



presentado en

Pista 1 ▪ Pista 2

Introducción a la IA generativa

Qué veremos en este módulo

- En qué se diferencia la IA generativa de otros tipos de IA
- Qué se necesita para crear un modelo de IA generativa
- La diferencia entre generadores de texto e imágenes
- Principales retos de la IA generativa

Recordatorio: no te preocupes.

-  Si es la primera vez que profundizas en los detalles técnicos, aprovecha para **familiarizarte** con los términos y conceptos.
-  Si ya conoce algunos de estos detalles, utilízalos para **aclarar** tus conocimientos.
-  Considera esta presentación como un **recurso**.
¡Consúltala cuando lo necesite!

IA generativa

Inteligencia
artificial

(datos)

Aprendizaje
automático

(big data)

Aprendizaje
profundo

(datos masivos, gran
capacidad de cálculo)

IA
generativa



Inteligencia
artificial

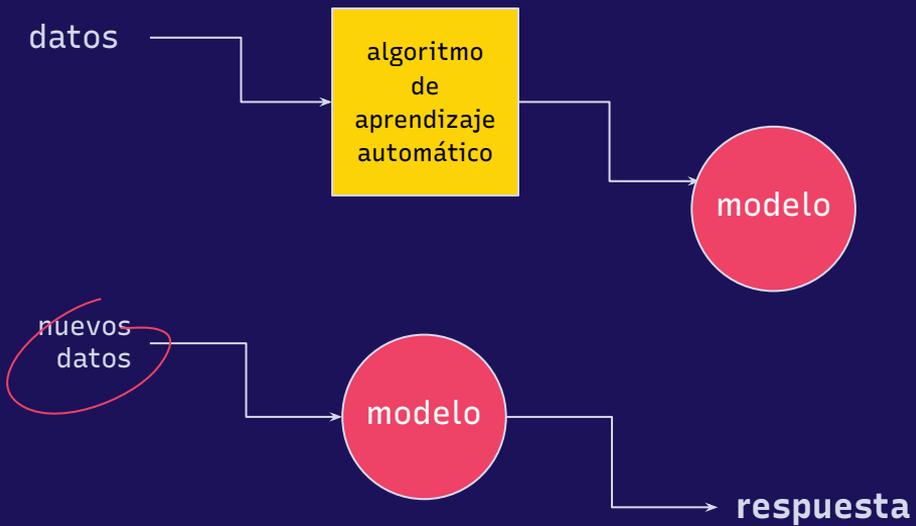
Aprendizaje
automático

Aprendizaje
profundo

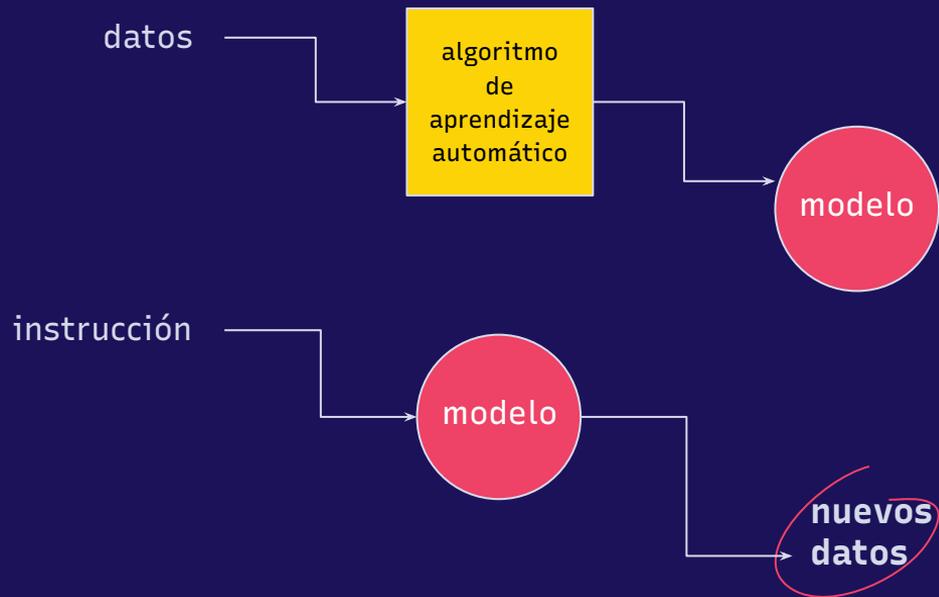
IA
generativa

IA
predictiva

IA predictiva



IA generativa



La era actual* de la
IA generativa se
basa en los
transformadores.

* Tenga en cuenta que esto podría cambiar algún día.

¿Qué son los transformadores?

