

*apresentado em*

Faixa 1 ▪ Faixa 2

# Introdução à IA generativa



# O que abordaremos neste módulo

- Como a IA generativa difere de outros tipos de IA
- O que é necessário para criar um modelo de IA generativa
- A diferença entre geradores de texto e de imagens
- Principais desafios da IA generativa



# Lembrete: não se preocupe.

 Se esta é a primeira vez que você se envolve com detalhes técnicos, use isso como uma forma de se **familiarizar** com termos e conceitos.

 Se você já conhece alguns desses detalhes, use isso como uma forma de melhorar sua compreensão.

 Trate este conjunto de slides como um recurso. Consulte-o sempre que necessário!



# IA generativa

Inteligência  
artificial

(dados)

Aprendizado  
de máquina

(big data)

Aprendizado  
profundo

(big data,  
computação em  
grande escala)

IA  
generativa



Inteligência  
artificial

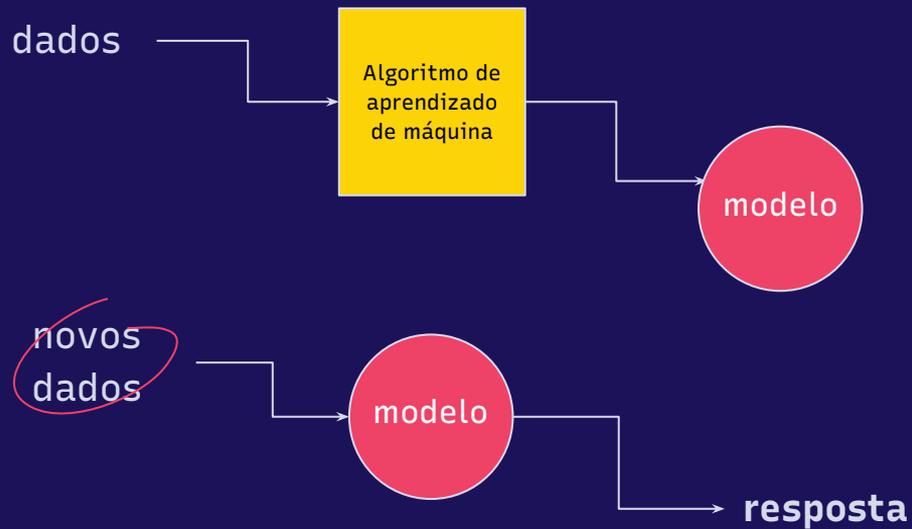
Aprendizado  
de máquina

Aprendizado  
profundo

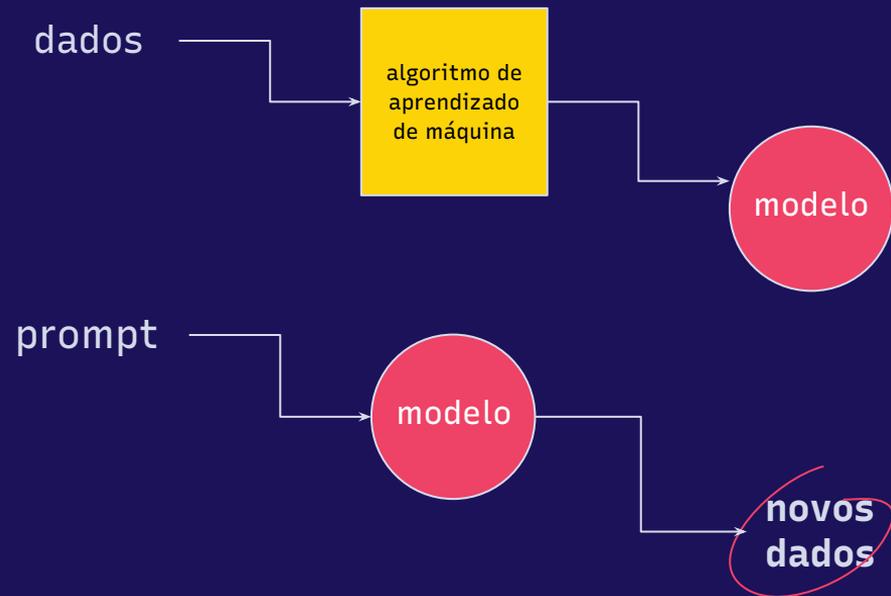
GenAI

IA preditiva

## IA preditiva



## IA generativa



A era atual\* da IA  
generativa é  
baseada em  
transformadores.

\* Lembre-se: isso pode mudar um dia!

