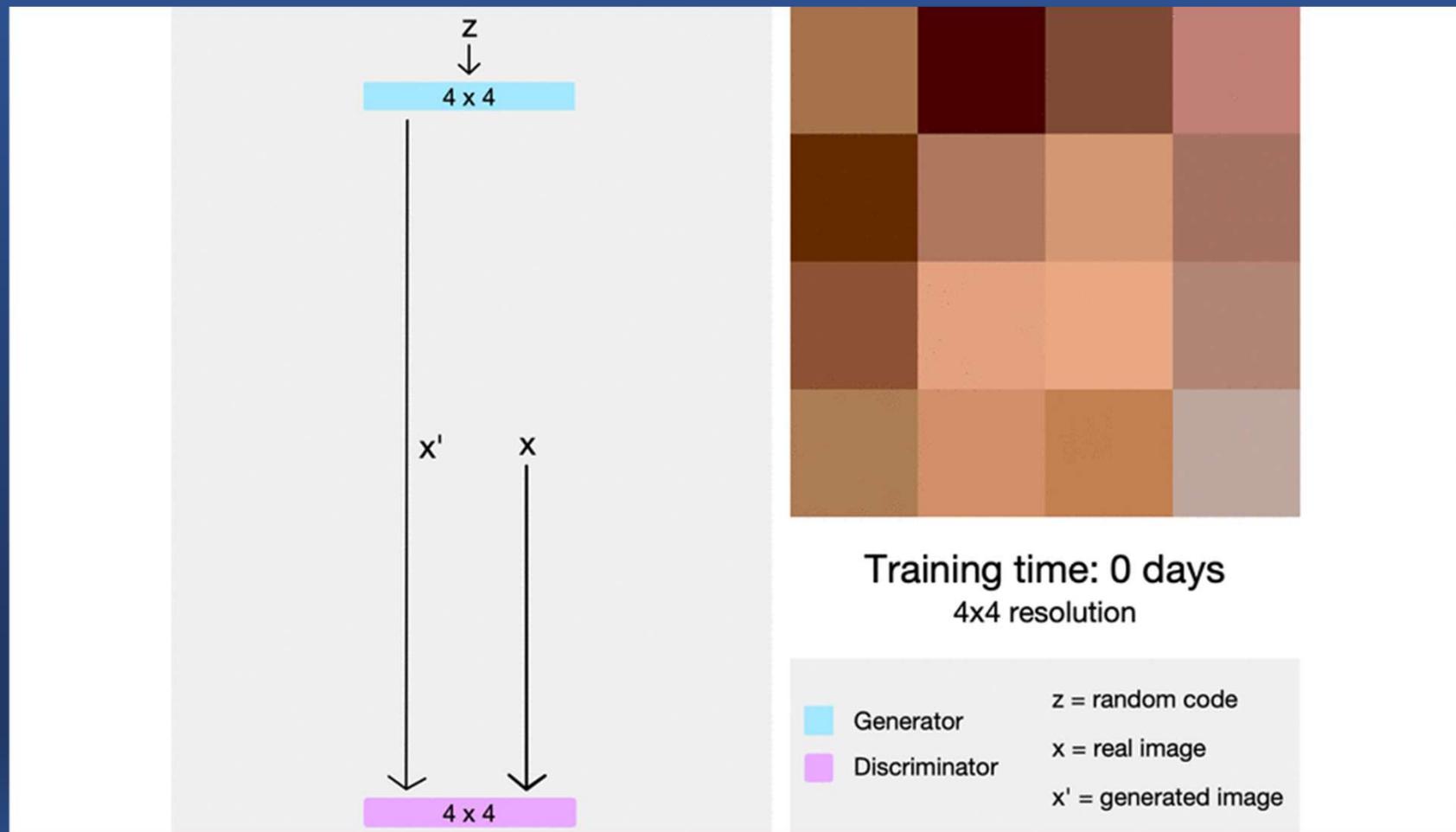
Deeplab (2018)