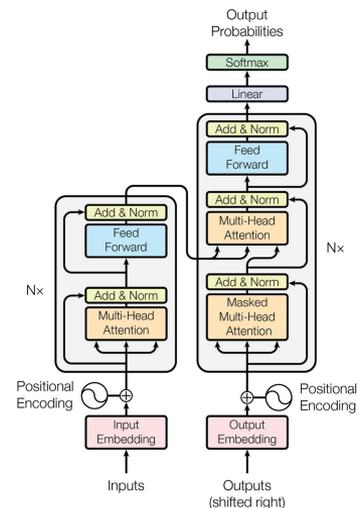
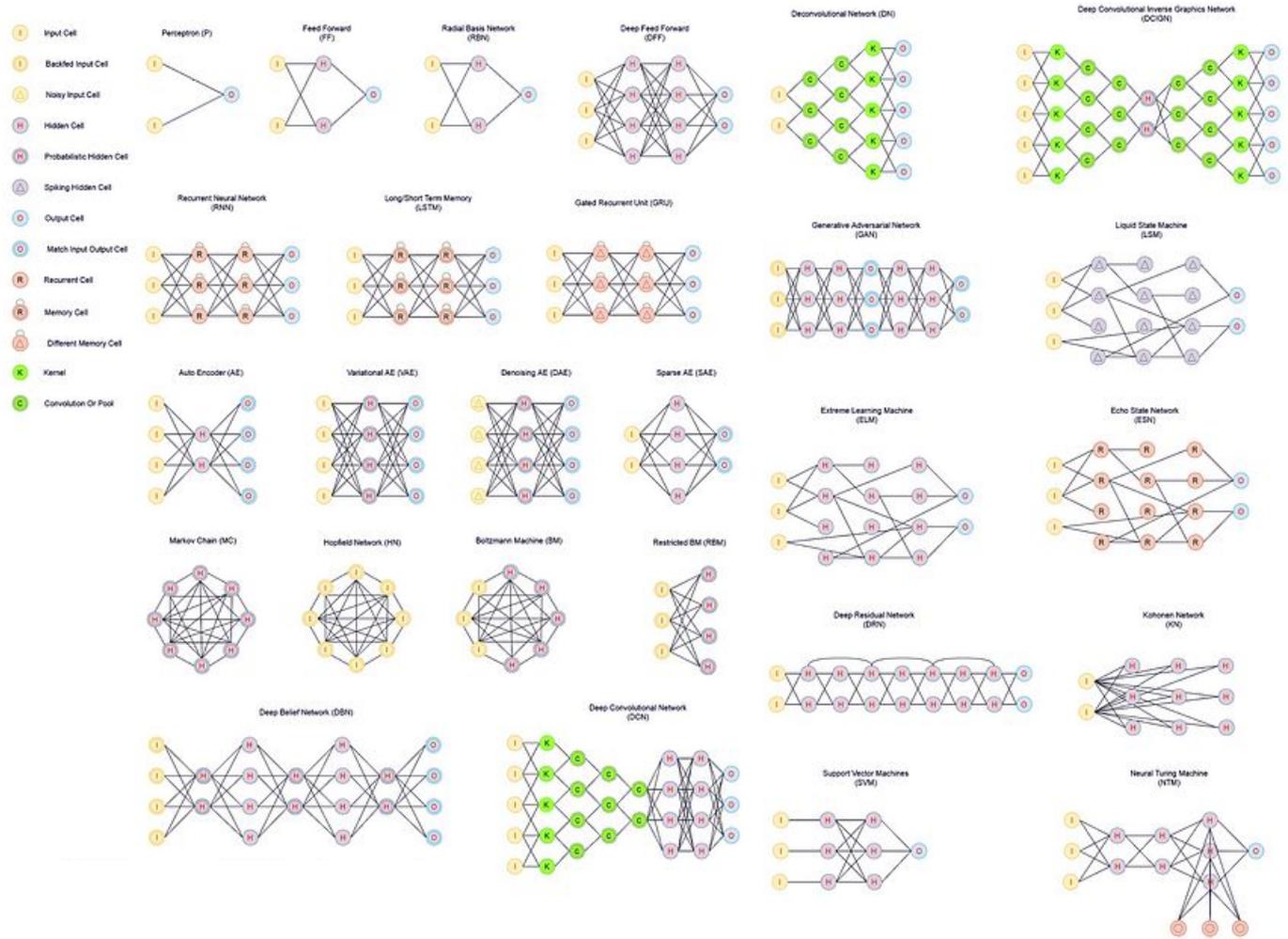


Figure 1: The Transformer - model architecture.



¿Qué hace que los transformadores sean tan especiales?

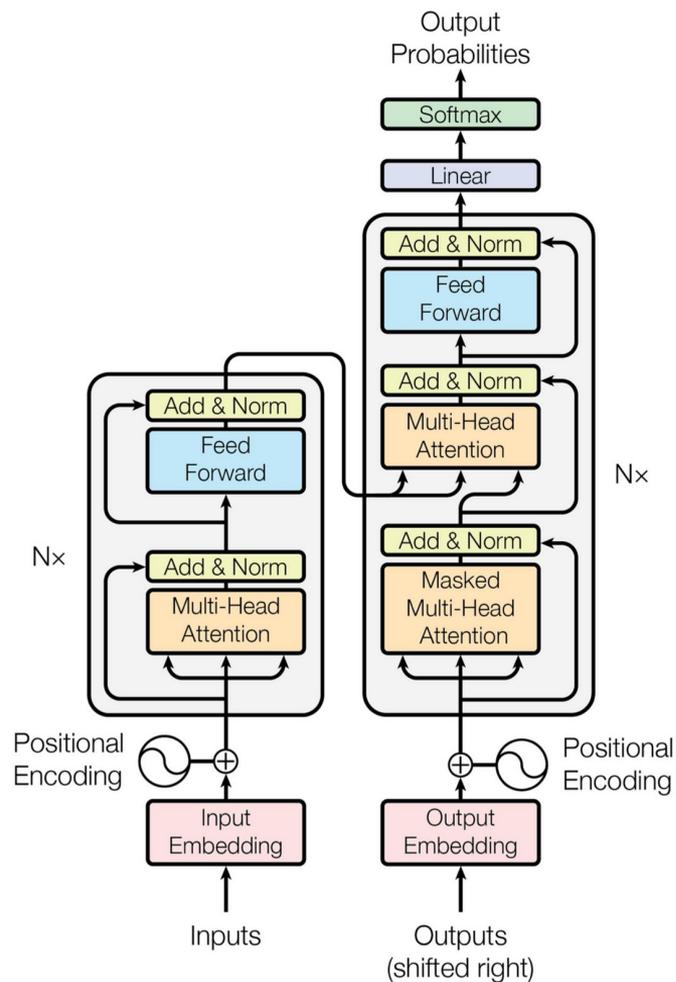


Figure 1: The Transformer - model architecture.