# O que são transformadores?

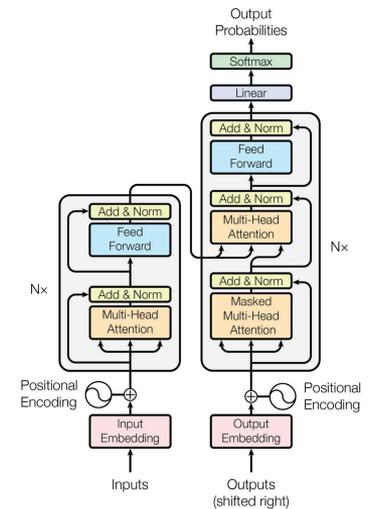
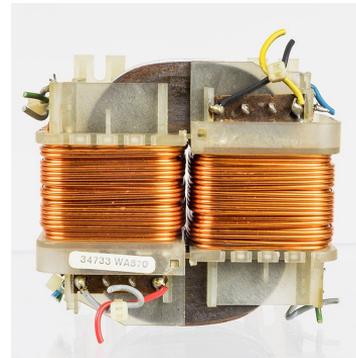
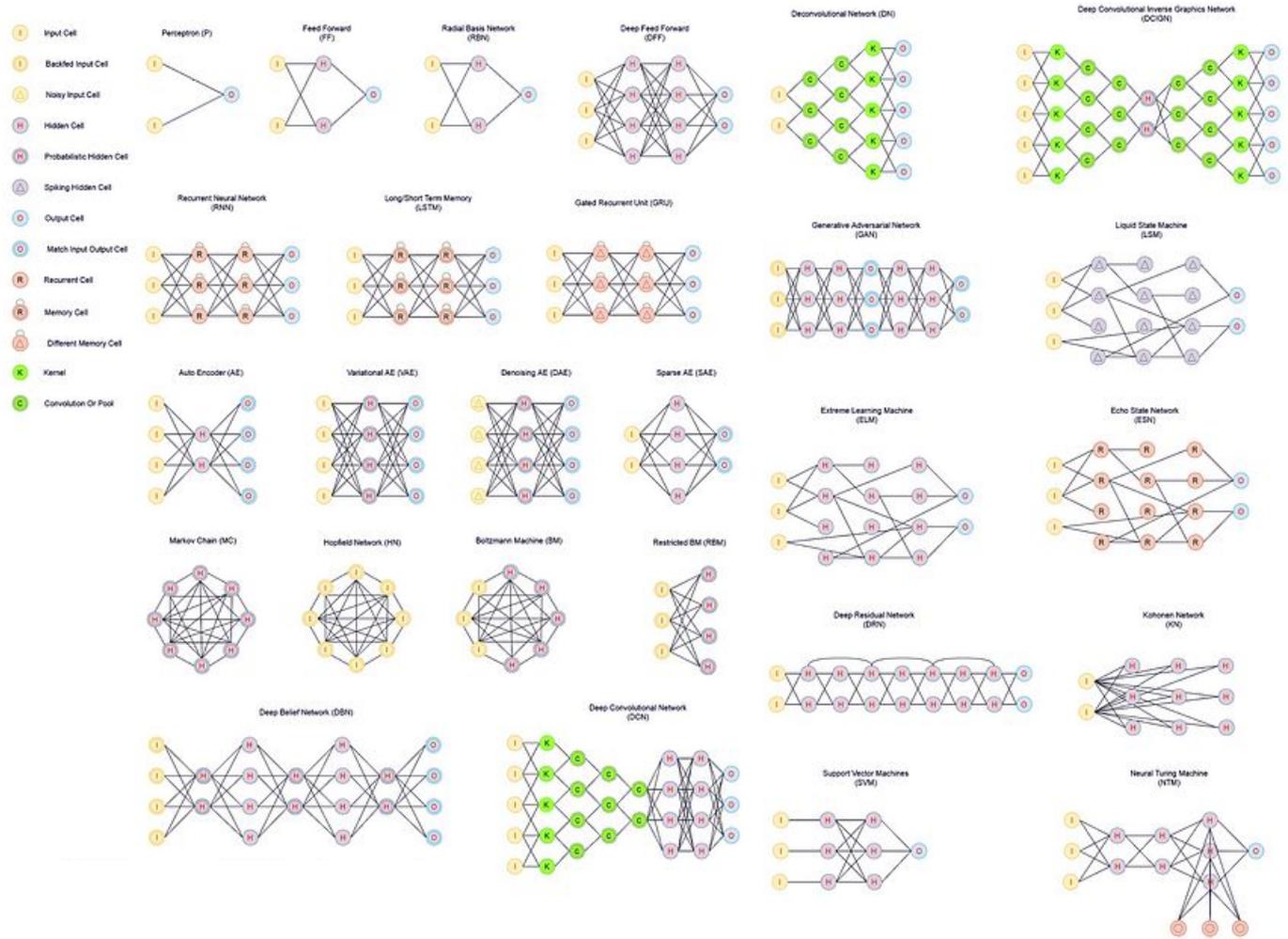


Figure 1: The Transformer - model architecture.



