

A decorative graphic consisting of white lines and small circles on a pink background, resembling a circuit board or a stylized frame. The lines are arranged in a U-shape, with a horizontal base and two vertical sides. The circles are placed at various points along the lines, some at the ends and some in the middle. The overall effect is a clean, modern, and tech-oriented design.

Intro to Generative AI

What we will cover in this module

- How generative AI differs from other kinds of AI
- What goes into making a generative AI model
- The difference between text and image generators
- Major challenges with generative AI



Reminder: don't stress.

 If this is your first time engaging in the technical details, use this as a way to familiarize yourself with terms & concepts.

 If you already know some of these details, use this as a way to clarify your understanding.

 Treat this slide deck as a resource. Refer back to it as needed!

generative AI

Artificial
intelligence

Machine
learning

Deep
learning

GenAI

(data)

(big data)

(massive data,
big compute)

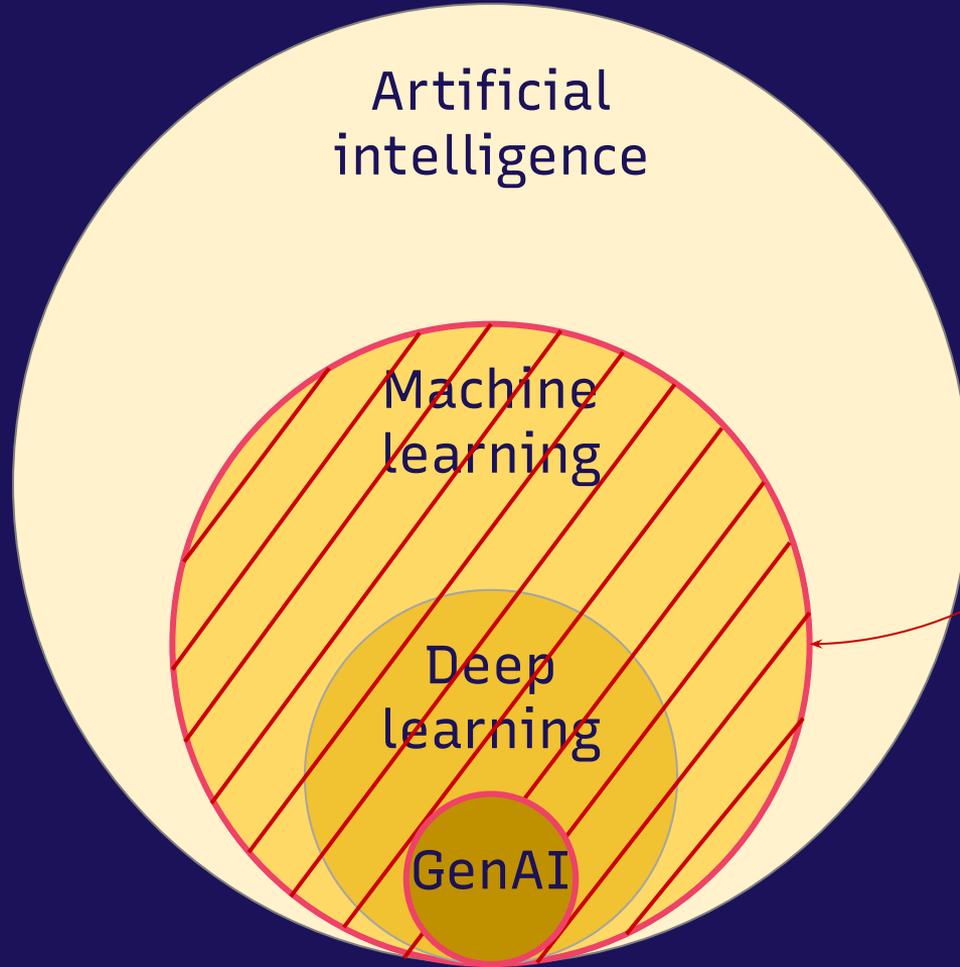
Artificial
intelligence

Machine
learning

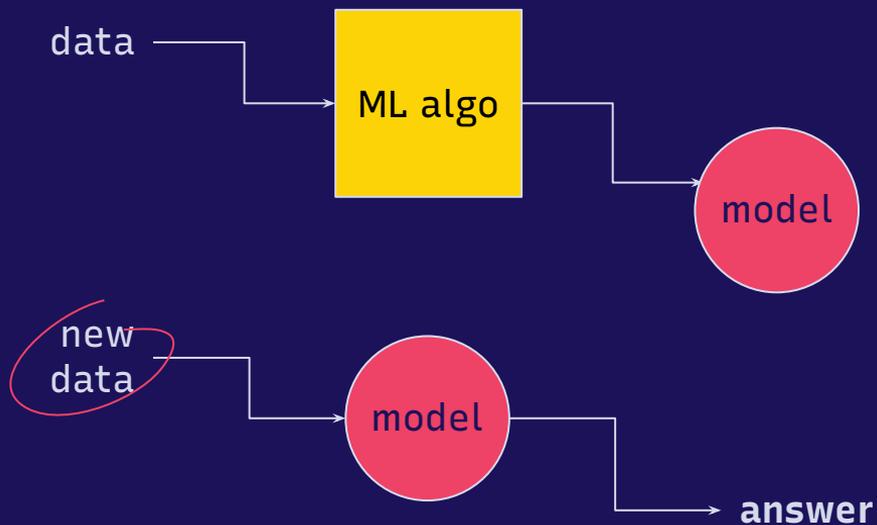
Deep
learning

GenAI

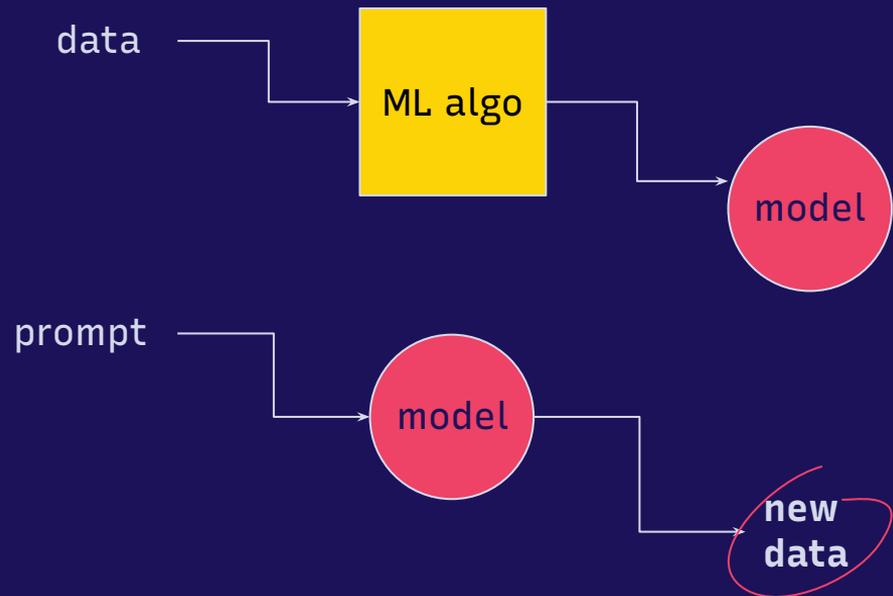
Predictive
AI



Predictive AI



Generative AI



The current*
generative AI era is
based on
transformers.

* Keep in mind: This could change one day!

What are transformers?

