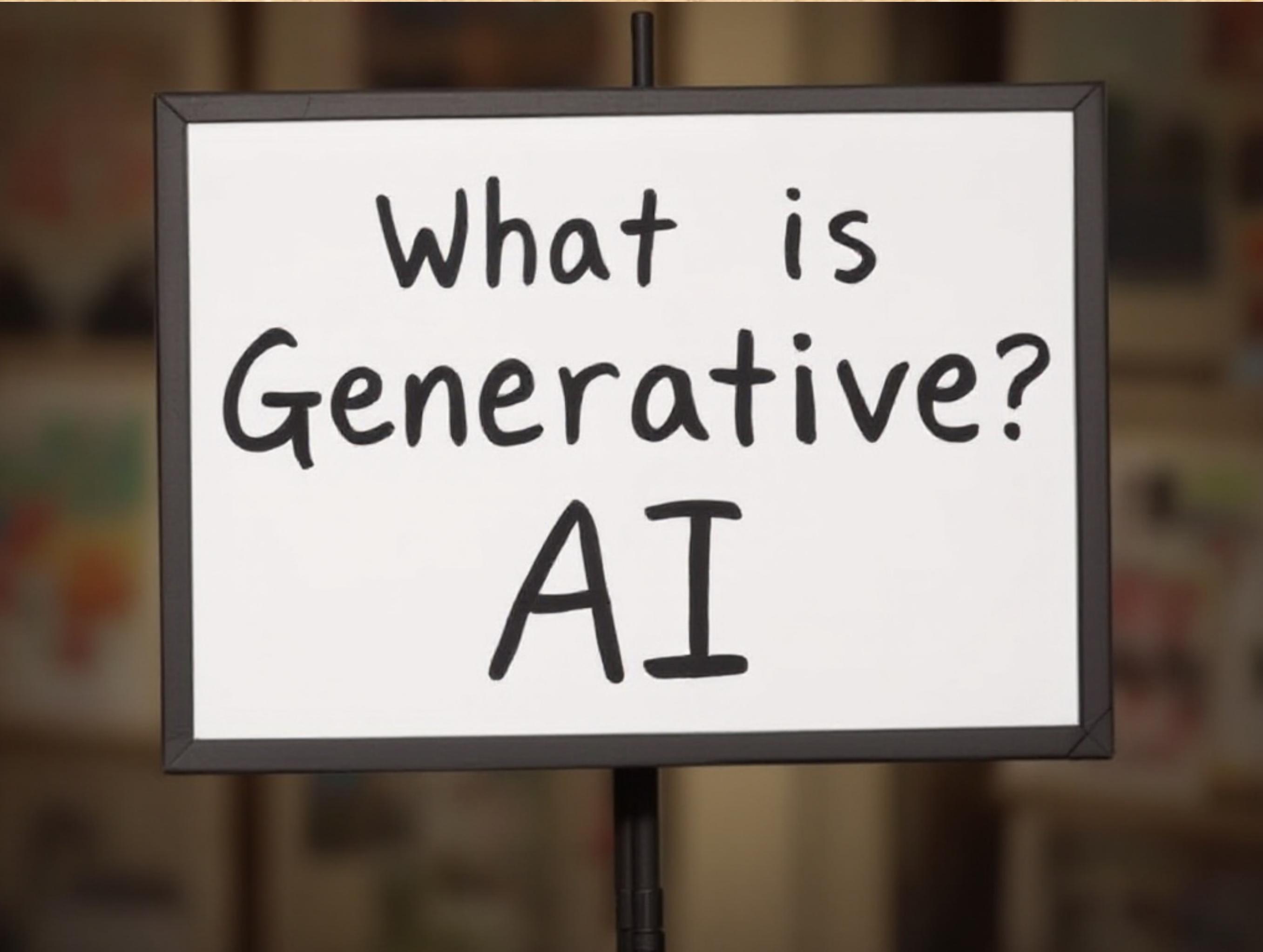


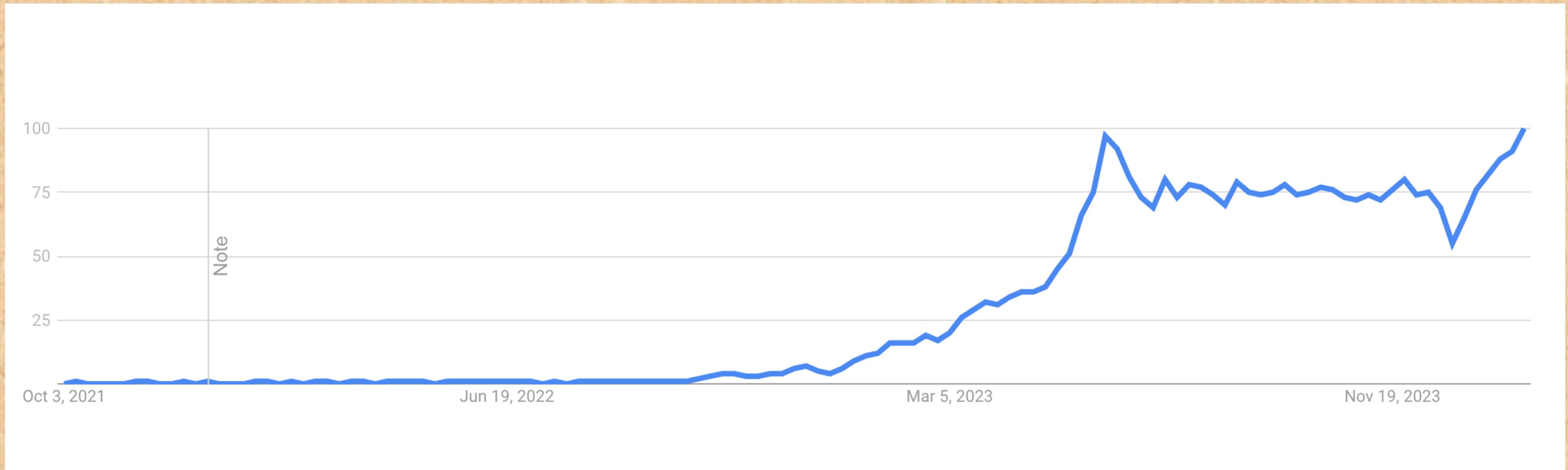
Module 1 - Topic 1

1.2 What is Generative AI?



What is
Generative?
AI

Which graph is this?



Hint: this is why we are here!

What is Generative AI?

Unlike traditional AI systems that were primarily used for prediction and analysis,

Generative AI is

- ~~ AI that creates new, original content
- ~~ Can generate text, images, audio, and more
- ~~ Works based on prompts or input data

Traditional AI vs Generative AI

	Traditional AI	Generative AI
Primary Goal	Designed to analyze data and make predictions.	Designed to create new data and content.
Applications	It is primarily used for classification, regression, etc.	It can generate text, images, audio, and other forms.
Adaptability	Requires extensive retraining to handle new data or scenarios.	Can continuously learn and improve from new data.
Automation	Can automate analytical tasks but lacks ability to automate creative work.	Can automate creative tasks like writing, art, etc.
Innovation	Constrained by rules and logic boundaries. Less potential for breakthrough innovations.	Can make novel connections and create new innovations. Unlocks new possibilities.