# O que torna os transformadores tão especiais?

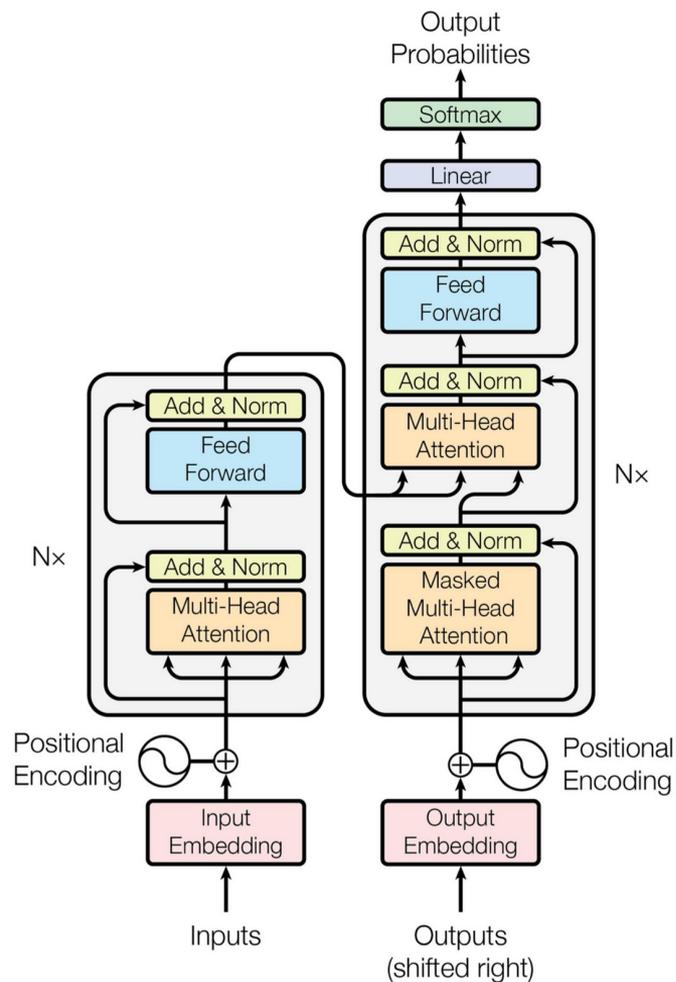


Figure 1: The Transformer - model architecture.

# Eles são bons em encontrar padrões de longo alcance em sequências de dados.

Devo comparar-te a um dia de verão?  
Tu és mais adorável e mais temperado.  
Os ventos fortes abalam os queridos botões de maio,  
E o verão tem uma duração muito curta.  
Às vezes, o olho do céu brilha com demasiado calor,  
E muitas vezes sua tez ourada se obscurece;  
E toda beleza, em algum momento, declina,  
Por acaso, ou pelo curso mutável da natureza, sem corte;  
Mas o teu eterno verão não se desvanecerá,  
Nem perderá a posse da beleza que possuis,  
Nem a morte se gabará de que tu vagueias em sua  
sombra,  
Quando em versos eternos ao Tempo tu cresceres.  
Enquanto os homens puderem respirar, ou os olhos  
puderem ver,  
Tanto tempo isso viverá, e isso te dará vida.

**Curto alcance**

Devo comparar-te a um dia de verão?  
Tu és mais adorável e mais temperado.  
Os ventos fortes abalam os queridos botões de maio,  
E o verão tem uma duração muito curta.  
Às vezes, o olho do céu brilha com demasiado calor,  
E muitas vezes sua tez ourada se obscurece;  
E toda beleza, em algum momento, declina,  
Por acaso, ou pela mudança do curso da natureza, sem  
cortes;  
Mas o teu eterno verão não se desvanecerá,  
Nem perderá a posse daquilo que te pertence  
Nem a morte se gabará de que tu vagueias em sua  
sombra,  
Quando em versos eternos ao Tempo tu cresceres.  
Enquanto os homens puderem respirar, ou os olhos  
puderem ver,  
Tanto tempo isso viverá, e isso te dará vida.