## ¿Y con deep learning?



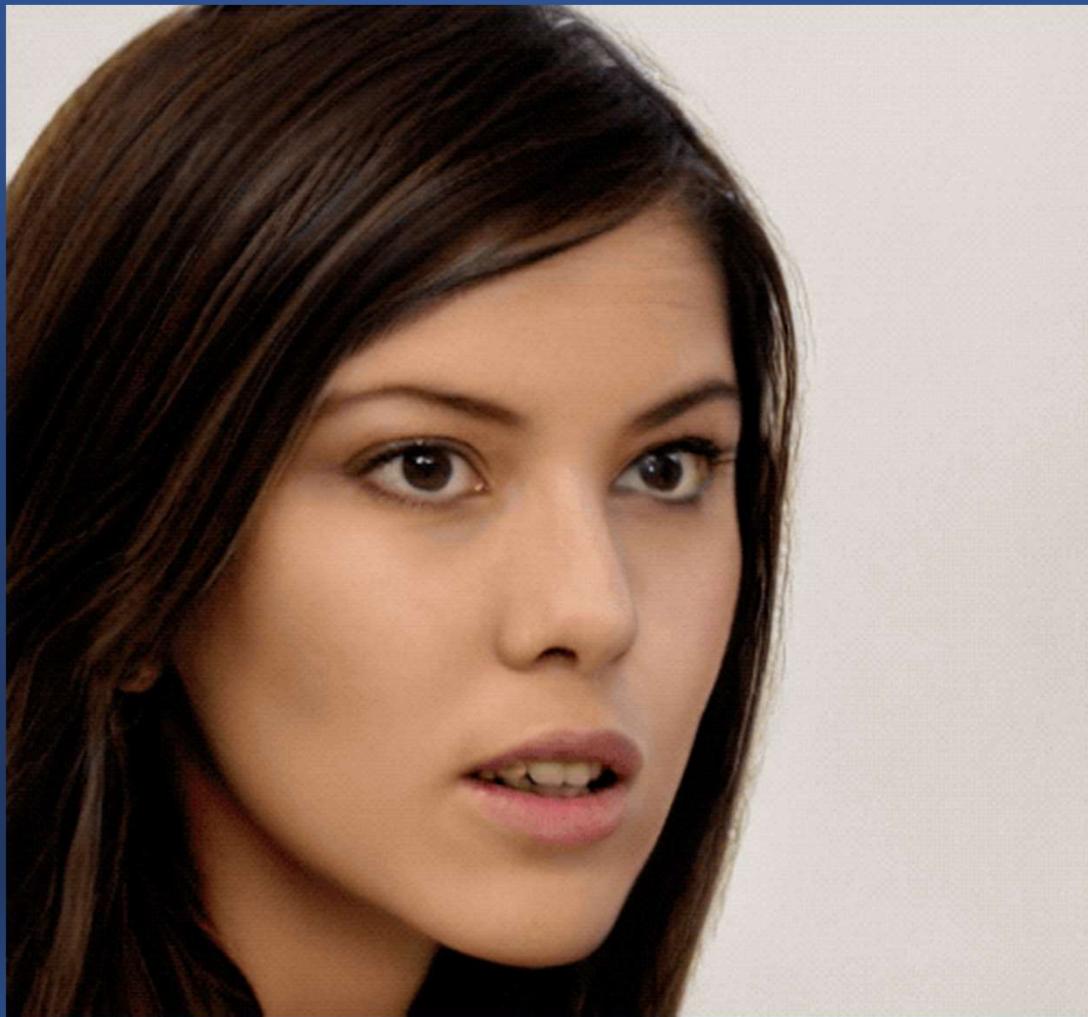
GAN

## ¿Y con deep learning?

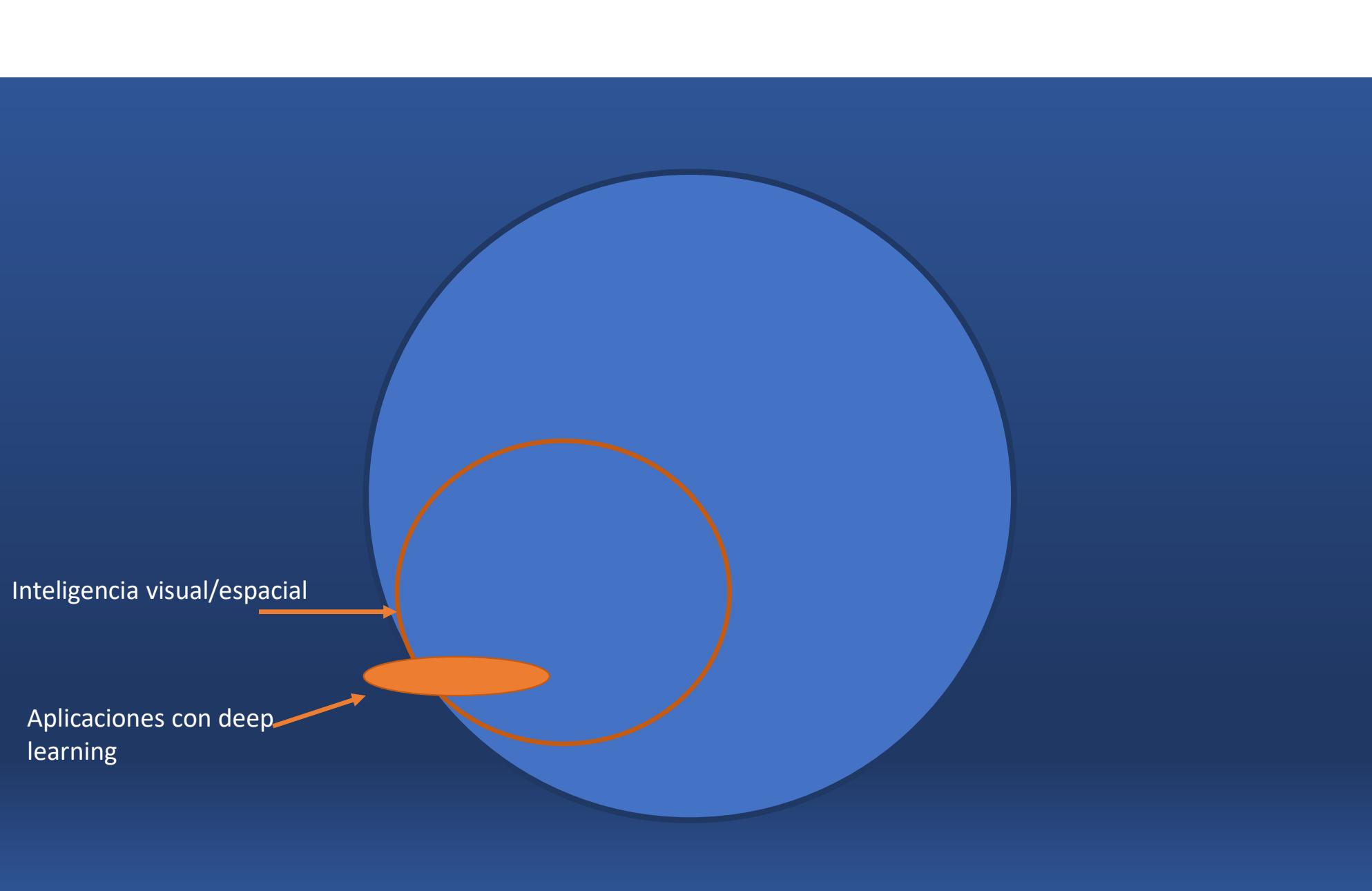


PGGAN (2018)

¿Y con deep learning?



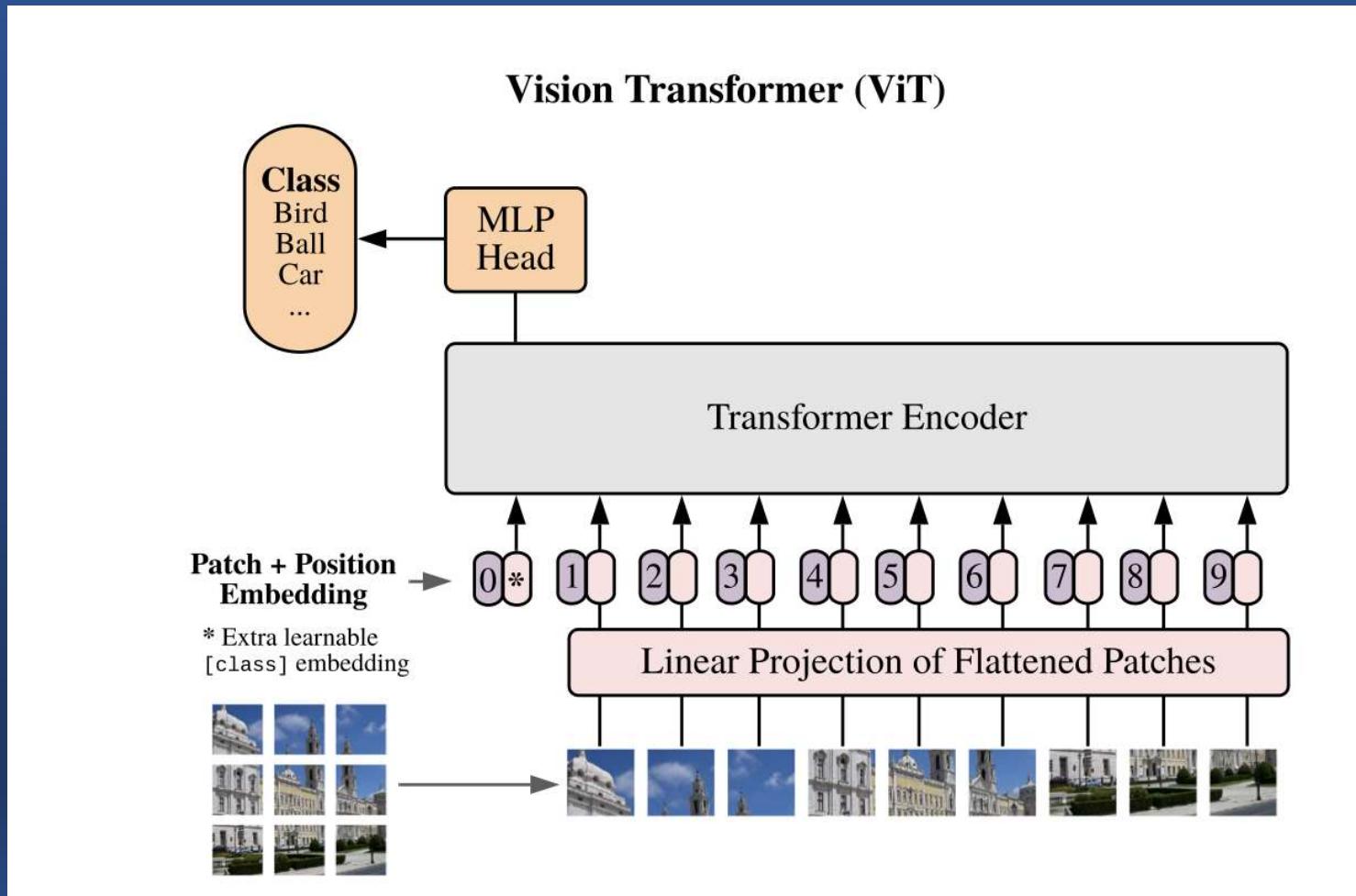
StyleGAN (2019)



Inteligencia visual/espacial

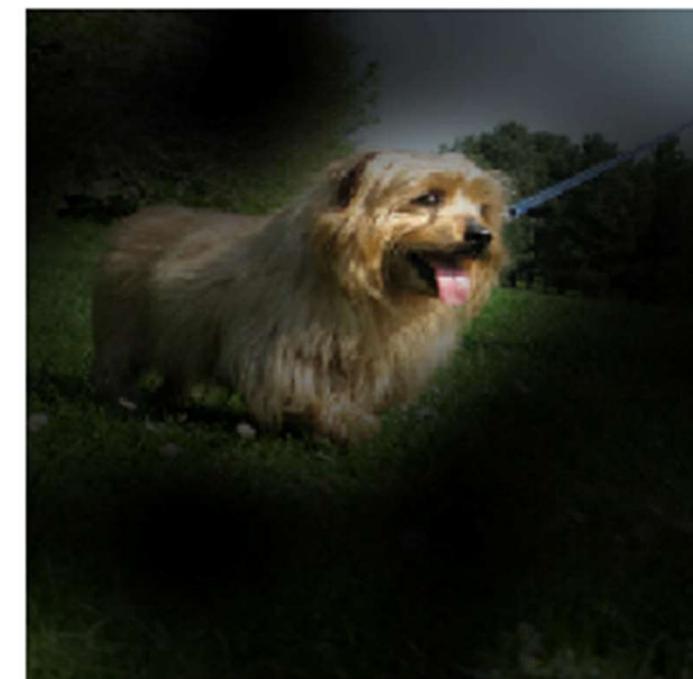
Aplicaciones con deep  
learning

# ¿Y con deep learning?



¿Y con deep learning?