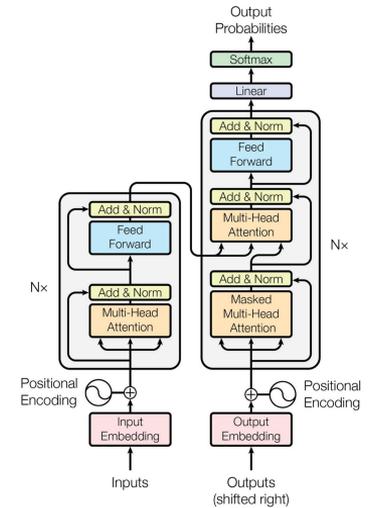
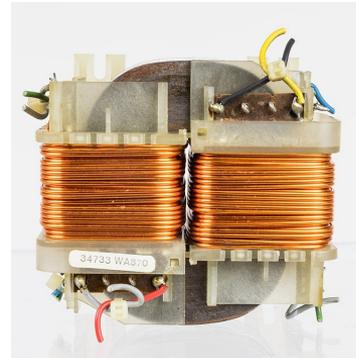
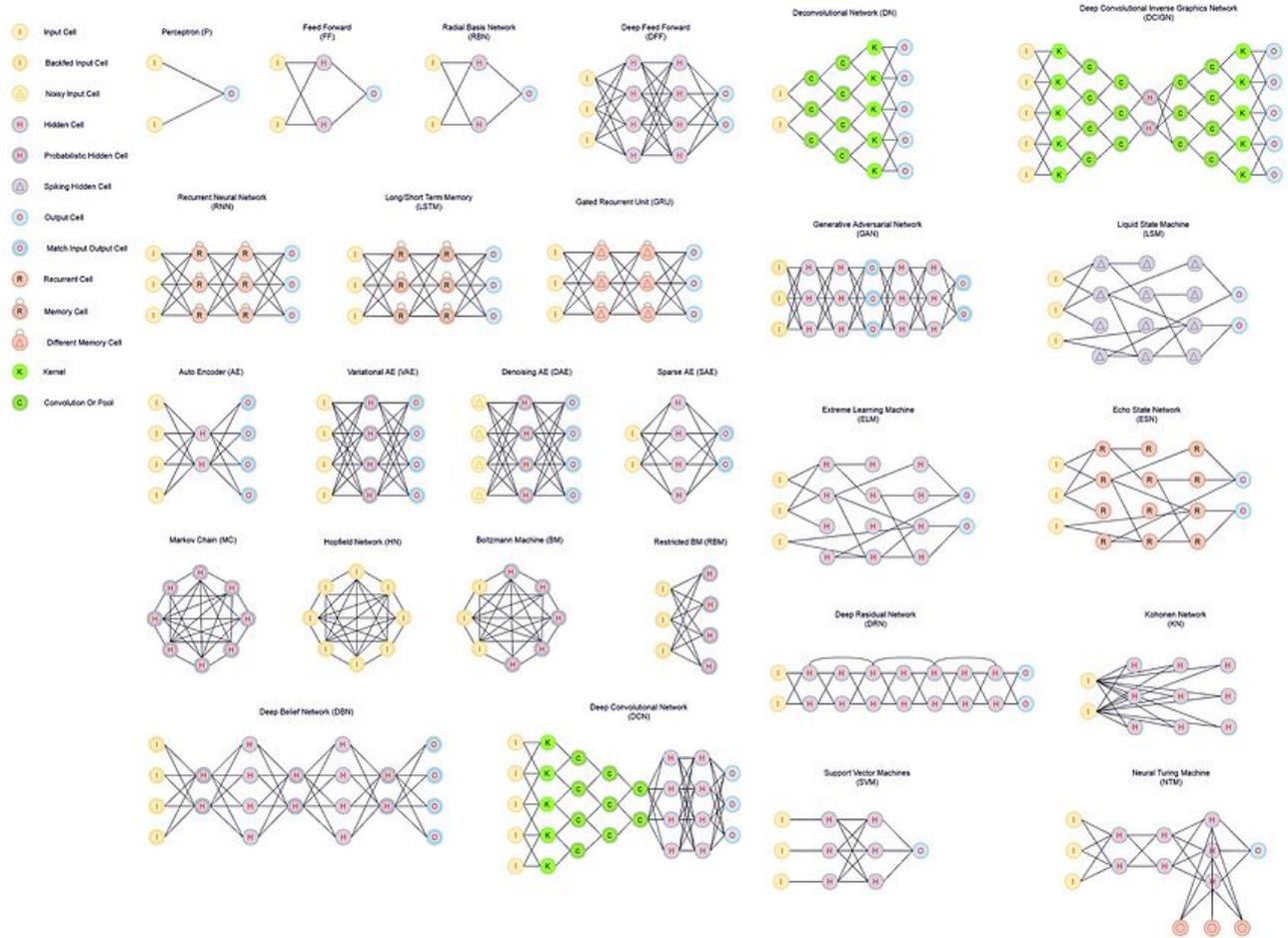


Figure 1: The Transformer - model architecture.



What makes transformers so special?