Son buenos para encontrar patrones de largo alcance en secuencias de datos.

Shall I compare thee to a summer's day?
Thou art more lovely and more temperate.
Rough winds do shake the darling buds of May,
And summer's lease hath all too short a date.
Sometime too hot the eye of heaven shines,
And often is his gold complexion dimmed;
And every fair from fair sometime declines,
By chance, or nature's changing course, untrimmed;
But thy eternal summer shall not fade,
Nor lose possession of that fair thou ow'st,
Nor shall death brag thou wand'rest in his shade,
When in eternal lines to Time thou grow'st.
So long as men can breathe, or eyes can see,
So long lives this, and this gives life to thee.

Corto alcance

Shall I compare thee to a summer's day?
Thou art more lovely and more temperate.
Rough winds do shake the darling buds of May,
And summer's lease hath all too short a date.
Sometime too hot the eye of heaven shines,
And often is his gold complexion dimmed;
And every fair from fair sometime declines,
By chance, or nature's changing course, untrimmed;
But thy eternal summer shall not fade,
Nor lose possession of that fair thou ow'st,
Nor shall death brag thou wand'rest in his shade,
When in eternal lines to Time thou grow'st.
So long as men can breathe, or eyes can see,
So long lives this, and this gives life to thee.

Largo alcance

NEW! Play Our Addictive New Puzzle Games! Today's Theme: Harry Potter • 1 hour ago

TECH

The New iOS8 Predictive Text Makes For Terrible Slam Poetry

We texted people using only the predictive text suggested words.