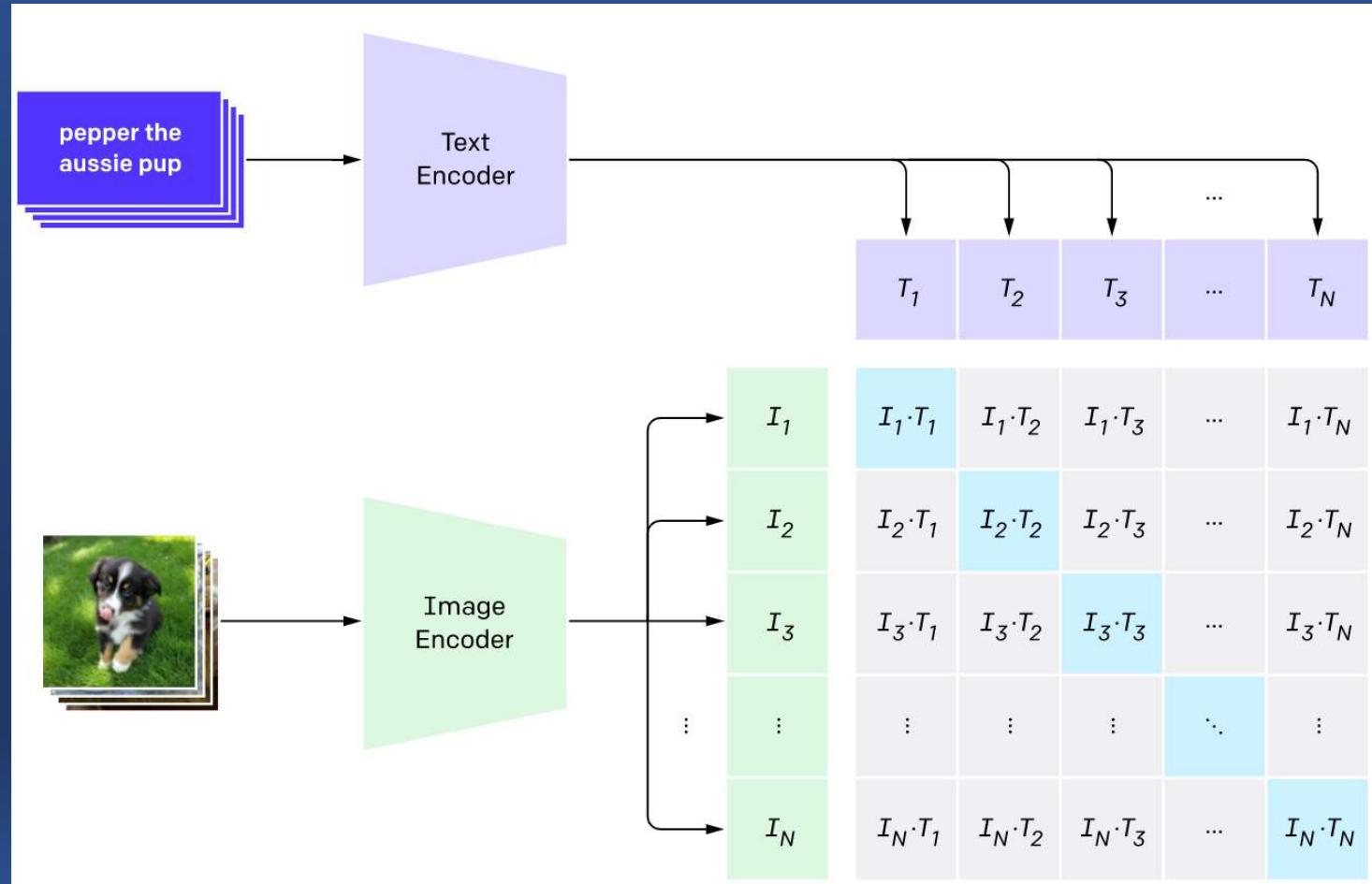
# Input      Attention



ViT (2020)

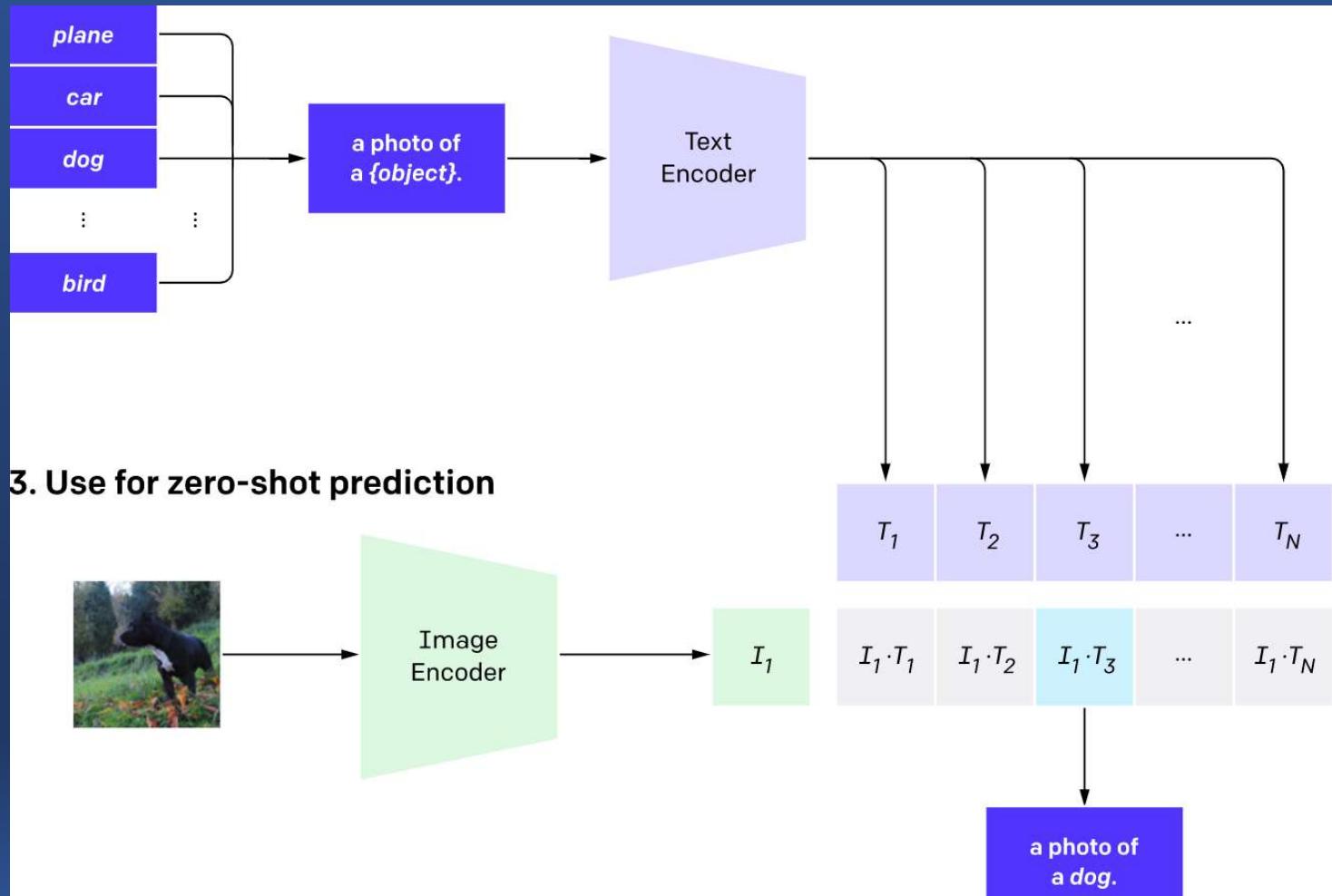
## CLIP (2021)

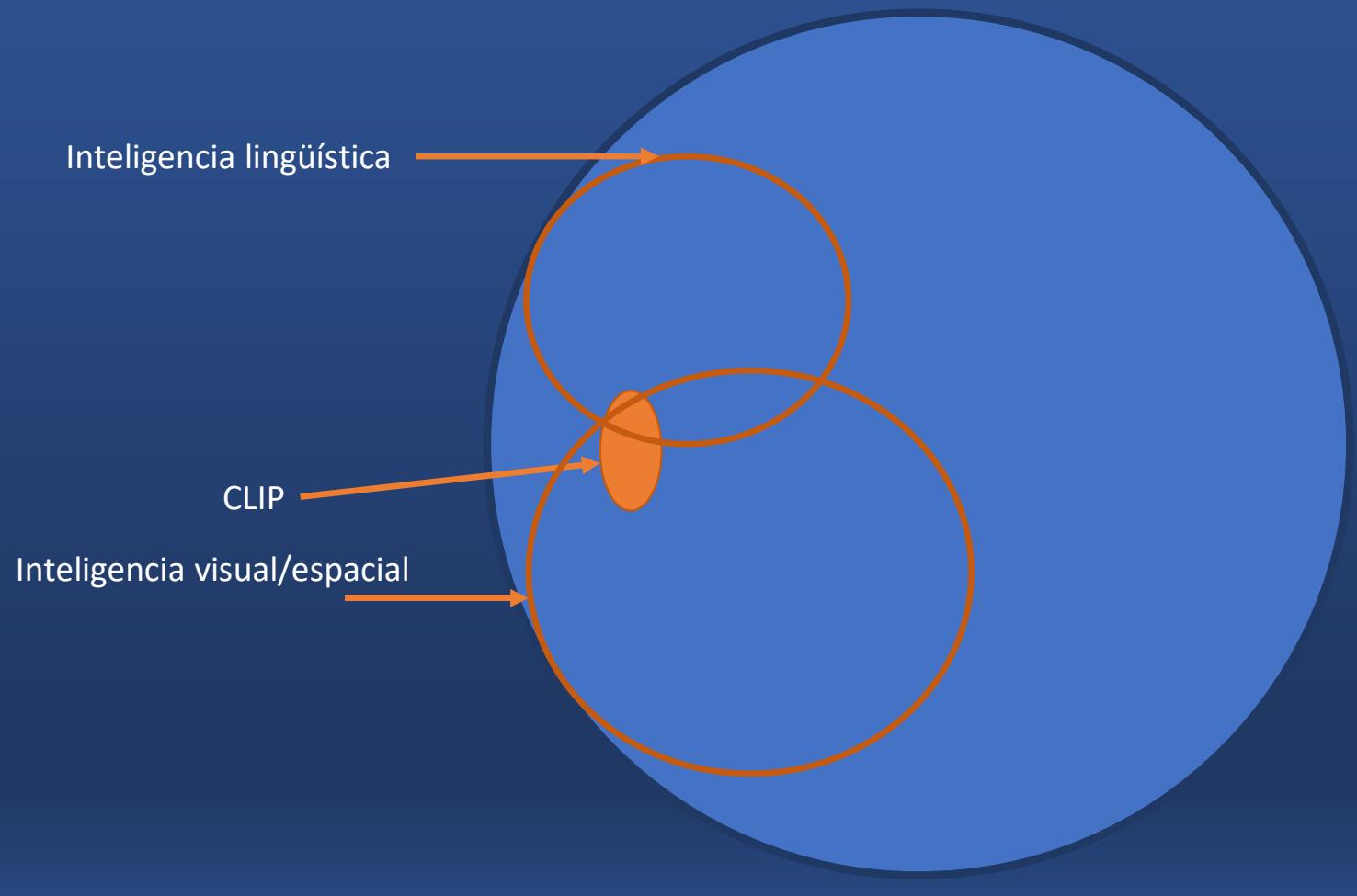
Crea embeddings de texto e imagen conjuntamente