**Longo prazo**

**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.

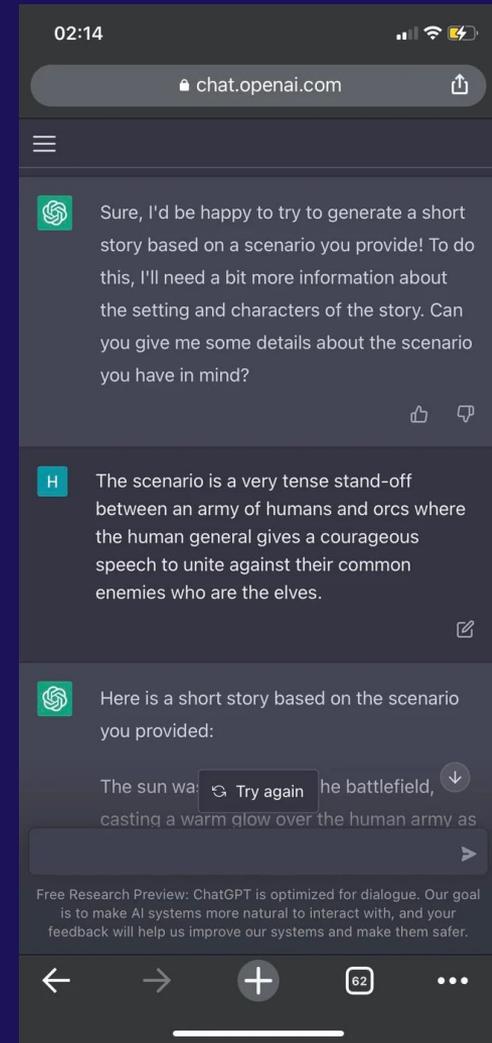


**Katie Notopoulos**  
BuzzFeed News Reporter

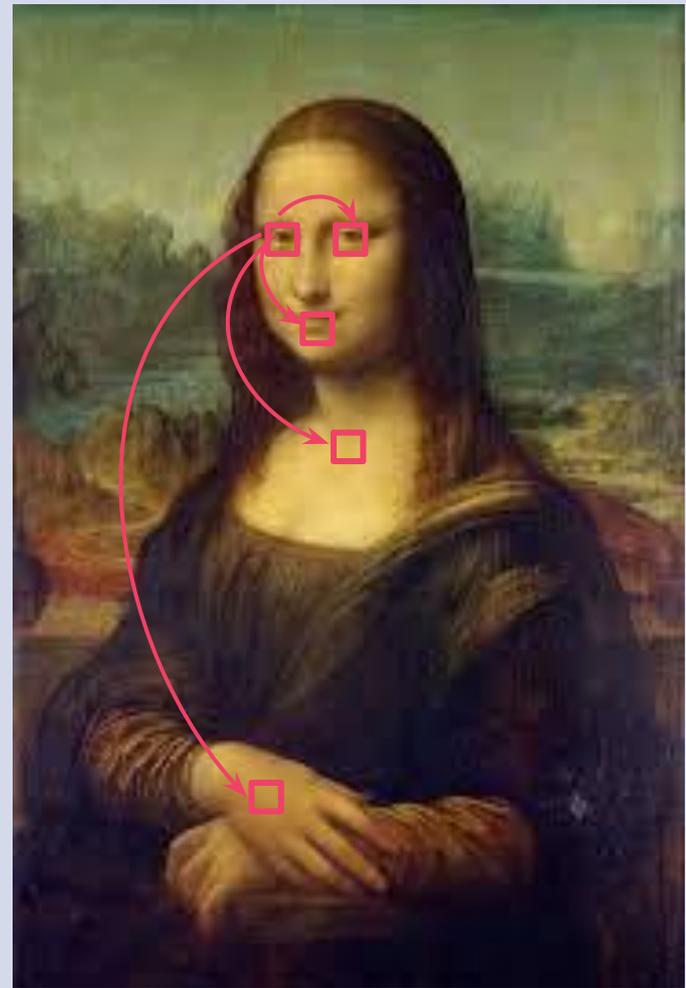
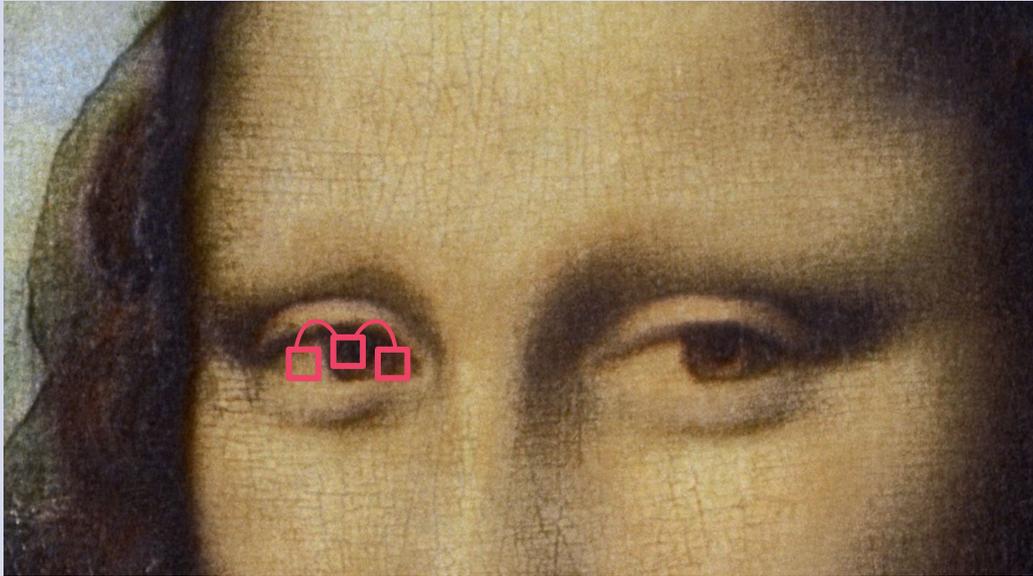
Posted on September 26, 2014 at 2:40 am

[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.



Eles são bons em encontrar padrões de longo alcance em sequências de dados.