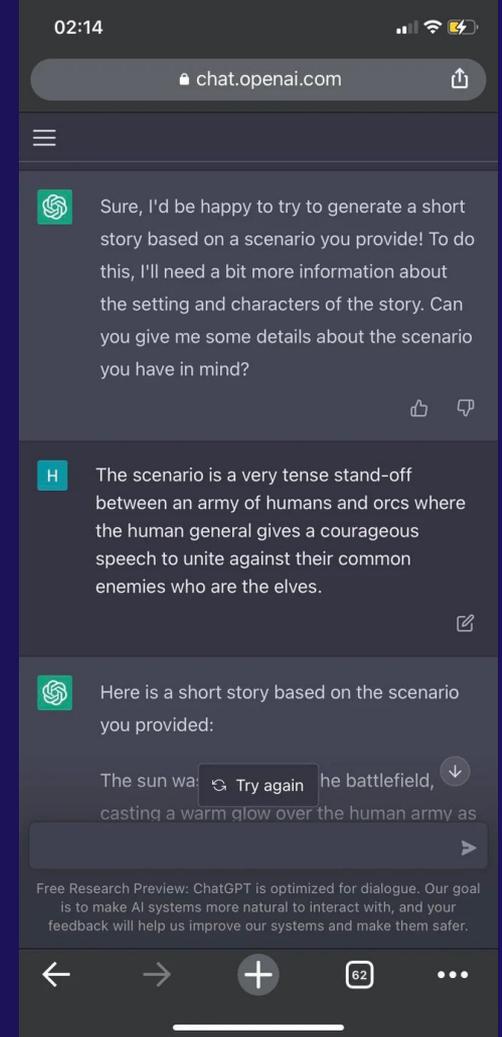


Katie Notopoulos
BuzzFeed News Reporter

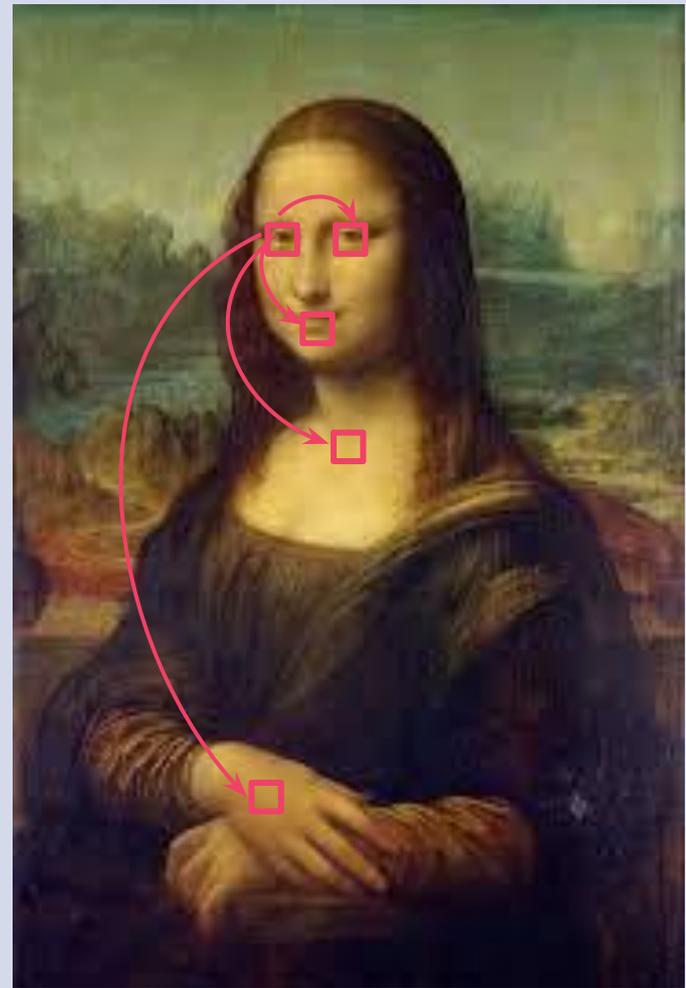
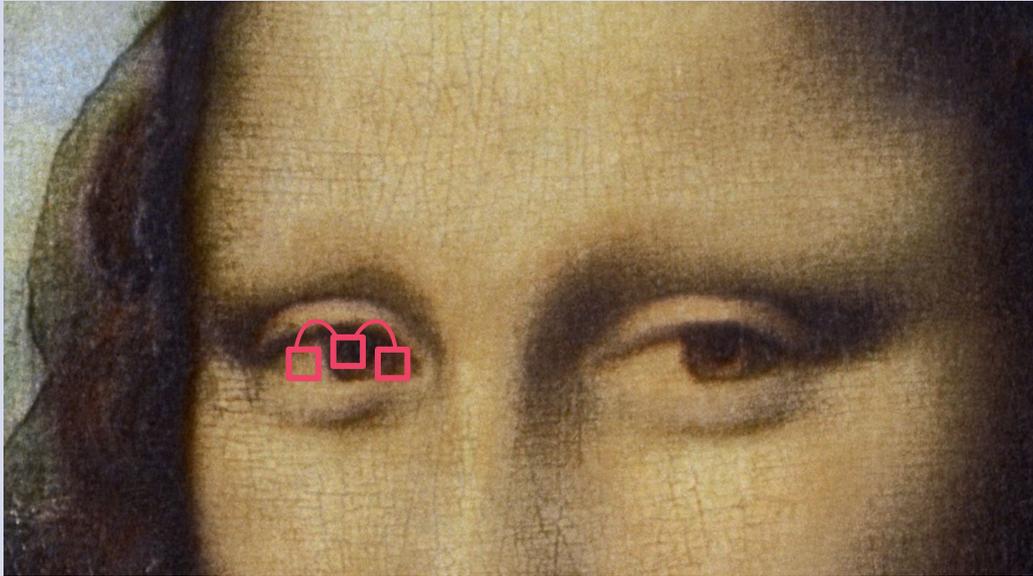
Posted on September 26, 2014 at 2:40 am

[View 19 comments](#)

One of the new features in iOS8 for iPhone is predictive text. This means when you're typing, it suggest what word it thinks you're gonna say next.



Son buenos para encontrar patrones de largo alcance en secuencias de datos.