## CLIP (2021)

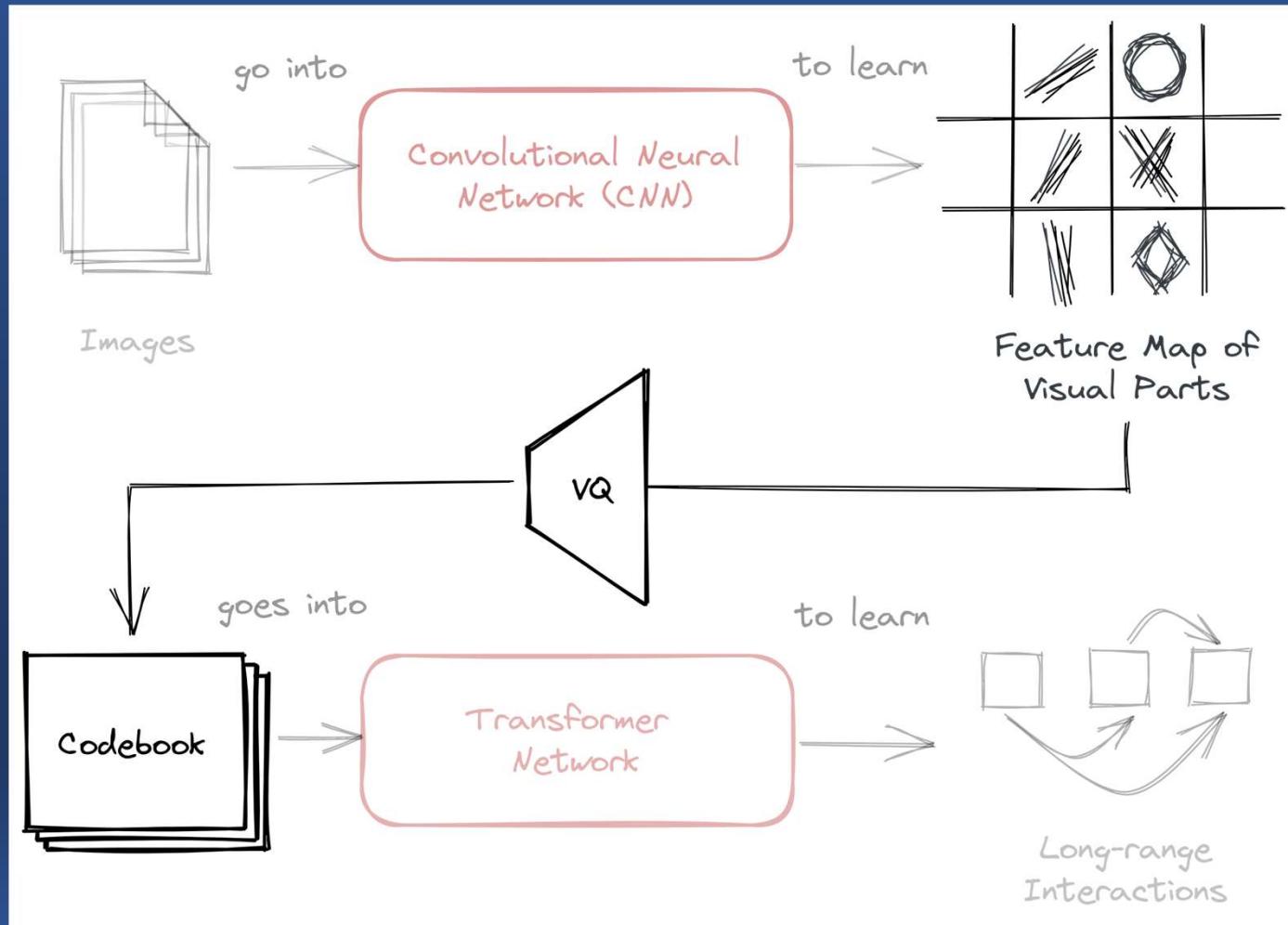
Puede realizar predicciones en datasets que nunca ha visto! (zero-shot prediction)





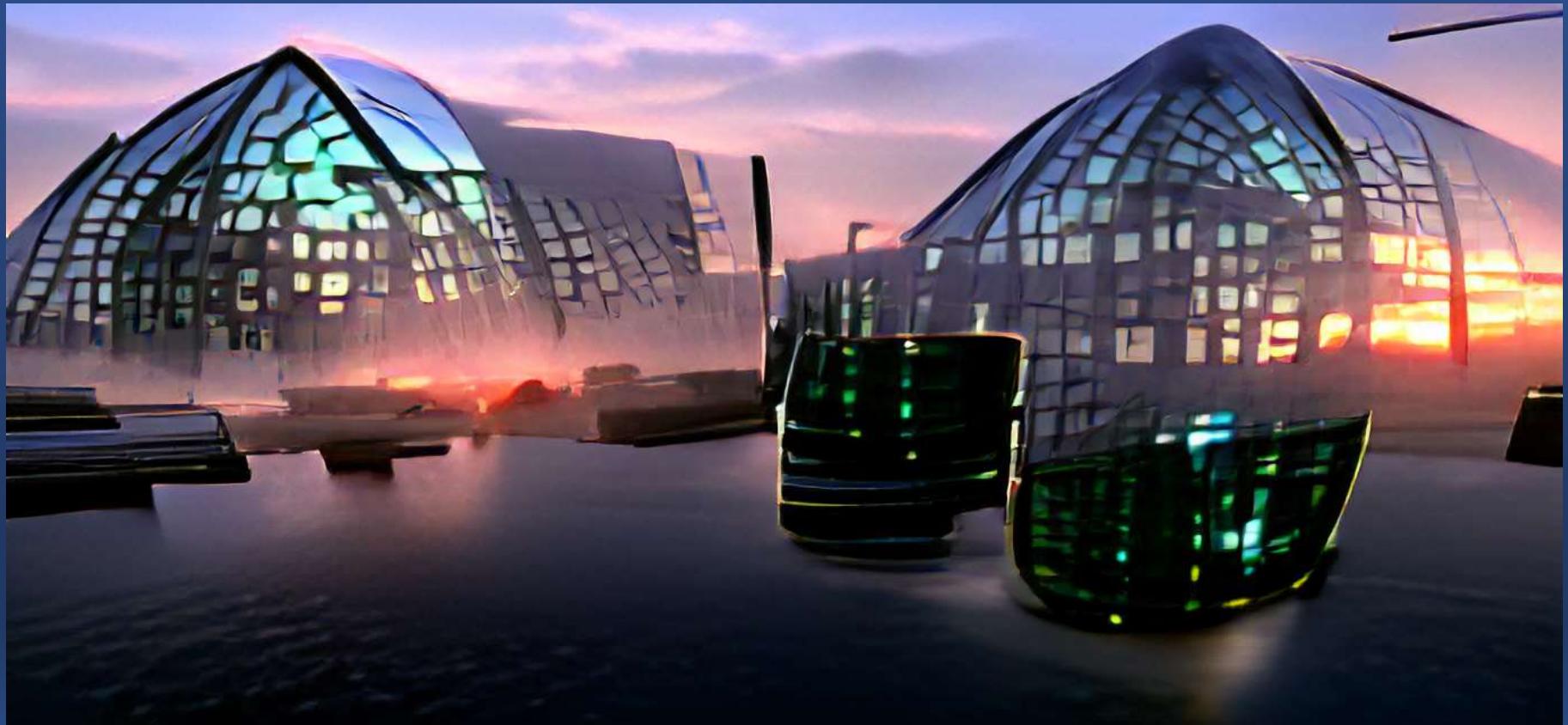
## VQGAN (2021)