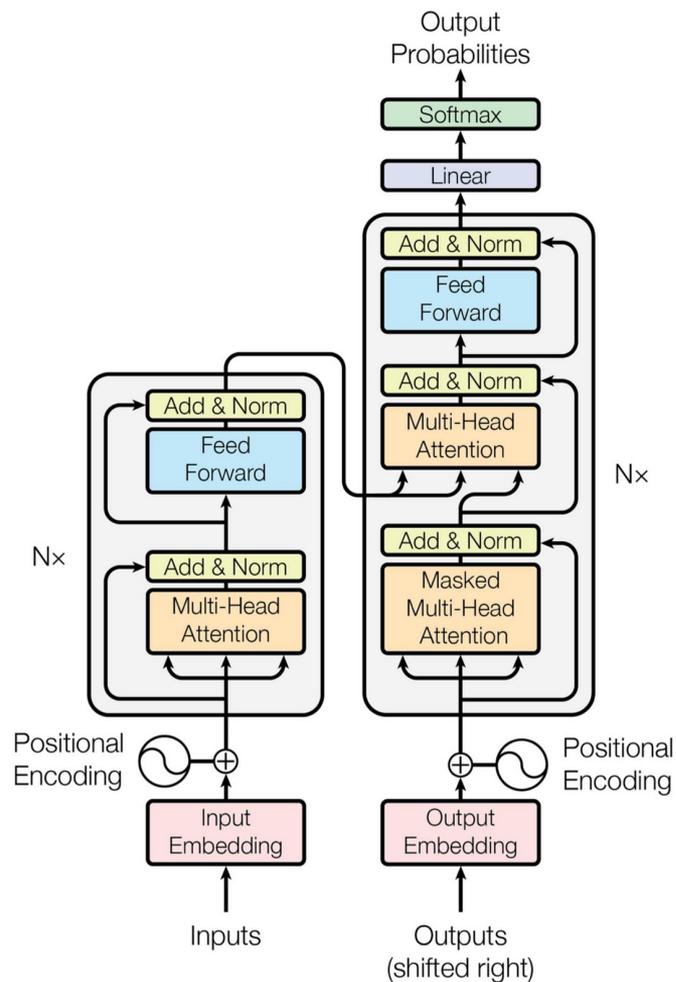


Figure 1: The Transformer - model architecture.

They are good at finding long-range patterns in sequences of data.

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.

Short-range

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.

Long-range



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.