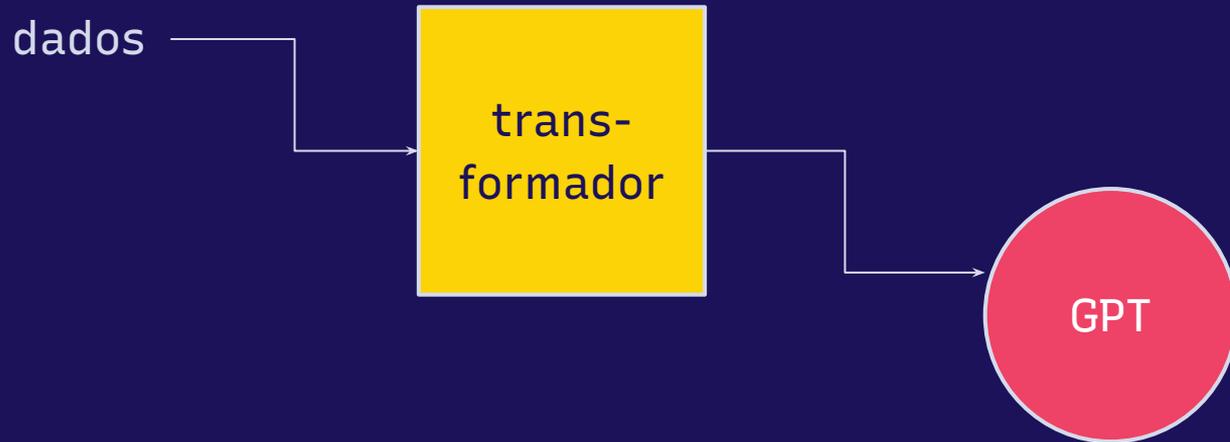
# ChatGPT

G: Generativo

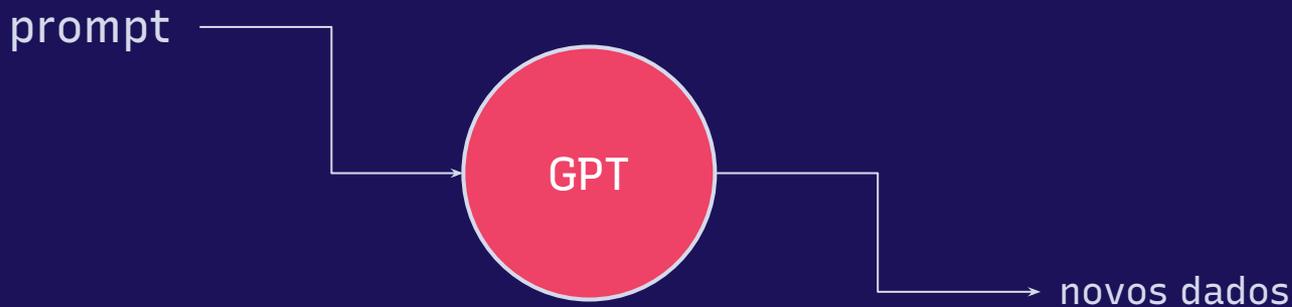
P: Pré-treinado

T: Transformador

## TREINAMENTO



## INFERÊNCIA



## TREINAMENTO

## PRÉ-TREINAMENTO

corpus gigante também conhecido como: a internet

trans-  
formador

## AJUSTE FINO

GPT  
base

## RLHF

GPT  
ajustado

GPT  
"mais  
seguro"

dados  
personalizados  
ou específicos  
para tarefas

contribuição  
humana

## TREINAMENTO

## PRÉ-TREINAMENTO

corpus gigante também conhecido como: a internet

trans-  
formador

## AJUSTE FINO

GPT  
base

## RLHF

GPT  
ajustado

GPT  
"mais  
seguro"

dados  
personalizados  
ou específicos  
para tarefas

contribuição  
humana

## TREINAMENTO

## PRÉ-TREINAMENTO

corpus gigante também conhecido como: a internet

trans-  
formador

## AJUSTE FINO

GPT  
base

## RLHF

GPT  
ajustado

GPT  
"mais  
seguro"

dados  
personalizados  
ou específicos  
para tarefas

contribuição  
humana

## TREINAMENTO

## PRÉ-TREINAMENTO

corpus gigante também conhecido como: a internet

trans-  
formador

## AJUSTE FINO

GPT  
base

## RLHF

GPT  
ajustado

GPT  
"mais  
seguro"

dados  
personalizados  
ou específicos  
para tarefas

contribuição  
humana

Aprendizagem não supervisionada

TREINAMENTO

PRÉ-TREINAMENTO

Aprendizagem supervisionada

corpus gigante também conhecido como: a internet

trans-  
formador

AJUSTE FINO

Aprendizado por reforço

RLHF

GPT base

dados personalizados ou específicos para tarefas

contribuição humana

GPT ajustado

GPT "mais seguro"

# INFERÊNCIA

# MODERAÇÃO

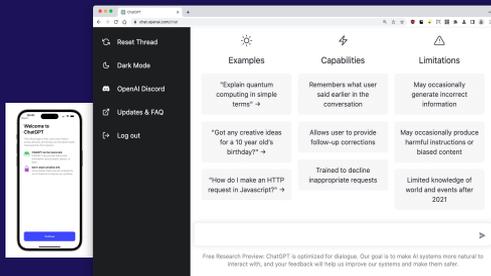
prompt

AJUSTE

novos dados

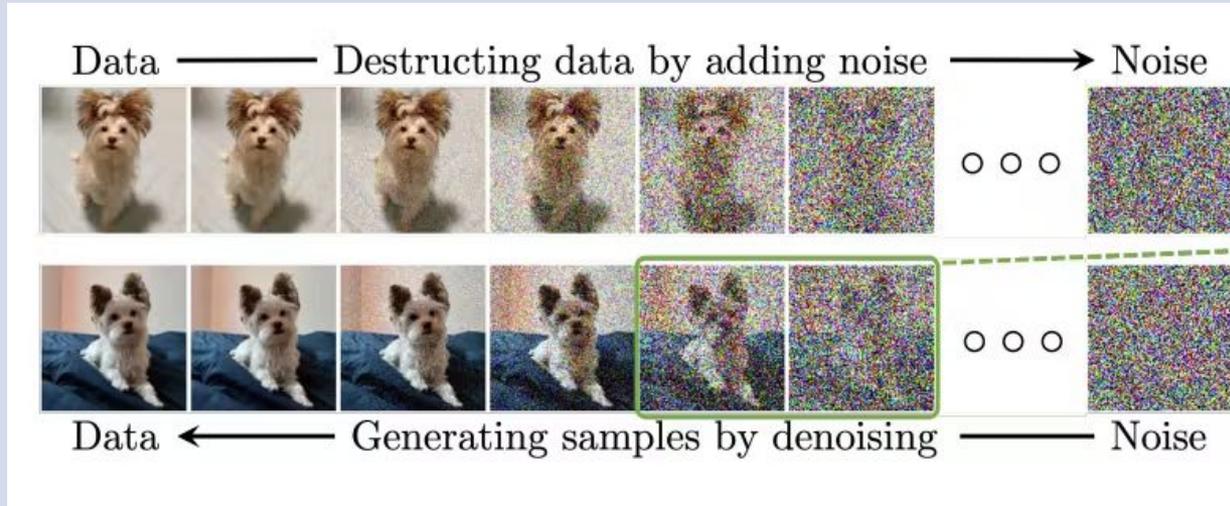
GPT  
"mais  
seguro"

# INTERFACE



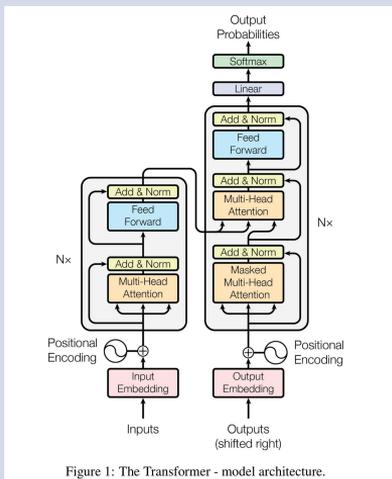
Os geradores de imagens são muito semelhantes, com uma diferença conceitual principal.