With Generative AI

You can be a CREATOR

~~ Text

~~ Images

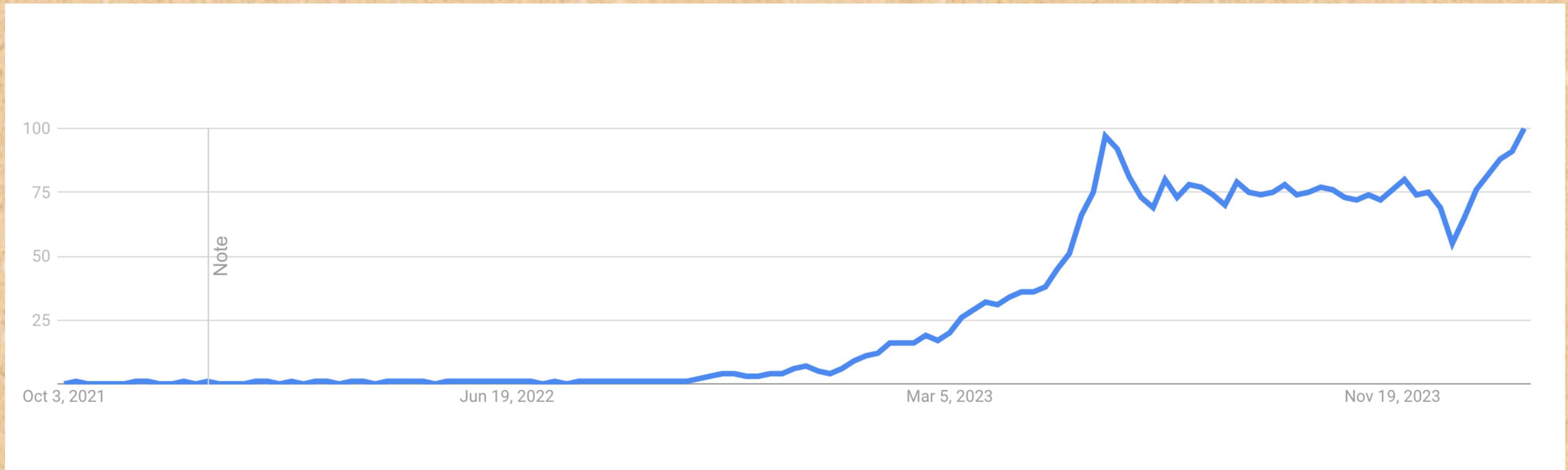
~~ Audio/Music

~~ Videos

But,

What has changed suddenly?

Which graph is this?



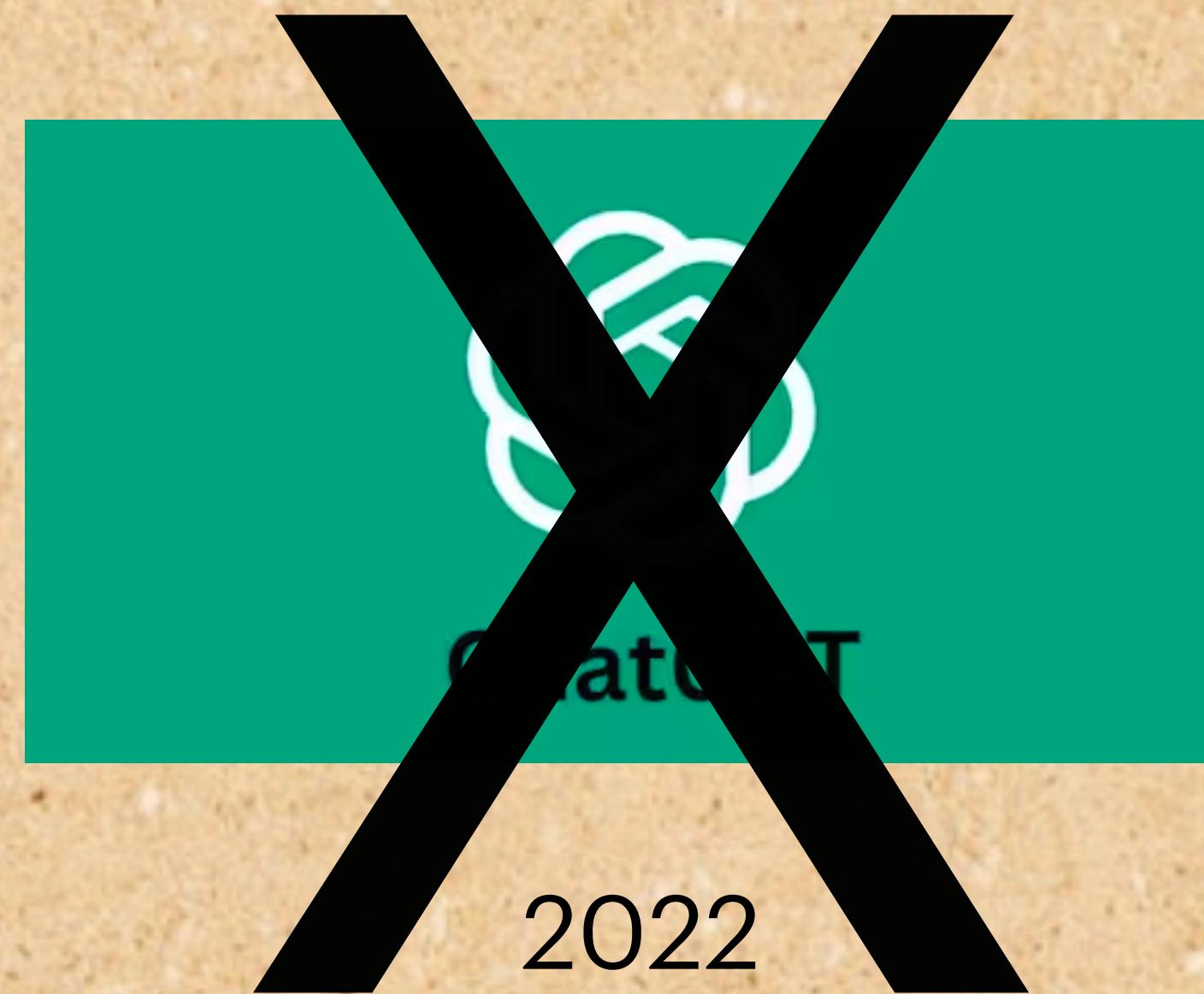
What has changed suddenly?