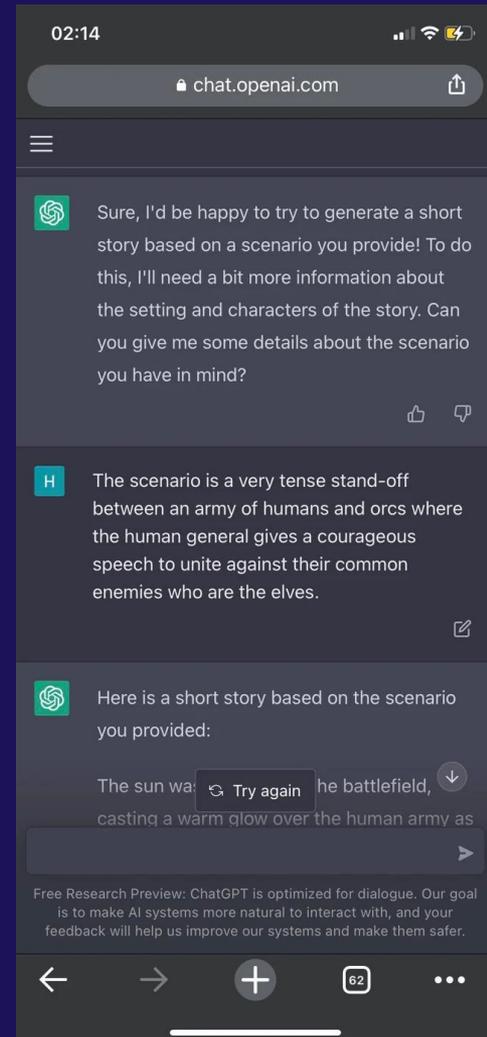


Katie Notopoulos
BuzzFeed News Reporter

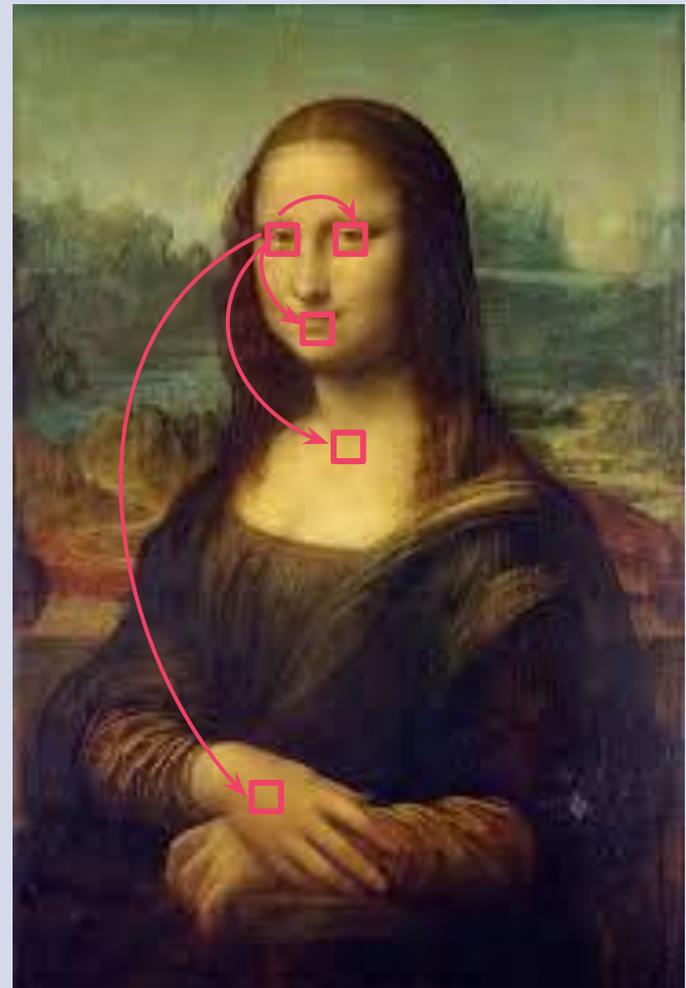
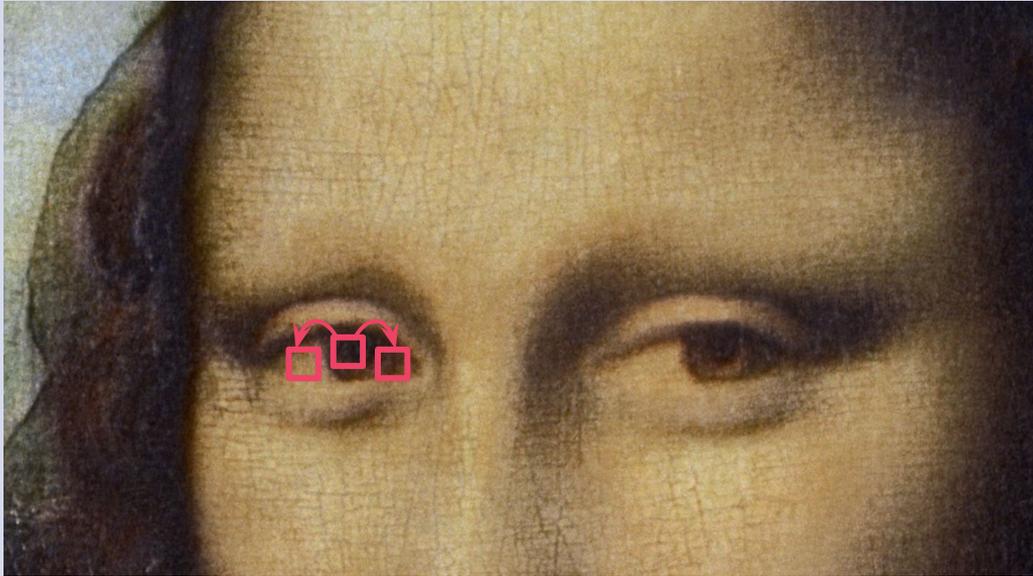
Posted on September 26, 2014 at 2:40 am

[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.



They are good at finding long-range patterns in sequences of data.