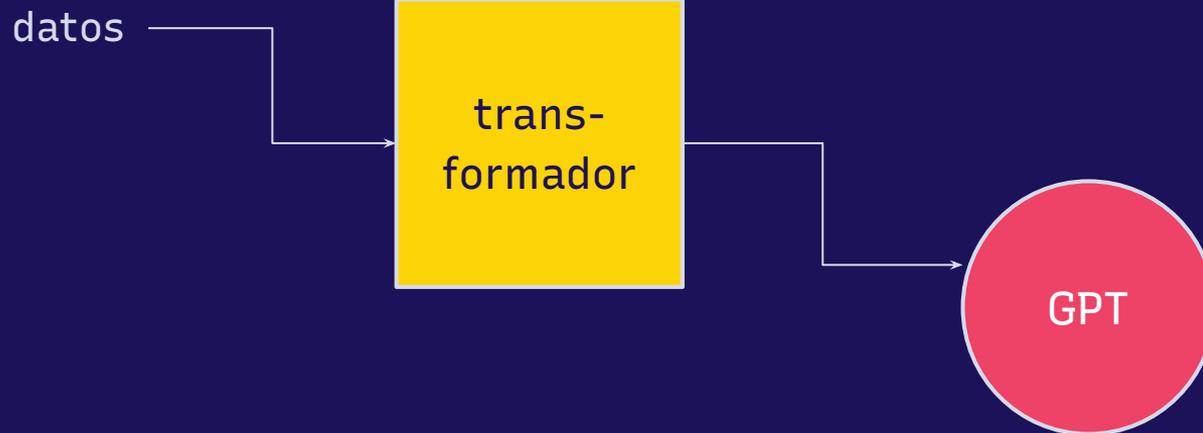
ChatGPT

G: Generativo

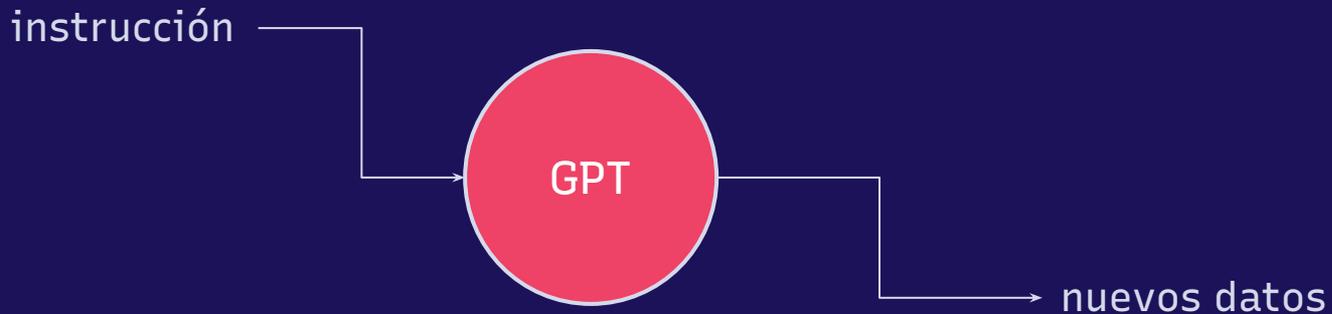
P: Preentrenado

T: Transformador

ENTRENAMIENTO



INFERENCIA



ENTRENAMIENTO

PREENTRENAMIENTO

AJUSTE FINO

RLHF

corpus gigante,
también conocido como:
Internet

trans-
formador

GPT
base

GPT
ajustado

GPT «más
seguro»

datos adaptados o
específicos
para cada tarea

entrada
humana

ENTRENAMIENTO

PREENTRENAMIENTO

AJUSTE FINO

RLHF

corpus gigante,
también conocido como:
Internet

trans-
formador

GPT
base

GPT
ajustado

GPT «más
seguro»

datos
adaptados o
específicos
para cada tarea

entrada
humana

ENTRENAMIENTO

PREENTRENAMIENTO

AJUSTE FINO

RLHF

corpus gigante,
también conocido como:
Internet

trans-
formador

GPT
base

GPT
ajustado

GPT «más
seguro»

datos adaptados o
específicos para cada tarea

entrada humana

ENTRENAMIENTO

PREENTRENAMIENTO

AJUSTE FINO

RLHF

corpus gigante,
también conocido como:
Internet

trans-
formador

GPT
base

GPT
ajustado

GPT «más
seguro»

datos
adaptados o
específicos
para cada tarea

entrada
humana

Aprendizaje no supervisado

ENTRENAMIENTO

PREENTRENAMIENTO

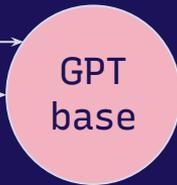
Aprendizaje supervisado

AJUSTE FINO

Aprendizaje por refuerzo

RLHF

corpus gigante, también conocido como: Internet



datos adaptados o específicos para cada tarea

entrada humana

INFERENCIA

MODERACIÓN

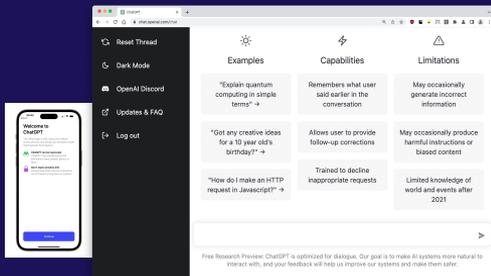
instrucción
(prompt)

nuevos
datos

ENRIQUECIMIENTO

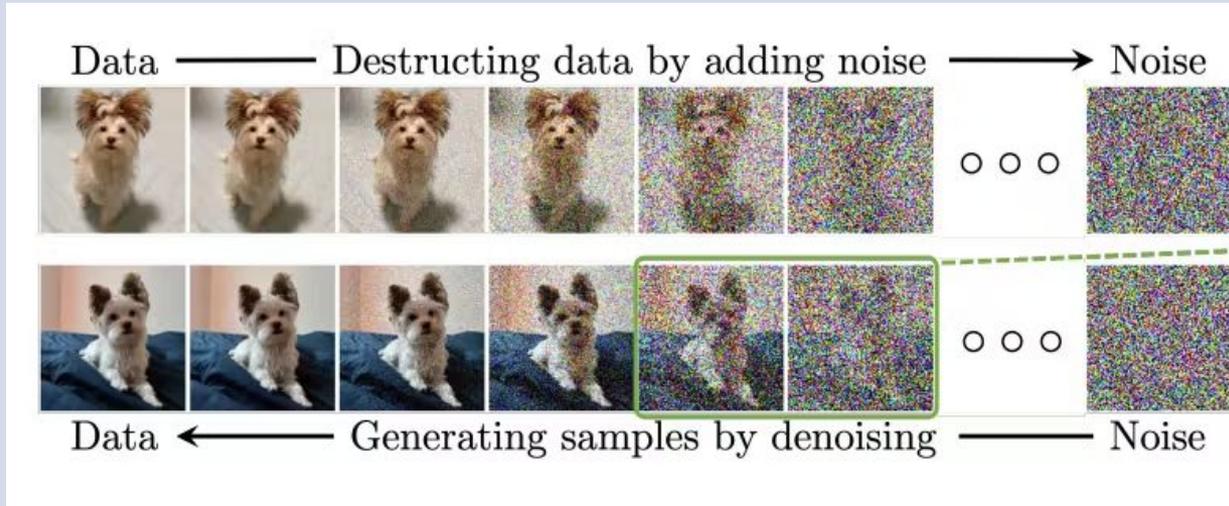
GPT
«más
seguro»

INTERFAZ



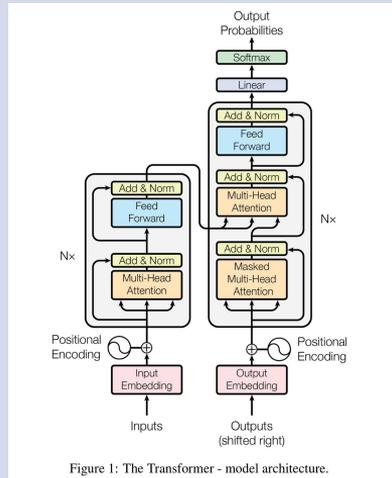
Los generadores de imágenes son muy similares, con una diferencia conceptual principal.