First AI Chatbot



2022

First AI Chatbot



Welcome to

```
EEEEEE LL     IIII   ZZZZZZ  AAAAAA
EE      LL     II     ZZ    AA    AA
EEEEEE LL     II     ZZZ   AAAAAAAA
EE      LL     II     ZZ    AA    AA
EEEEEE LLLLLL IIII   ZZZZZZ  AA    AA
```

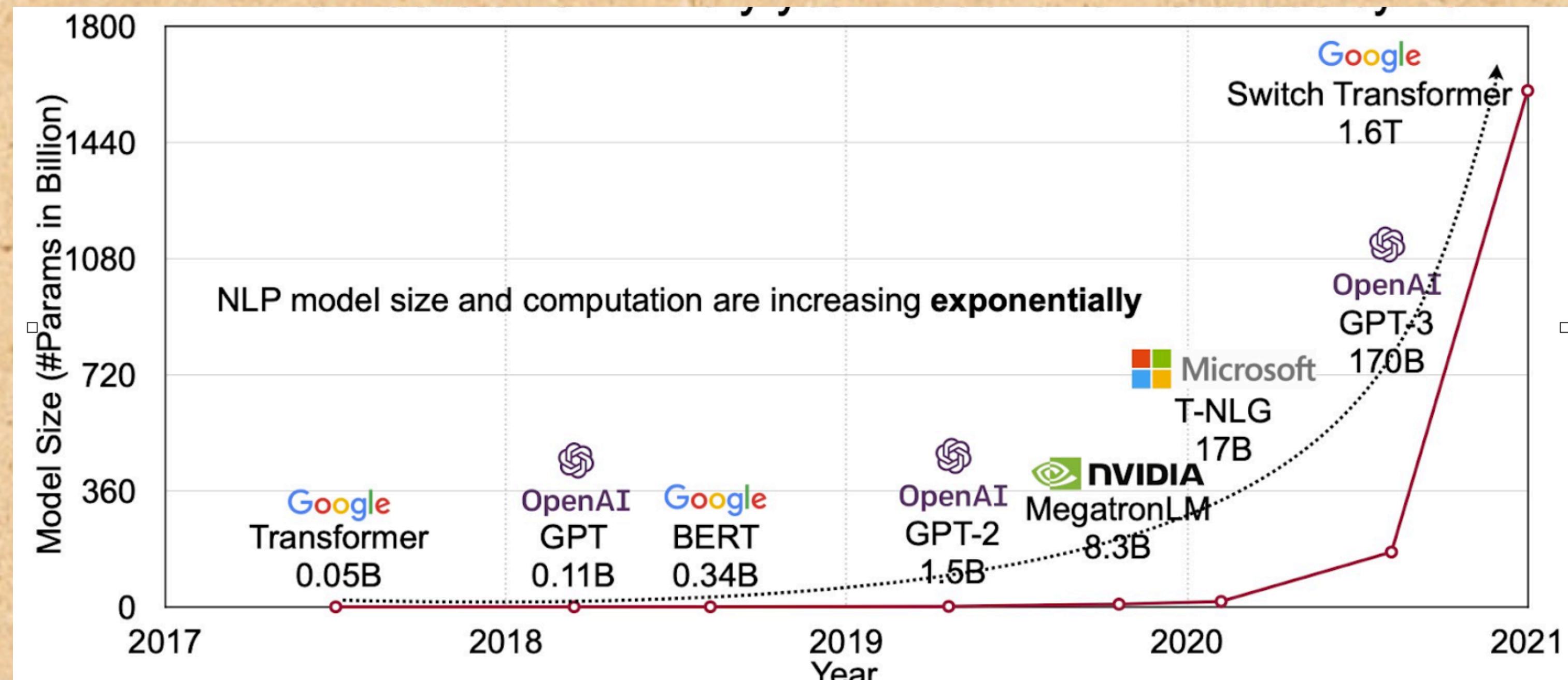
Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

```
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:
```