ChatGPT

G: Generative

P: Pre-trained

T: Transformer

TRAIN

data

trans-
former

GPT

INFERENCE

prompt

GPT

new data

TRAIN

PRE-TRAINING

FINE-TUNING

RLHF

giant
corpus
aka: the
internet

trans-
former

Base
GPT

tailored or
task-specific
data

Tuned
GPT

human
input

"Safer"
GPT

TRAIN

PRE-TRAINING

FINE-TUNING

RLHF

giant
corpus
aka: the
internet

trans-
former

Base
GPT

tailored or
task-specific
data

Tuned
GPT

human
input

"Safer"
GPT

TRAIN

PRE-TRAINING

FINE-TUNING

RLHF

giant
corpus
aka: the
internet

trans-
former

Base
GPT

tailored or
task-specific
data

Tuned
GPT

human
input

"Safer"
GPT

TRAIN

PRE-TRAINING

FINE-TUNING

RLHF

giant
corpus
aka: the
internet

trans-
former

Base
GPT

tailored or
task-specific
data

Tuned
GPT

human
input

"Safer"
GPT

Unsupervised learning

Supervised learning

Reinforcement learning

TRAIN

PRE-TRAINING

FINE-TUNING

RLHF

giant corpus aka: the internet

trans-former

Base GPT

tailored or task-specific data

Tuned GPT

human input

"Safer" GPT

INFERENCE

MODERATION

prompt

AUGMENTATION

new data

“Safer”
GPT

INTERFACE

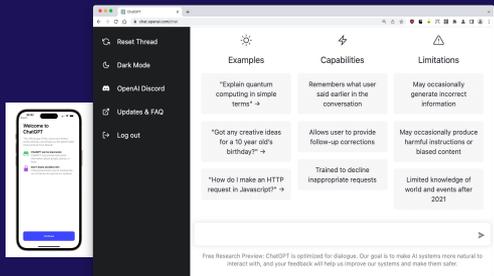
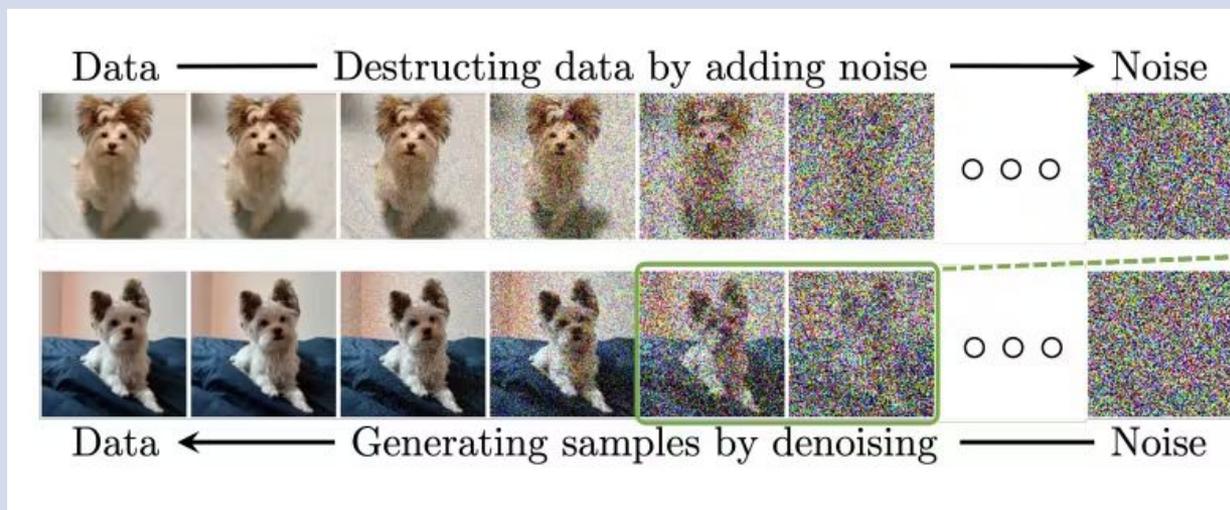


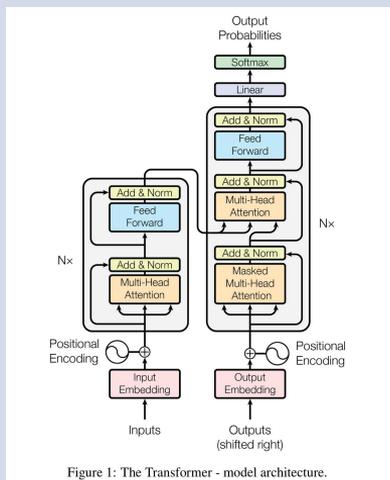
Image generators
are really similar,
with one main
conceptual
difference.

They use diffusion.



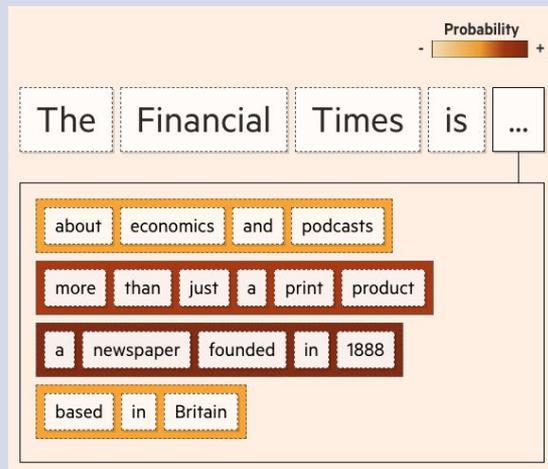
Diffusion is an objective.

Architecture



The architecture used to train an LLM is:
the transformer.