# Eles usam difusão.



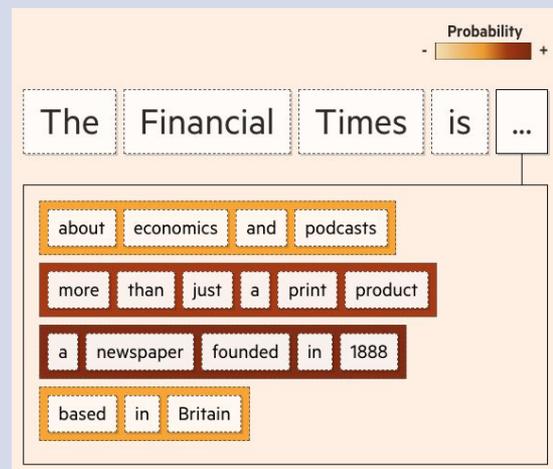
# A difusão é um objetivo.

## Arquitetura



A arquitetura usada para treinar um LLM é: **o transformador.**

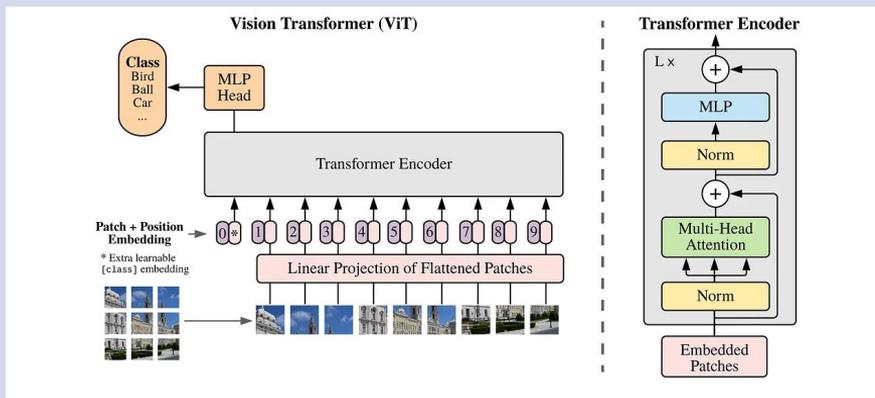
## Objetivo



O objetivo utilizado para treinar um LLM é: **prever a(s) próxima(s) palavra(s).**

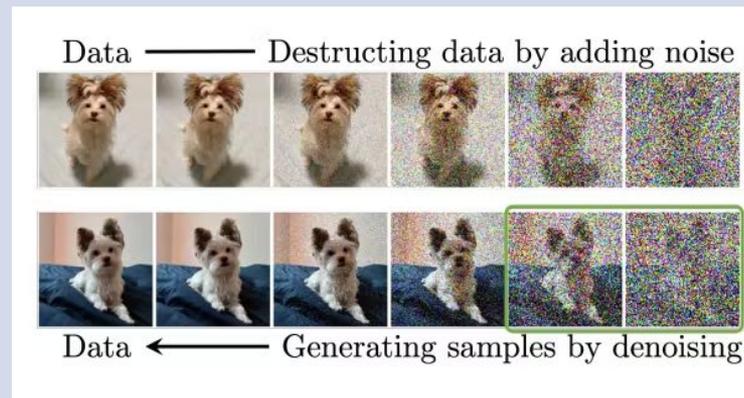
# A difusão é um objetivo.

## Arquitetura



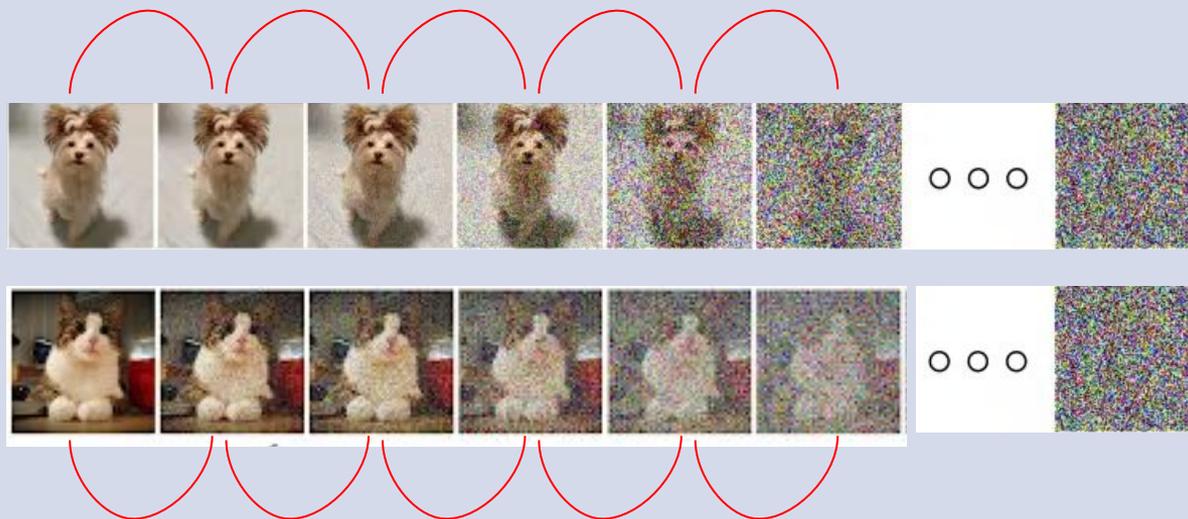
A arquitetura usada para treinar um gerador de imagens é: o **transformador de visão**.