Utilizan difusión.



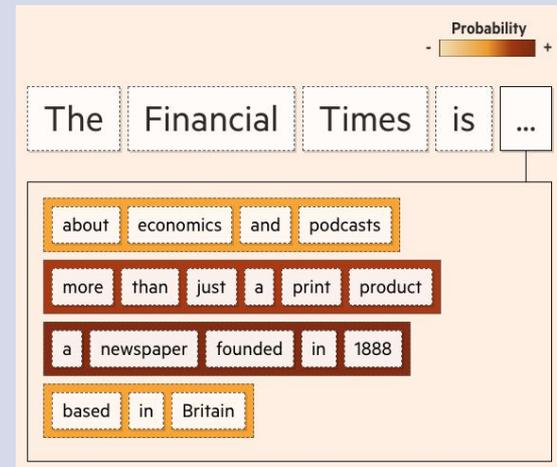
La difusión es un objetivo.

Arquitectura



La arquitectura utilizada para entrenar un LLM es: **el transformador.**

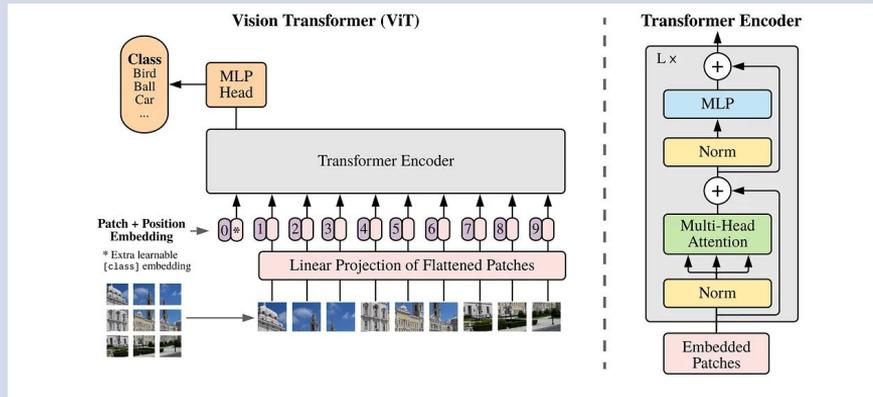
Objetivo



El objetivo utilizado para entrenar un LLM es: **predecir la(s) siguiente(s) palabra(s).**

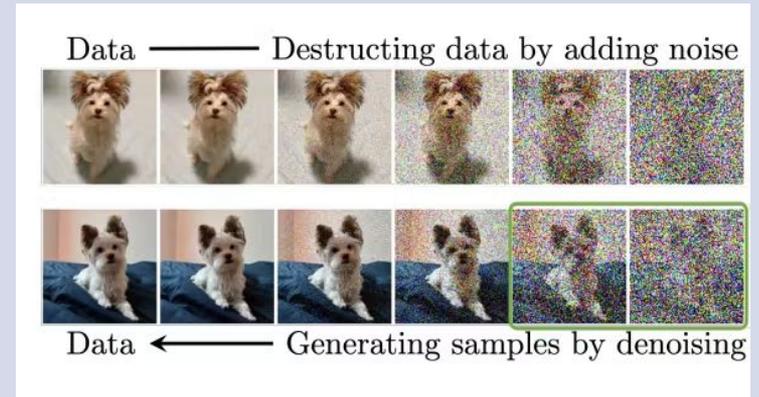
La difusión es un objetivo.

Arquitectura



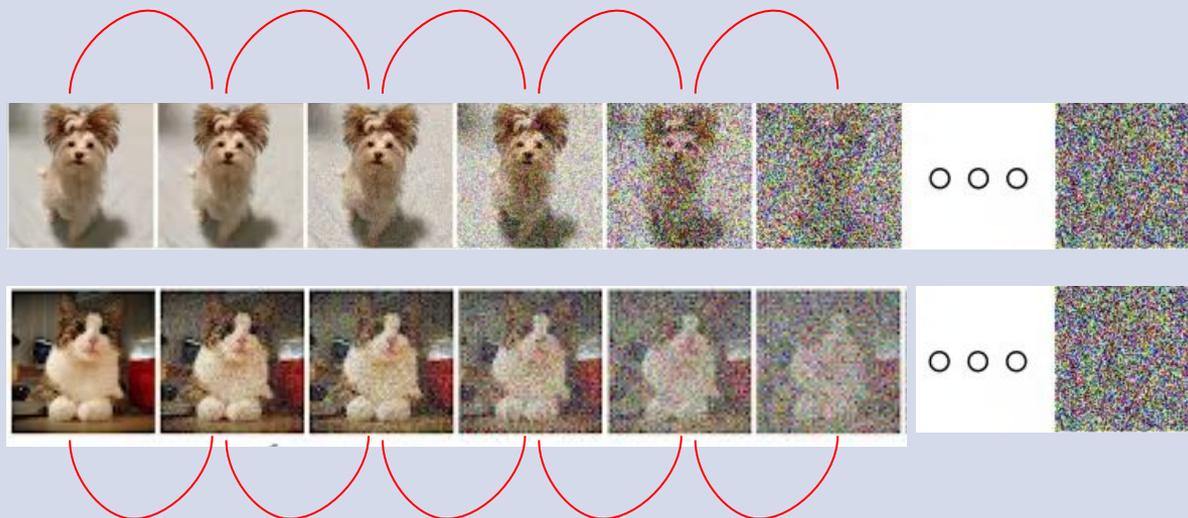
La arquitectura utilizada para entrenar un generador de imágenes es: **el transformador de visión**.