Es capaz de usar transformers para aprender las interacciones a larga distancia



## VQGAN + CLIP

CLIP guía la generación de VQGAN para minimizar la diferencia entre su resultado y el texto



Ciudad futurística geométrica hecha de cristal, al amanecer

# Dall-e (2021)

Capaz incluso de hacer un razonamiento visual

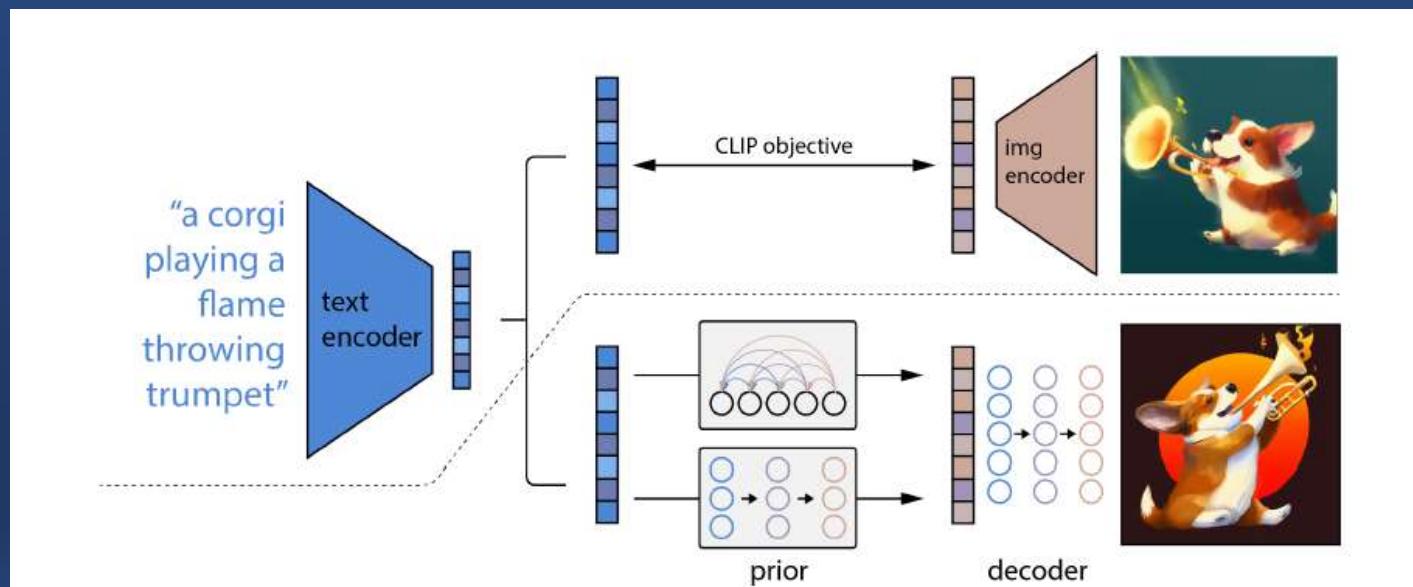


Una foto de la comida de China

## Dall-e 2: características

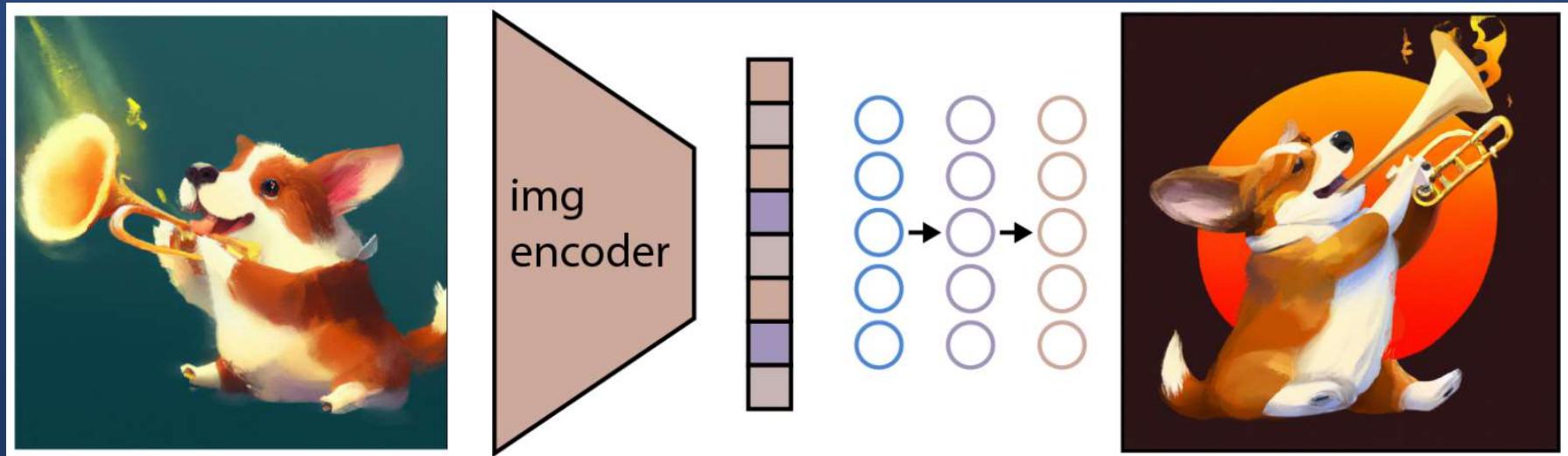
3.500 millones de parámetros (sorprendentemente, Dall-e tenía 12.000 millones)

Basado en CLIP, usa un modelo llamado unCLIP, que transforma una representación en una imagen y un modelo “prior” que busca las relaciones lógicas en la imagen



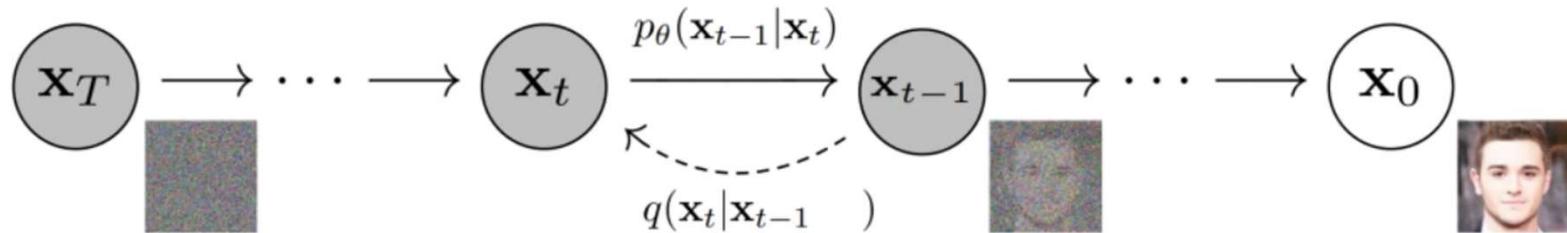
## Dall-e 2: características

Se crea un modelo que es capaz de invertir la codificación de CLIP, de forma que se obtengan imágenes con características similares a partir de los encodings the CLIP. Este modelo se llama GLIDE



## Dall-e 2: características

Un r  
añad

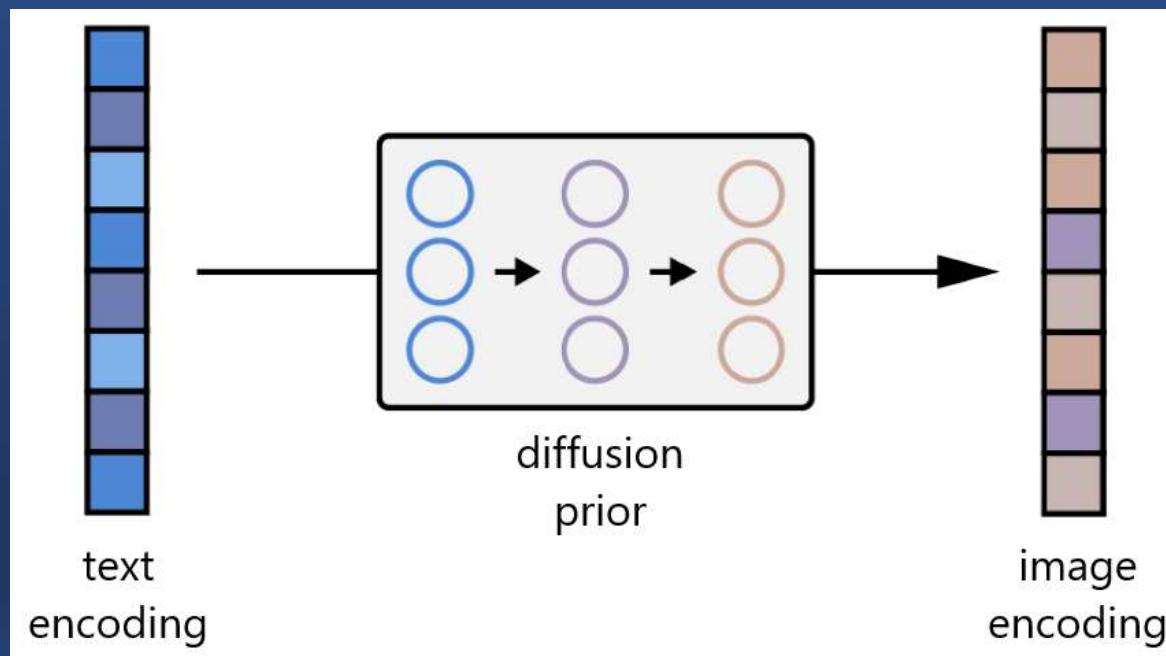


Adding textual  
conditioning to the  
diffusion process

La diferencia es que aquí se añade la condición de respetar los embeddings del texto, a través del transformer