1965

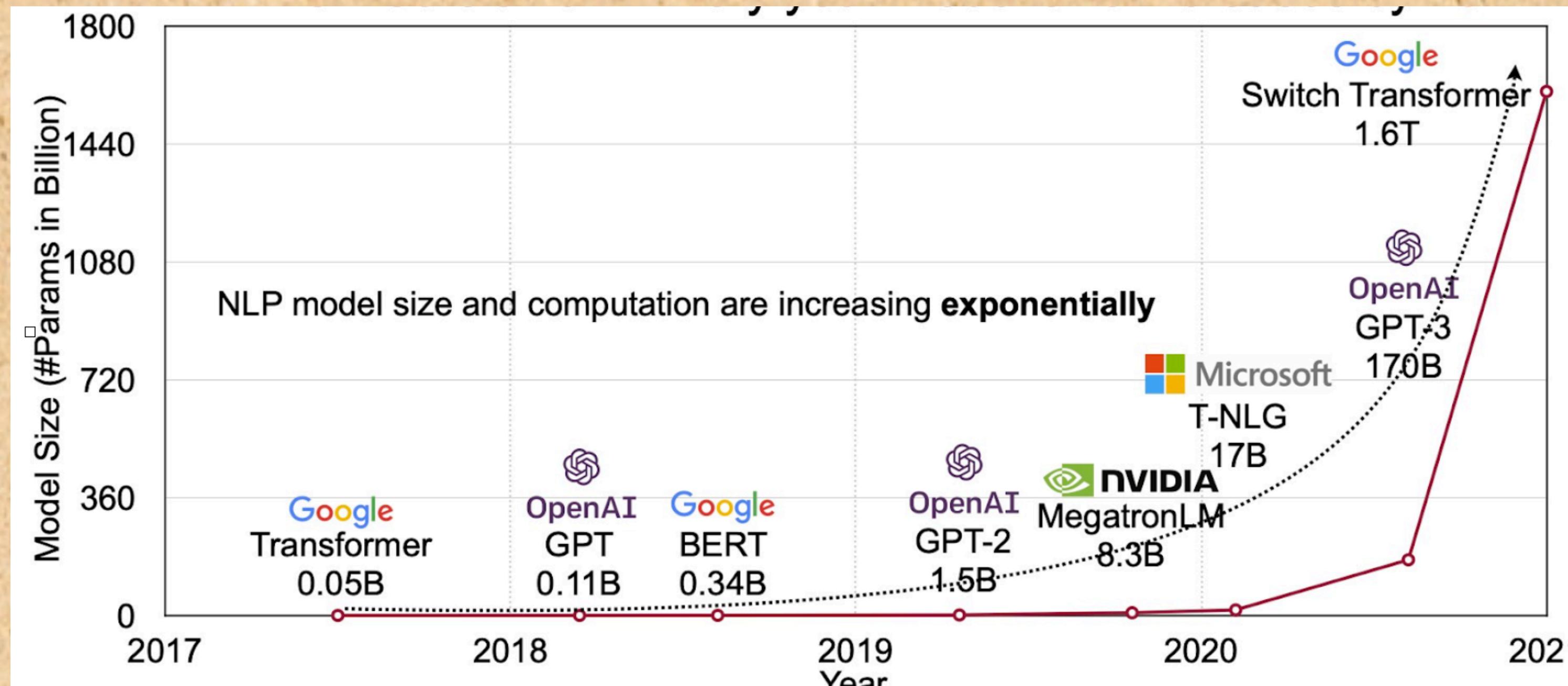
What has changed?

First language model (ELIZA) was introduced was introduced in 1964



What has changed?

Attention Is All You Need - proposes new architecture for LLMs



Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Ilia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including

Attention Is All You Need - proposes Transformer architecture

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including