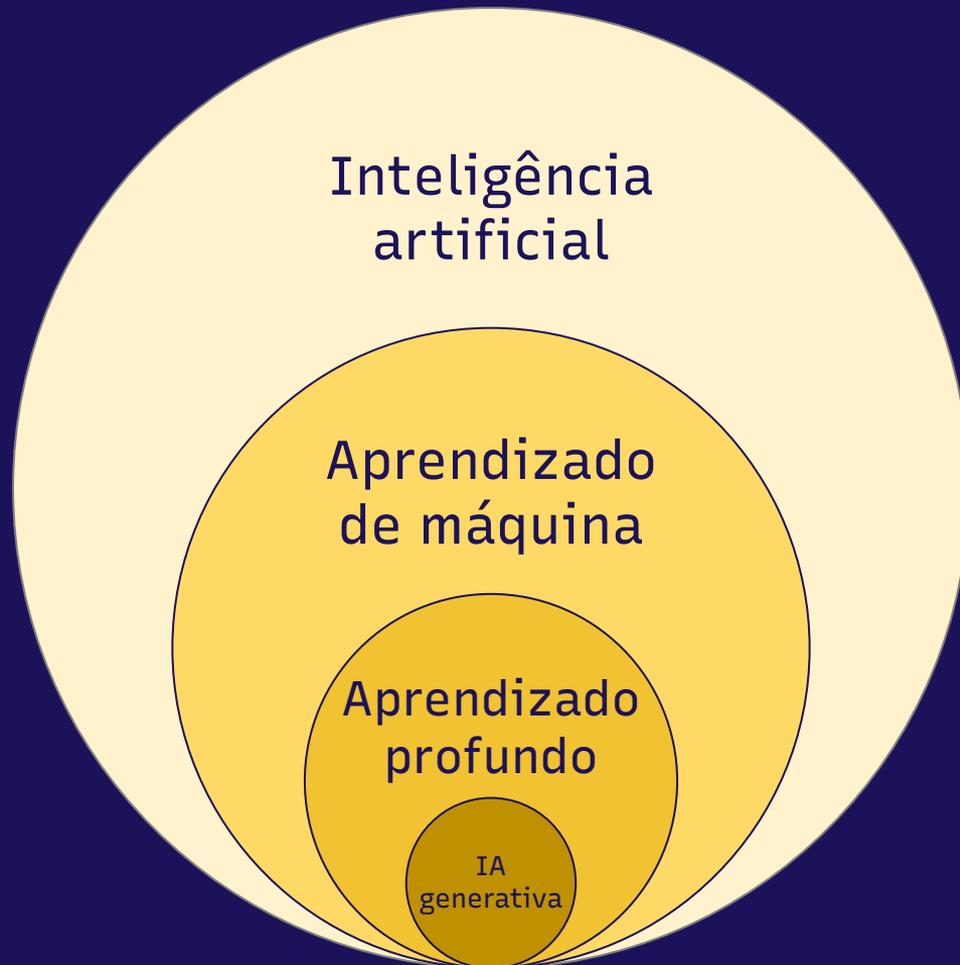
## Objetivo



O objetivo utilizado para treinar um gerador de imagens é: **difusão, ou recuperar uma imagem corrompida**.

# Difusão é encontrar o sinal no ruído.





## TREINAMENTO

## PRÉ-TREINAMENTO

corpus gigante também conhecido como: a internet

trans-formador

## AJUSTE FINO

GPT base

## RLHF

GPT ajustado

GPT "mais seguro"

**Ninguém sabe o que há aqui.**

desinformação  
erros factuais  
equívocos  
teorias da conspiração  
especulações  
lendas urbanas  
contos  
ficção científica  
fanfics  
tweets  
desabafos  
hipérbole  
racismo  
abusos