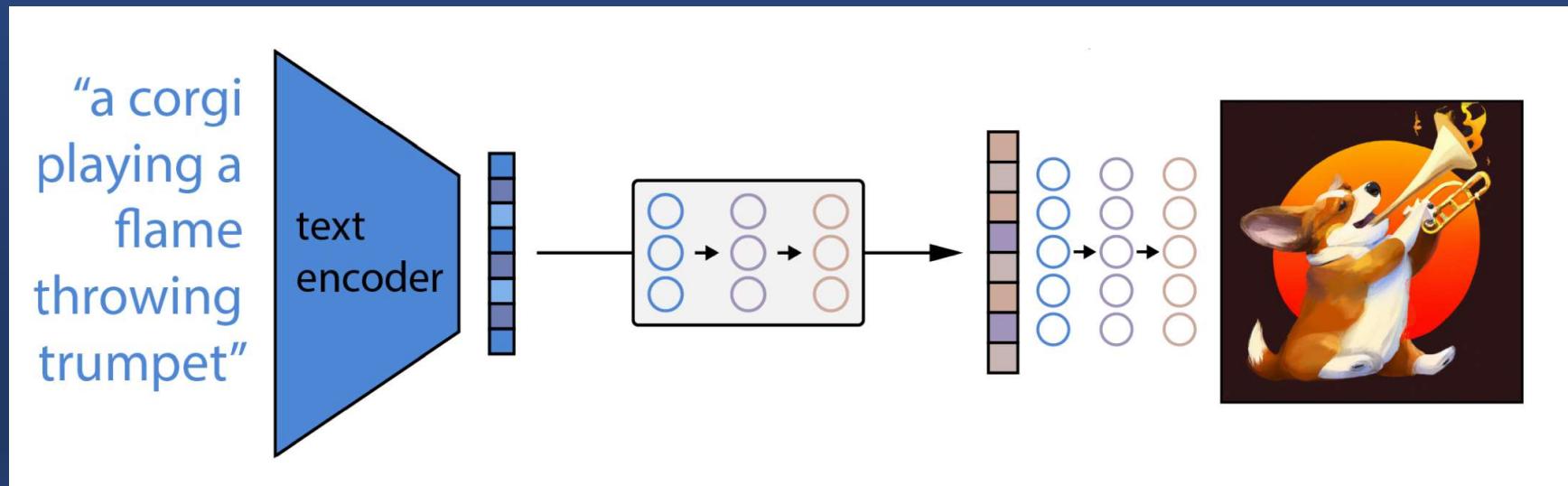
## Dall-e 2: características

Este modelo de diffusion se encargará de transformar los encodings de texto en encodings de imagen.



## Dall-e 2: características

El proceso se divide en 3 partes: generación de los encodings de texto con CLIP, diffusion del encoding de texto a un encoding de imagen similar y modelo de difusión inversa para generar la imagen a partir del encoding

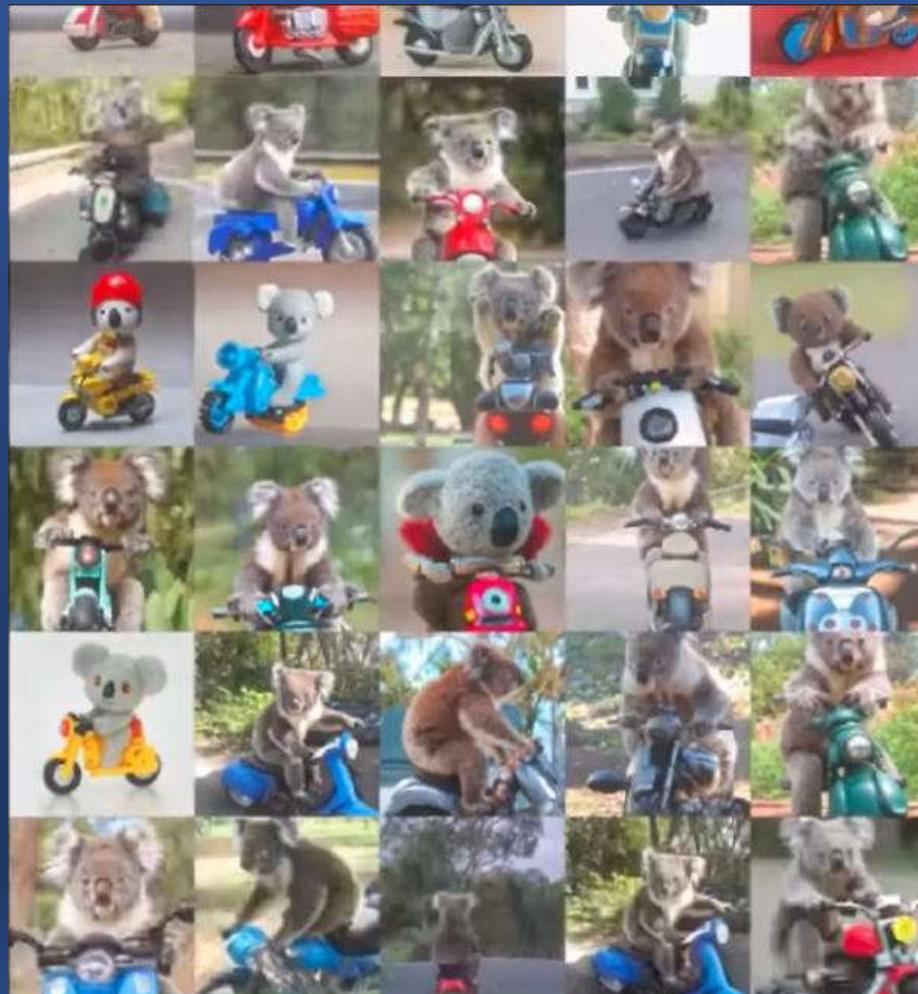


## Dall-e 2 características

De forma sencilla:

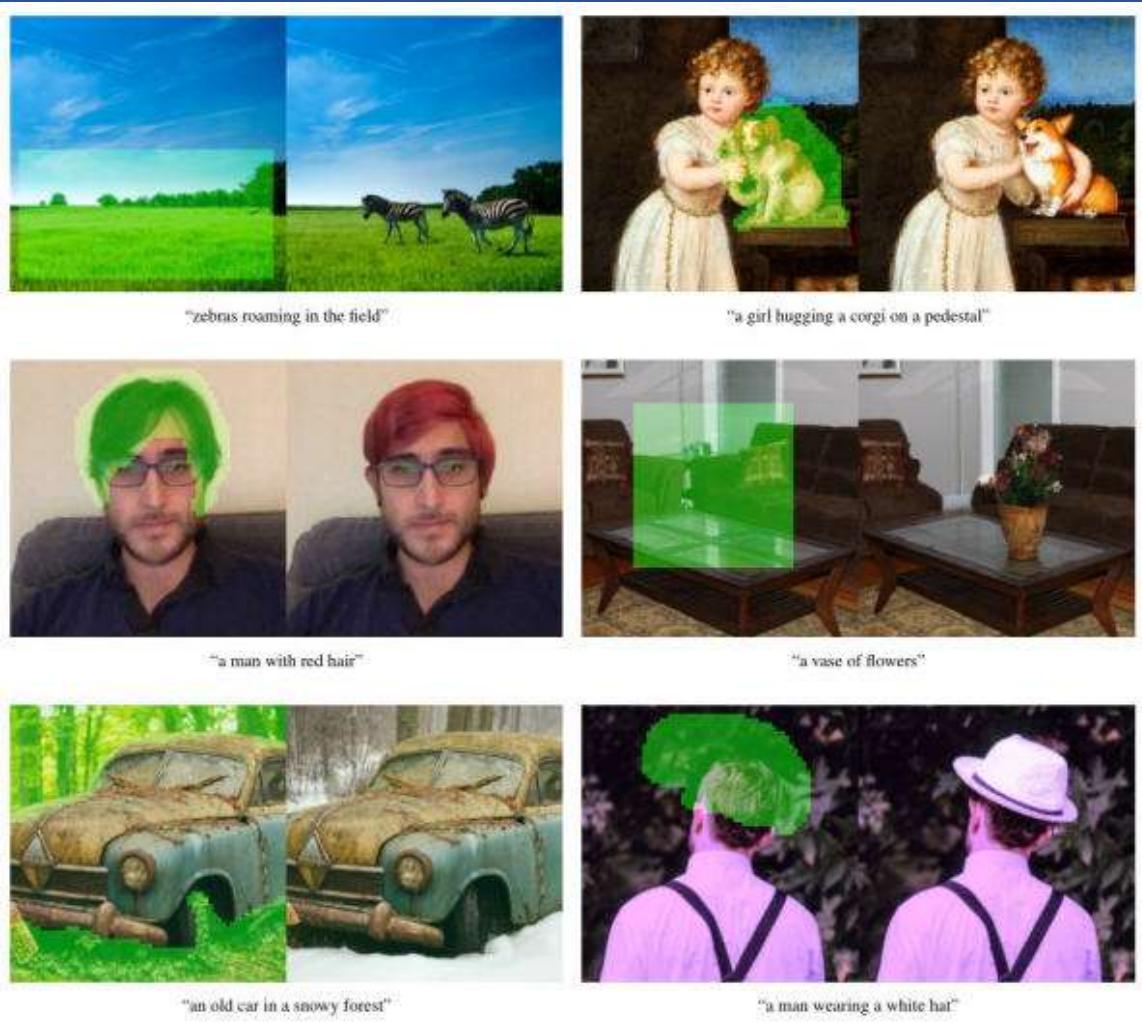
- Crea una representación de la imagen a partir de un texto (como nosotros hacemos al dibujar, en realidad, solo que la representación mental es muy diferente de un embedding)
- Transforma esta representación en una imagen (para eso sirve unCLIP). Se pueden generar muchas variaciones
- Se compara la imagen generada con el embedding
- Fue entrenado con un dataset de 650 millones de pares imagen-descripción