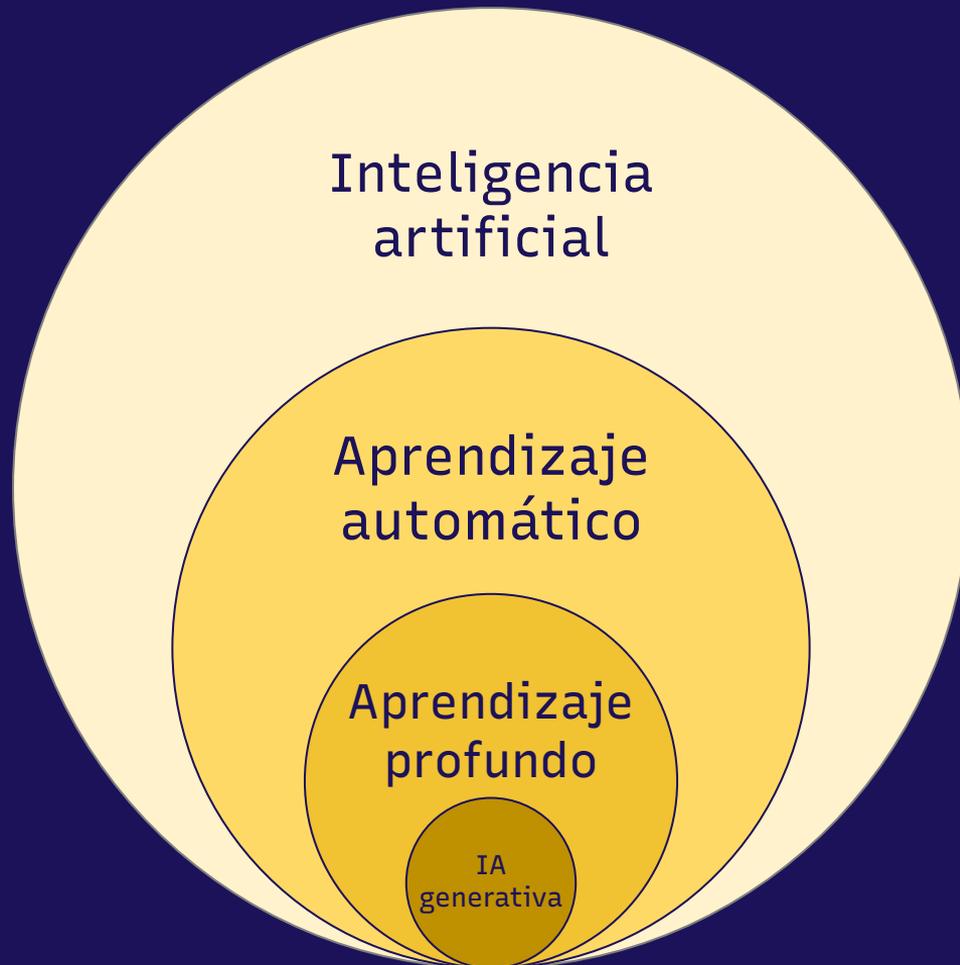
Objetivo



El objetivo utilizado para entrenar un generador de imágenes es: **la difusión, o recuperar una imagen dañada**.

La difusión consiste en encontrar la señal en el ruido.





ENTRENAMIENTO

PREENTRENAMIENTO

corpus gigante, también conocido como: Internet

transformador

AJUSTE FINO

GPT base

RLHF

GPT ajustado

GPT «más seguro»

datos adaptados o específicos para cada tarea

entrada humana

Nadie sabe lo que hay aquí.

desinformación
errores fácticos
conceptos erróneos
teorías conspirativas
especulaciones
leyendas urbanas
relatos cortos
ciencia ficción
ficción de fans
tuits
diatribas
exageración
racismo
abuso



SIGN IN

SUBSCRIBE

[ABOUT](#) [RSS](#) [SUPPORT/FAQ](#) [PODCAST](#) [FOIA](#) [FORUM](#) [ARCHIVE](#) [MERCH](#) [ADVERTISE](#) [THANKS](#) [PRIVACY](#)[AI](#)

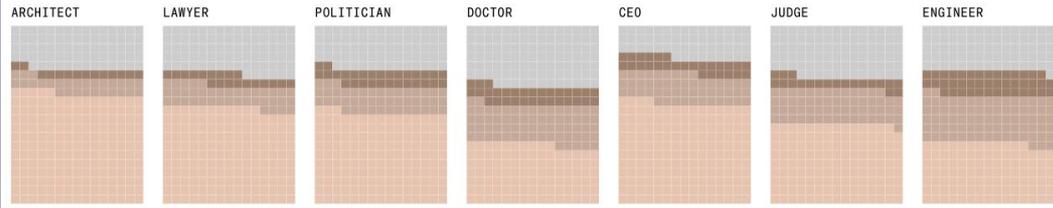
Largest Dataset Powering AI Images Removed After Discovery of **Child Sexual Abuse Material**

The model is a massive part of the AI-ecosystem, used by Stable Diffusion and other major generative AI products. The removal follows discoveries made by Stanford researchers, who found thousands instances of suspected child sexual abuse material in the dataset.