dados personalizados ou específicos para tarefas

contribuição humana



SIGN IN

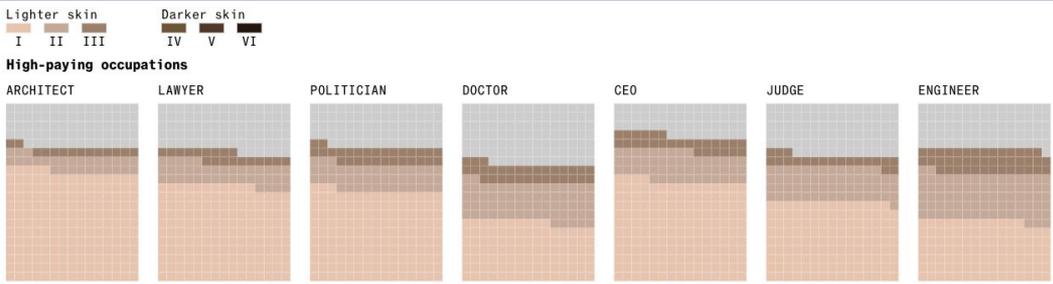
SUBSCRIBE

[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**

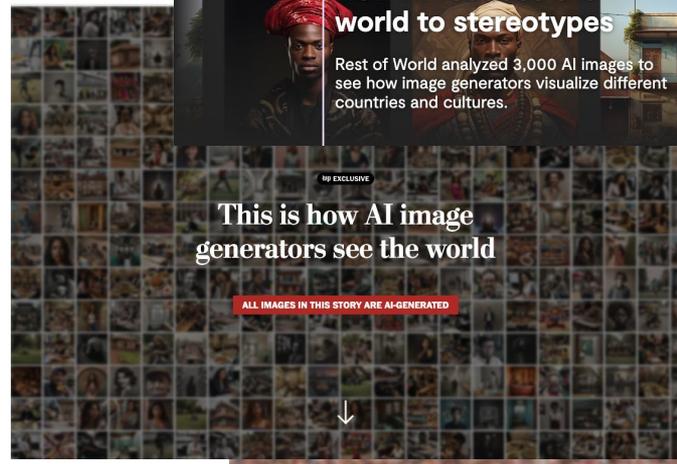
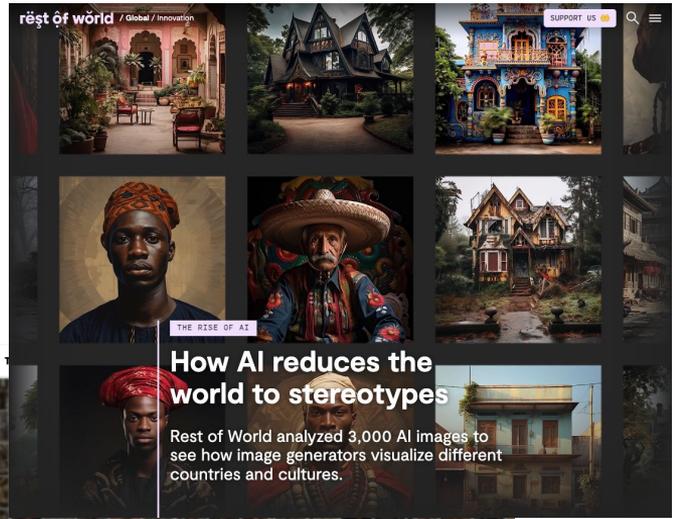
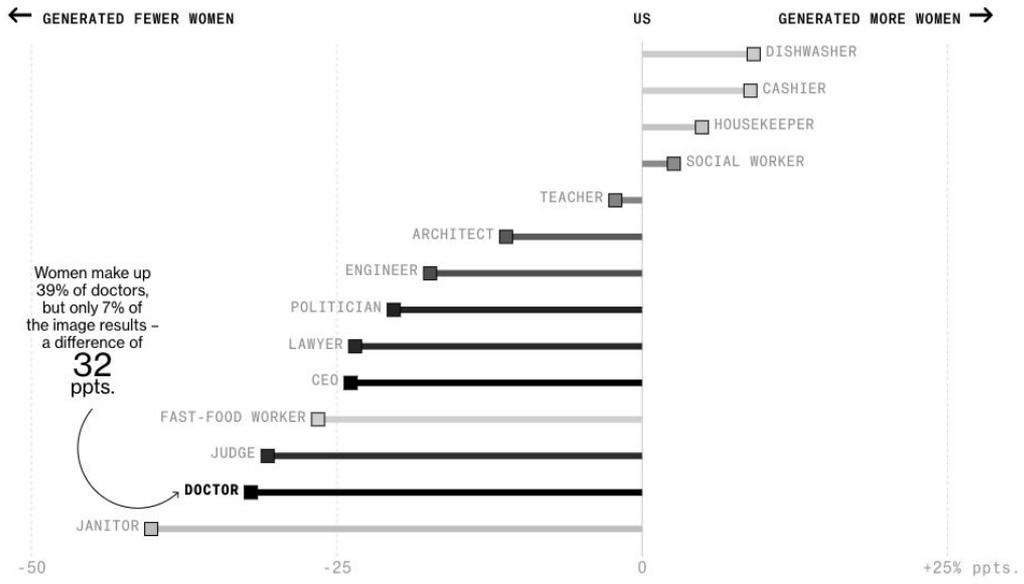
The model is a massive part of the AI-ecosystem, used by Stable Diffusion and other major generative AI products. The removal follows discoveries made by Stanford researchers, who found thousands instances of suspected child sexual abuse material in the dataset.

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



**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

A IA generativa trata da remixagem de dados. NÃO se trata de fornecer informações.



BBC News sends Breaking News alerts on a single subject



Breaking News

Tue 20:04

CEO shooting suspect Luigi Mangione in angry outburst outside court as he fights extradition to New York - follow live

Apple sends an AI-generated summary of BBC News headlines - which can create inaccuracies



BBC News

2h ago

Luigi Mangione shoots himself; Syrian mother hopes Assad pays the price; South Korea police raid Yoon Suk Yeol's office.

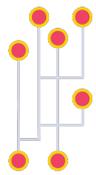
# 51%

**das respostas do ChatGPT, Copilot, Gemini e  
Perplexity sobre as notícias estavam INEXATAS  
- Pesquisa da 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 start of the word "AI".

# 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**