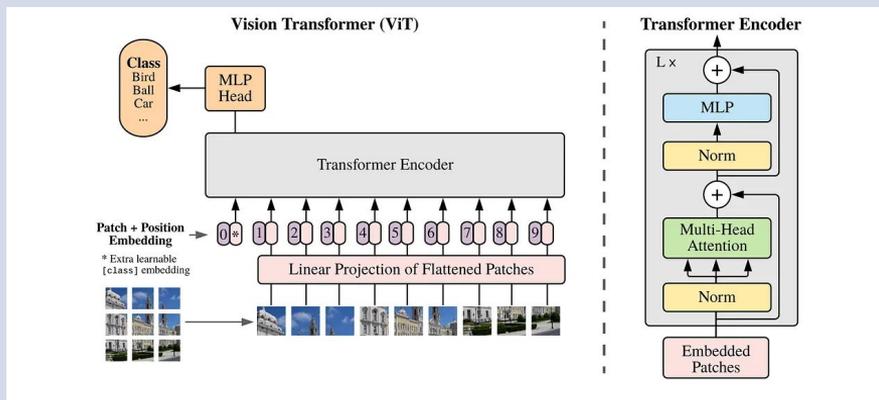
Objective



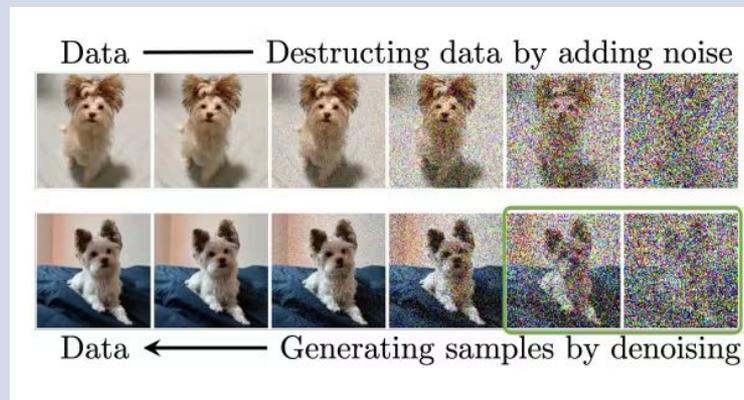
The objective used to train an LLM is:
to predict the next word(s).

Diffusion is an objective.

Architecture



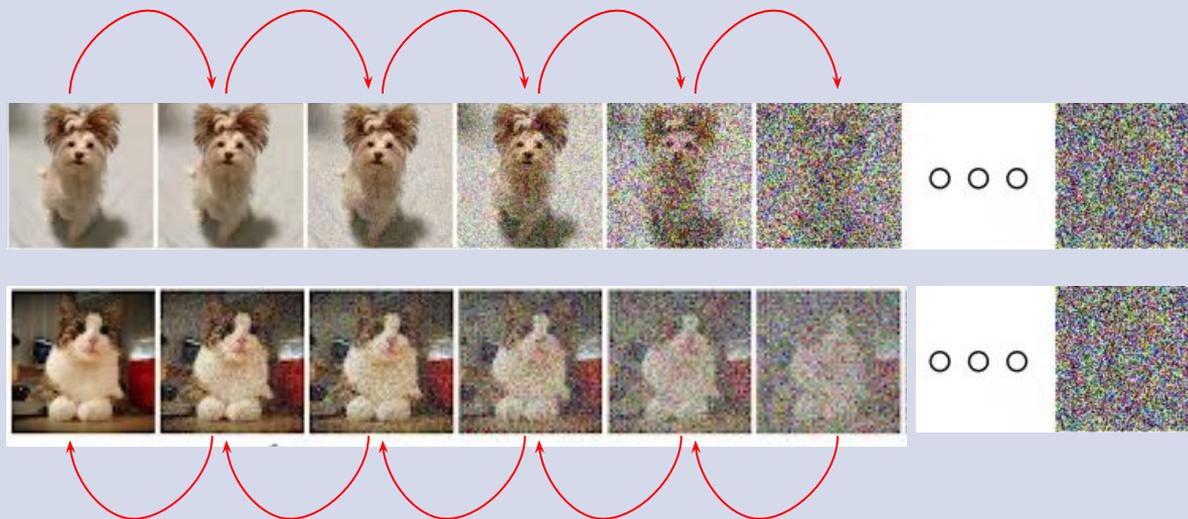
Objective



The architecture used to train an image generator is:
the vision transformer.

The objective used to train an image generator is:
diffusion, or to recover a corrupted image.

Diffusion is about finding signal in the noise.



Artificial
intelligence

Machine
learning

Deep
learning

GenAI

TRAIN

PRE-TRAINING

FINE-TUNING

RLHF

giant corpus

No one knows what's in here.

disinformation
factual errors
misconceptions
conspiracy theories
speculation
urban legends
short stories
science fiction
fan fiction
tweets
rants
hyperbole
racism
abuse

trans-
former

Base
GPT

tailored or
task-specific
data

Tuned
GPT

human
input

"Safer"
GPT



SIGN IN

SUBSCRIBE

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

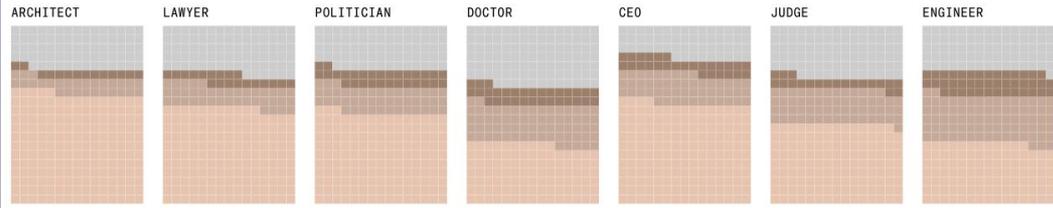
Largest Dataset Powering AI Images Removed After Discovery of **Child Sexual Abuse Material**