## Dall-e 2 multiple generation



Koalas riding motorcycles

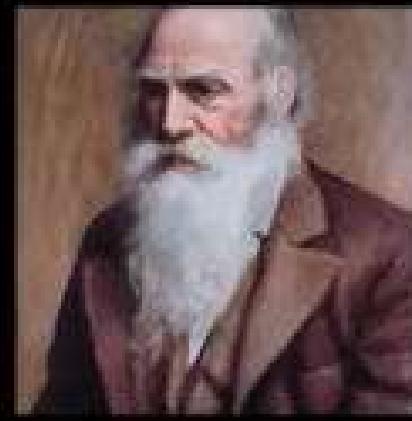
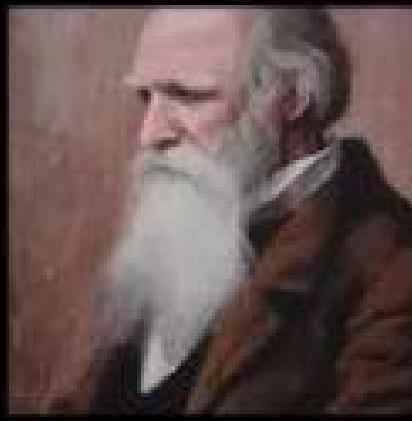
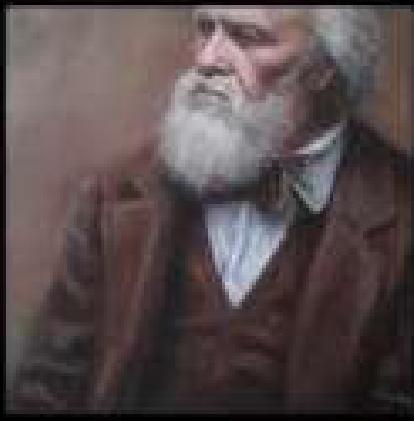
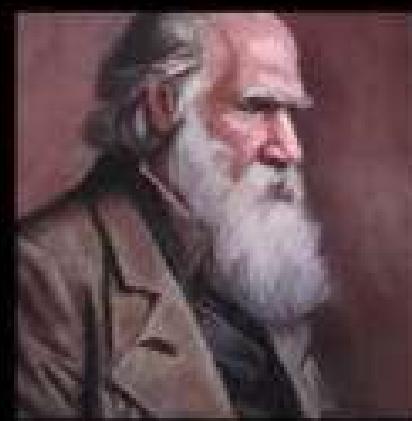
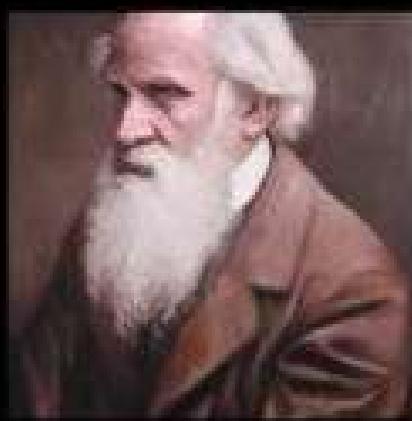
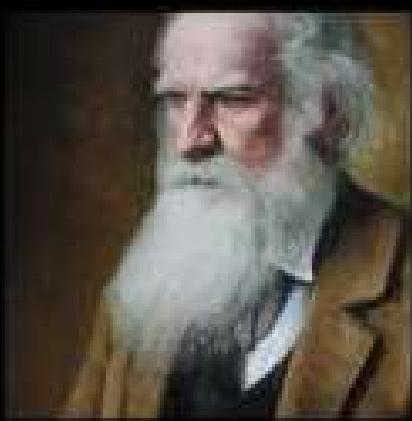
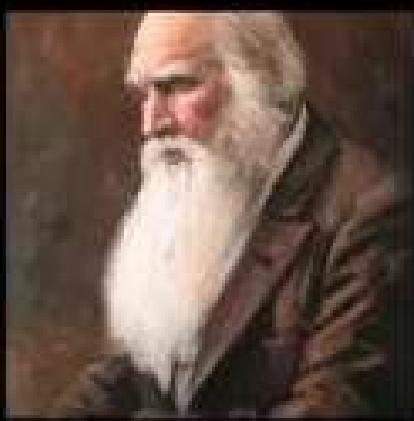
## Dall-e 2 Inpainting