 SAMANTHA COLE · DEC 20, 2023 AT 7:00 AM

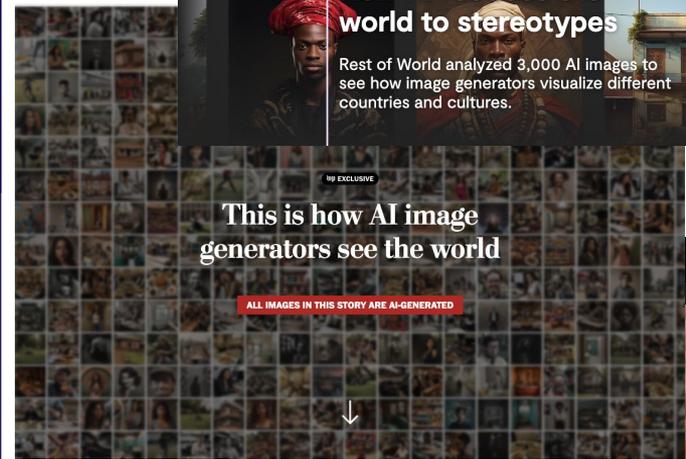
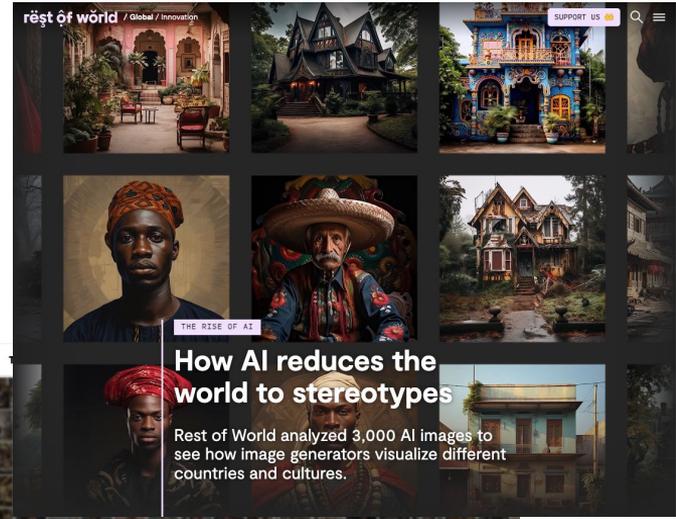
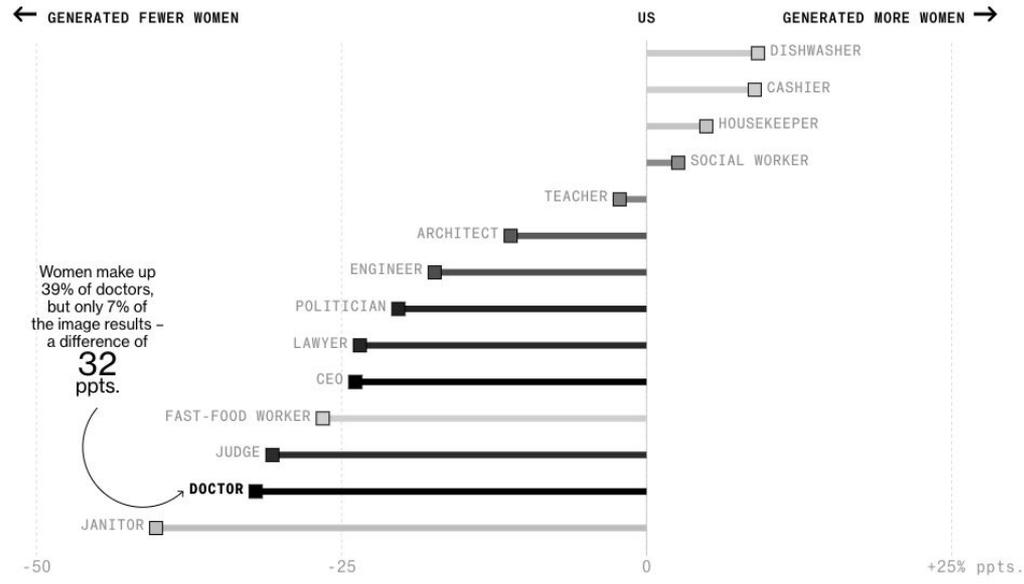
Lighter skin
I II III
Darker skin
IV V VI

High-paying occupations



Working Women Misrepresented Across the Board
Stable Diffusion results compared to US demographics for each occupation

Average US income in 2022
\$20K ————— \$242K



**HUMANS ARE BIASED.
GENERATIVE AI
IS EVEN WORSE**

Stable Diffusion's text-to-image model amplifies stereotypes about race and gender - here's why that matters

La IA generativa consiste en remezclar datos. NO se trata de recuperar información.

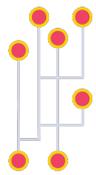


51%

**de las respuestas de ChatGPT, Copilot, Gemini y
Perplexity sobre las noticias eran INEXACTAS
-Investigación de la BBC**

A graphic for the AI Spotlight Series. It features the text "AI Spotlight Series" in a bold, pink, sans-serif font. To the left of the text is a vertical line of four circles: the top one is yellow, the second and third are light blue, and the bottom one is light blue. A horizontal line extends from the top yellow circle to the right, ending at the "AI" text.

AI Spotlight Series

A logo for The AI Accountability Network, consisting of a network of grey lines connecting several yellow and red circular nodes.

**The AI
Accountability
Network**

The Pulitzer Center logo, which is a stylized blue circle containing a white 'P' and 'C' intertwined.

Pulitzer Center