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

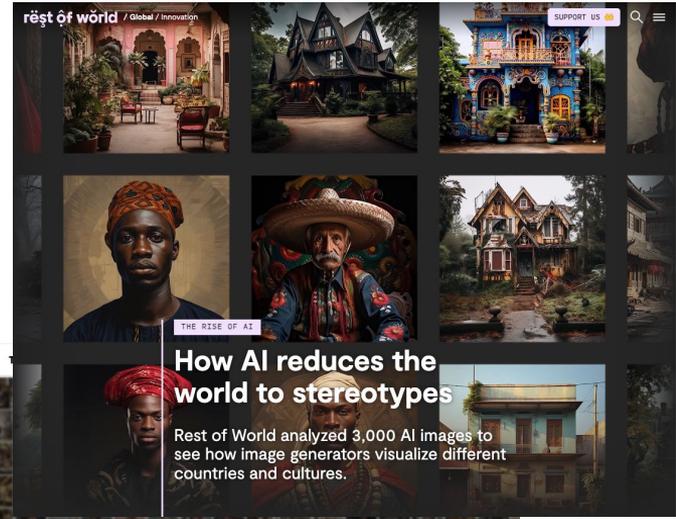
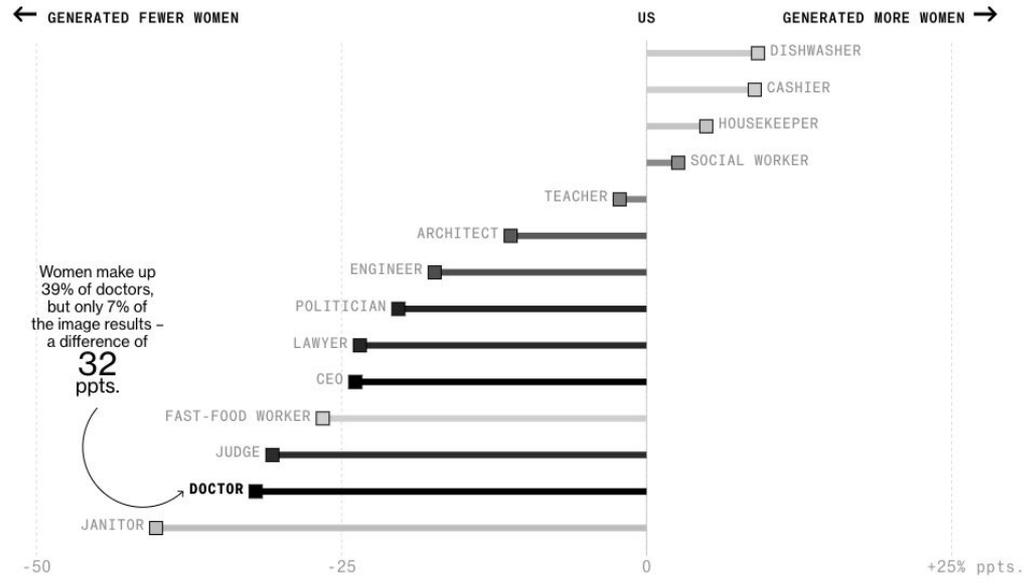
Lighter skin
I II III
Darker skin
IV V VI

High-paying occupations

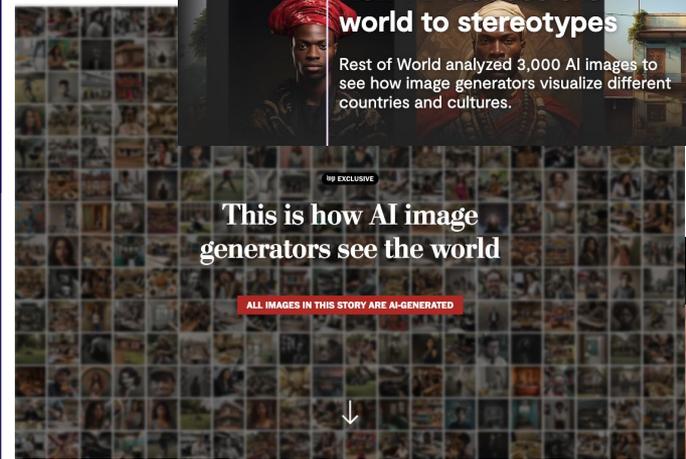


Working Women Misrepresented Across the Board
Stable Diffusion results compared to US demographics for each occupation

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



ER



**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

Generative AI
is about
remixing data.
It is NOT about
retrieving
information.



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.

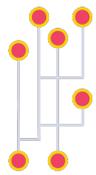
51%

**of ChatGPT, Copilot, Gemini, and Perplexity's
answers about the news were INACCURATE**

-BBC Research

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 with four circular nodes. The top node is yellow, the second is light blue, the third is light blue, and the bottom is light blue. A horizontal line extends from the top node to the right, ending in a yellow circle.

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