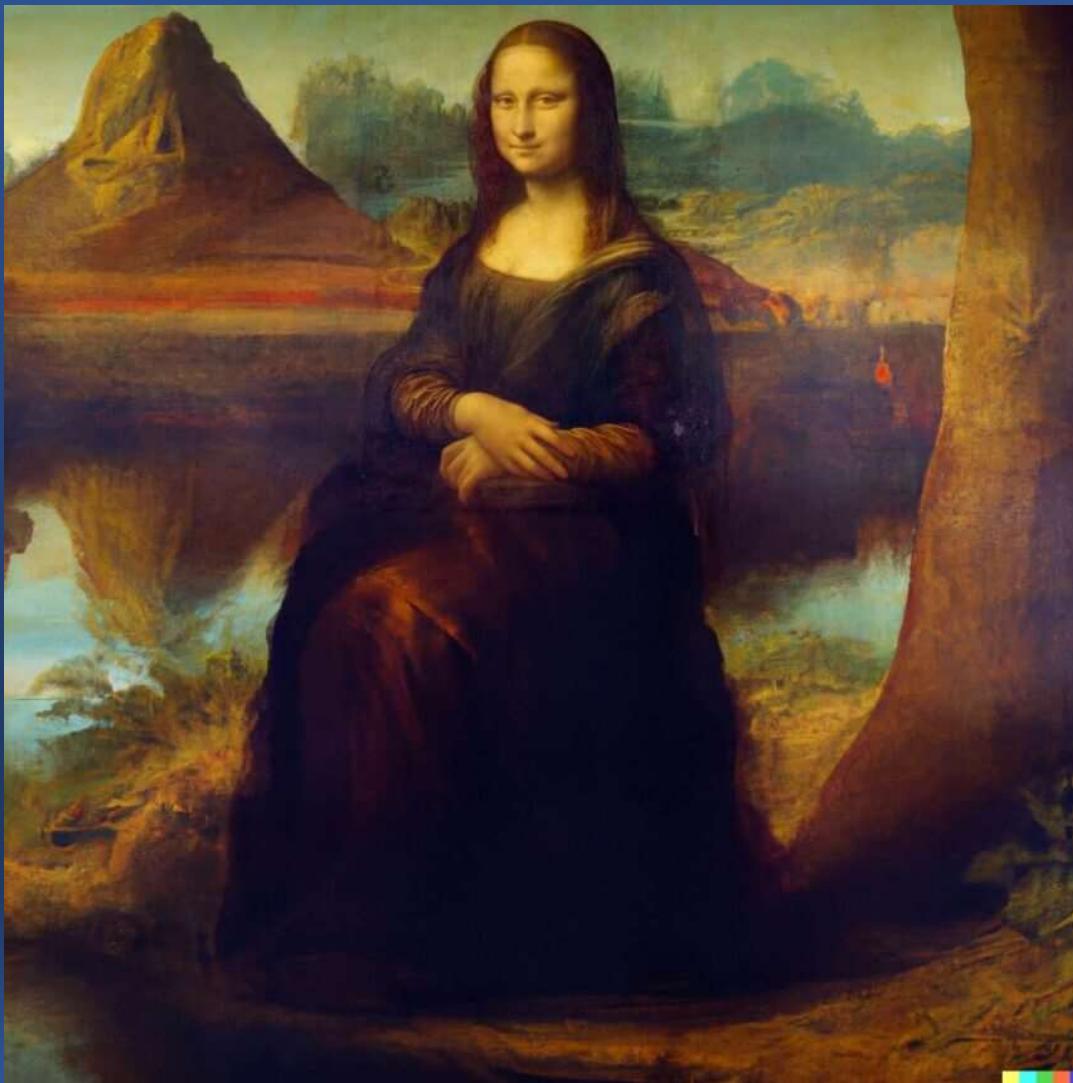
## Dall-e 2 Inpainting



## Dall-e 2 Variaciones de una imagen



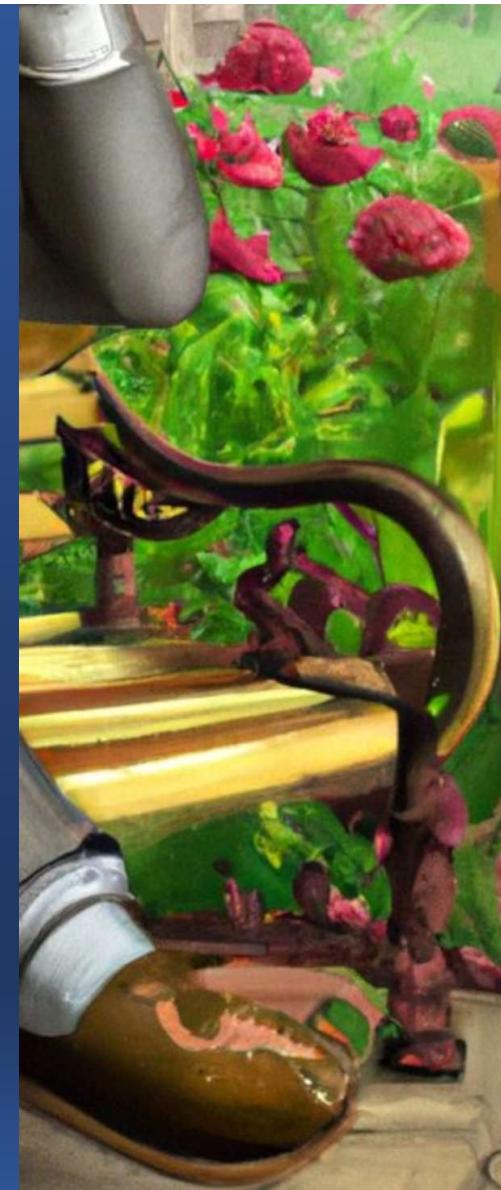
## Dall-e 2 Extensión de una imagen



## Dall-e 2 Errores



A rabbit detective sitting on a park bench and reading a newspaper in a Victorian setting



## Dall-e 2 Ejemplos



Pintura de Rafael de la Madonna y el Niño, comiendo pizza

## Dall-e 2 Ejemplos



Pintura medieval representando que la conexión WiFi no funciona

## Dall-e 2 Ejemplos

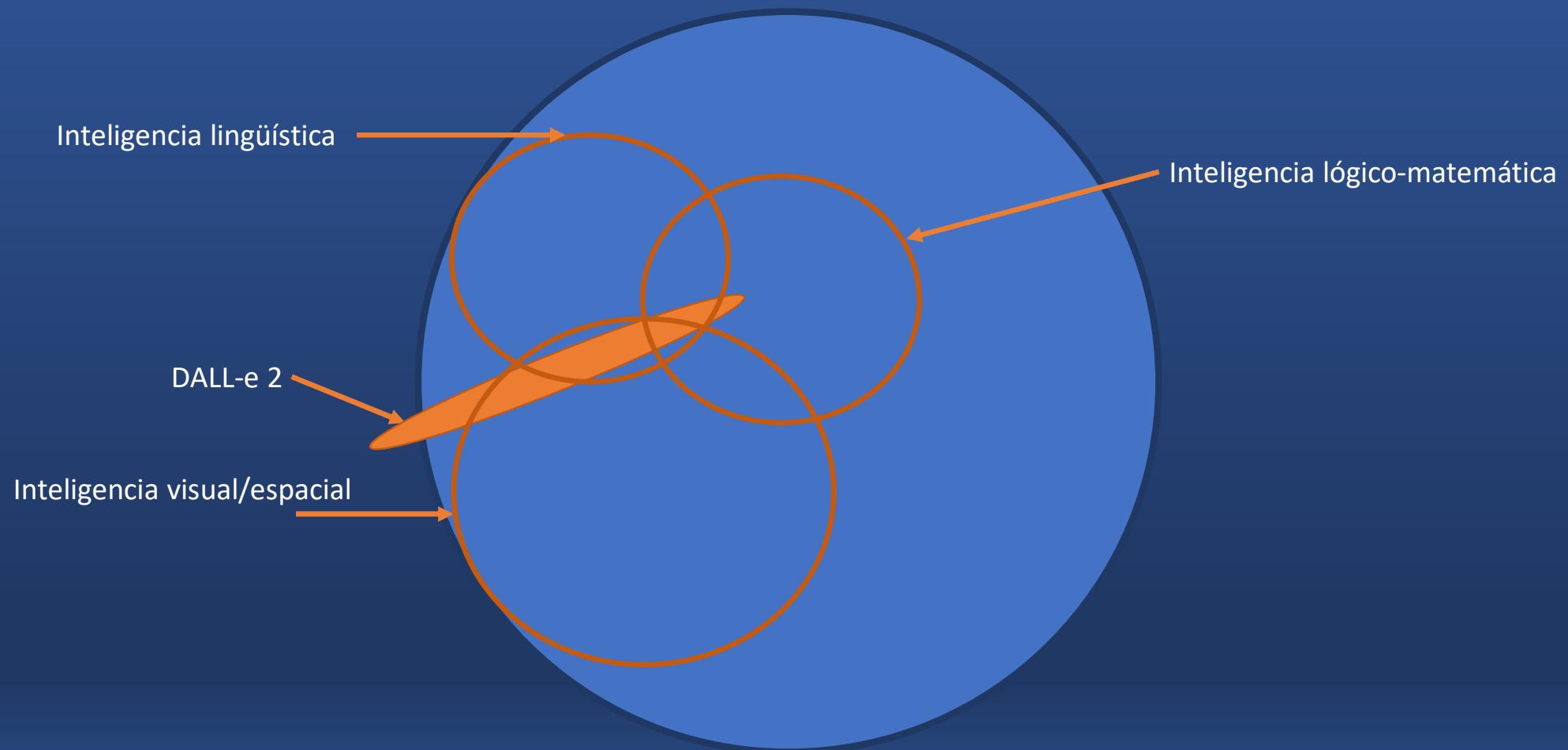


Osos de peluche trabajando en investigación en IA en la luna en los 80

## Dall-e 2 Ejemplos



Una vaca con una cazadora vaquera grabando un éxito de hip-hop en un estudio



## Conclusiones

CLIP, Dall-e y dall-e 2 son los primeros modelos que son capaces de realizar realmente un razonamiento en varios ámbitos de la inteligencia.

Si hay suficientes datos de entrenamiento, parece factible superar las capacidades humanas en muchos aspectos.

No existe una red neuronal de propósito general y estamos muy lejos de ella, pero poco a poco las redes se van generalizando en sus funciones, y en los últimos años incluso son capaces de manejar dominios diferentes de la inteligencia

Sólo las grandes compañías pueden permitirse entrenar redes que superen a las mejores en cada especialidad actualmente

Estoy agradecido por su atención

DALL·E mini  
AI model generating images from any prompt!

I am grateful for your attention Run

The interface shows a text input field with the prompt "I am grateful for your attention" and a "Run" button. Below the input are nine generated images arranged in a 3x3 grid. The images feature various scenes of people interacting with a bright, glowing sun or sunset, such as a silhouette of a person walking away, a person meditating on a beach, a woman sitting on a dock, a girl standing on a rocky cliff, a mermaid on a rock, and hands reaching up towards the sun.

## ¿Alguna pregunta?

DALL·E mini

AI model generating images from any prompt!

do you have any question? Run

The image displays a 3x3 grid of nine distinct 3D-rendered question marks. The colors and styles vary across the grid:

- Row 1: An orange question mark with a small orange dot at the bottom left; a yellow question mark with a small yellow dot at the bottom right; a red question mark with a small blue dot at the top right.
- Row 2: A red question mark with a small red dot at the bottom left; a black question mark with a small black dot at the bottom right; a blue question mark with a small blue dot at the bottom right.
- Row 3: A blue question mark with a small yellow dot at the top right; a large black question mark with a small black dot at the bottom right; a white question mark with a small white dot at the bottom right.

# No dudéis en preguntar

