

GPT3

- Atishya Jain

```
[example] an input that says "search" [toCode] Class App extends React Component... </div> } } }  
[example] a button that says "I'm feeling lucky" [toCode] Class App extends React Component...  
[example] an input that says "enter a todo" [toCode]
```



The background features a dark teal to black gradient. It is overlaid with a complex network of thin, light-colored lines that form various geometric shapes, primarily triangles and polygons. Some of these shapes are filled with a semi-transparent teal or light blue color. Scattered throughout the scene are numerous small, glowing dots in shades of teal, blue, and orange, creating a bokeh effect. The overall aesthetic is futuristic and digital.

Let's Dissect it


Auto Regressive

Let's Dissect it

Auto Regressive

Byte Pair Encoding

Let's Dissect it

The background features a dark teal color with a network of thin white lines connecting various points. Some points are highlighted with larger, glowing teal circles, while others are smaller blue dots. The overall aesthetic is futuristic and digital.

aaabdaaabac

aaabdaaac

Merge aa

ZabdZabac
Z=aa

aaabdaaac

Merge aa

ZabdZabac
Z=aa

Merge ab

ZYdZYac
Y=ab
Z=aa

Merge aa

aaabdaaac

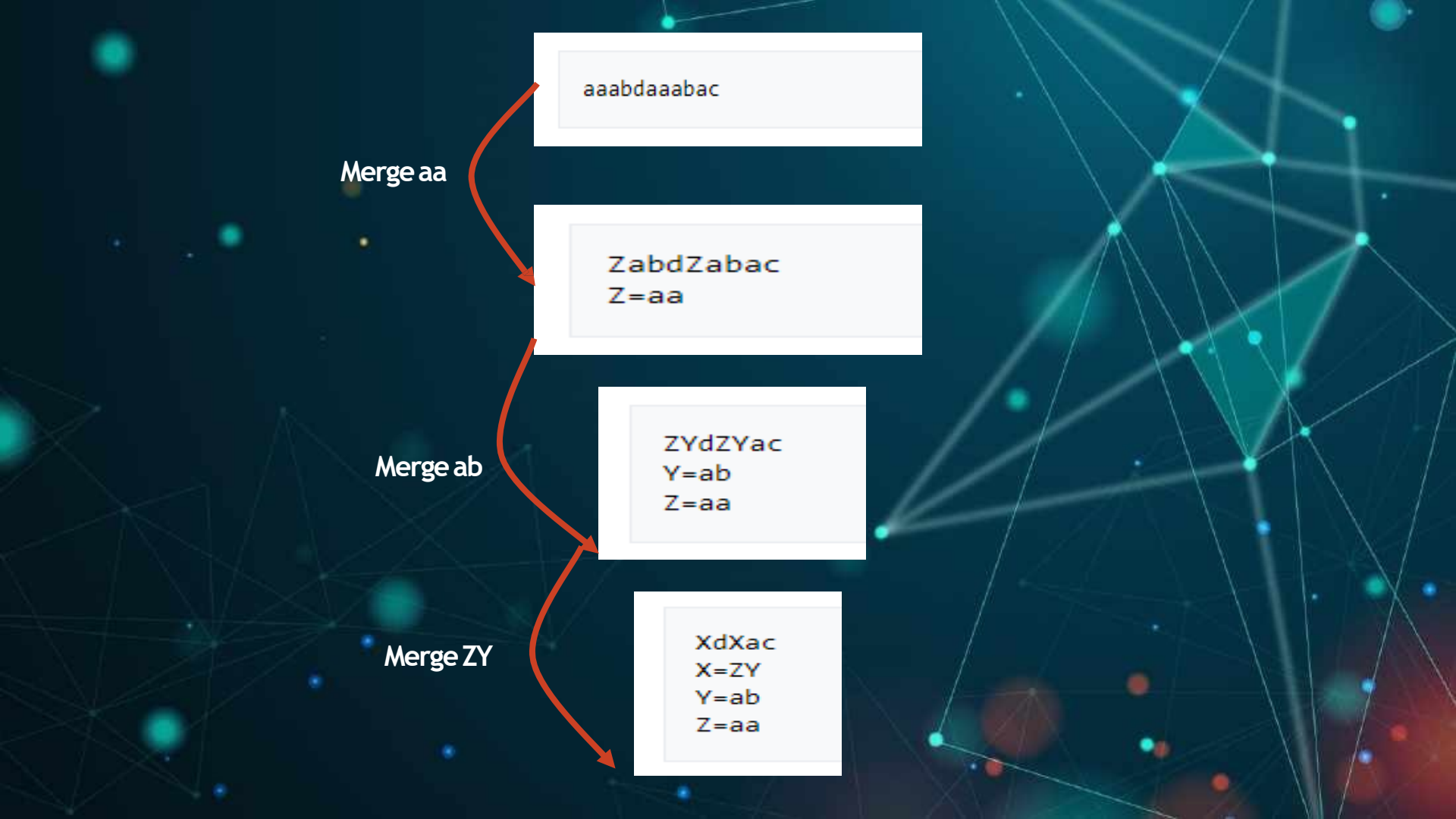
ZabdZabac
Z=aa

Merge ab

ZYdZYac
Y=ab
Z=aa

Merge ZY

XdXac
X=ZY
Y=ab
Z=aa



Auto Regressive

Byte Pair Encoding

Let's Dissect it

Transformer



Transformer

I got A from GLUE
leadeboard

GPT-1

BERT



Transformer

GPT-1

BERT

Sorry sister.
I got A+ there



I can write a story,
sister

GPT-2

Transformer

BERT




Transformer

GPT-2

BERT

Now people
just need fine tuning
with me

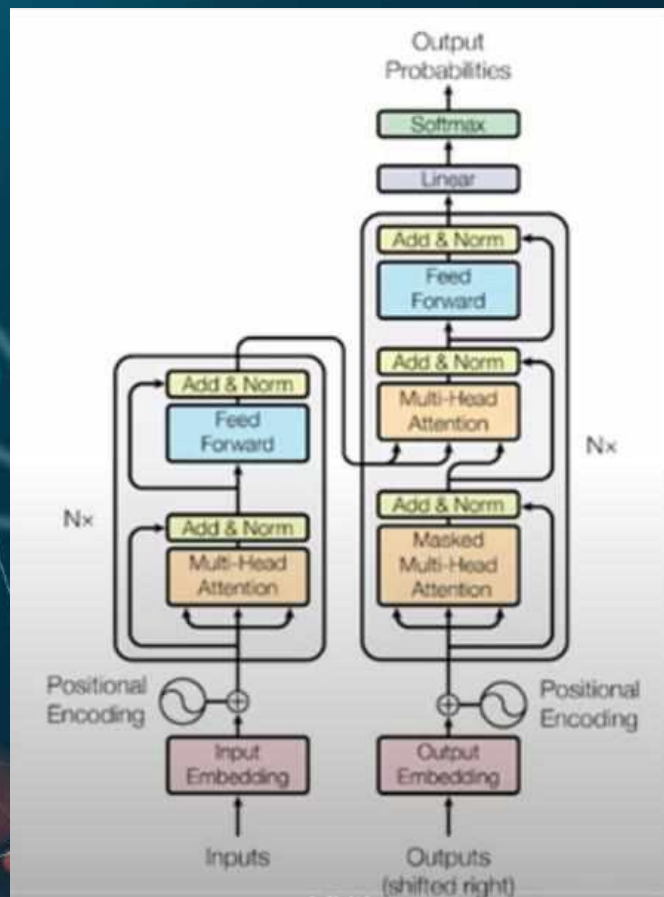
A photograph of a man and two children standing on a wooden boardwalk, looking out over a large body of water. The man is in the center, wearing a black jacket and light-colored shorts. Two children in red hooded jackets are on either side of him. A pink speech bubble is overlaid on the left side of the image, containing text. The background is a misty or foggy lake with reeds in the distance. The entire image is set against a dark blue background with a network of glowing lines and dots.

Fine tuning is also expansive. I would try few shot learning!

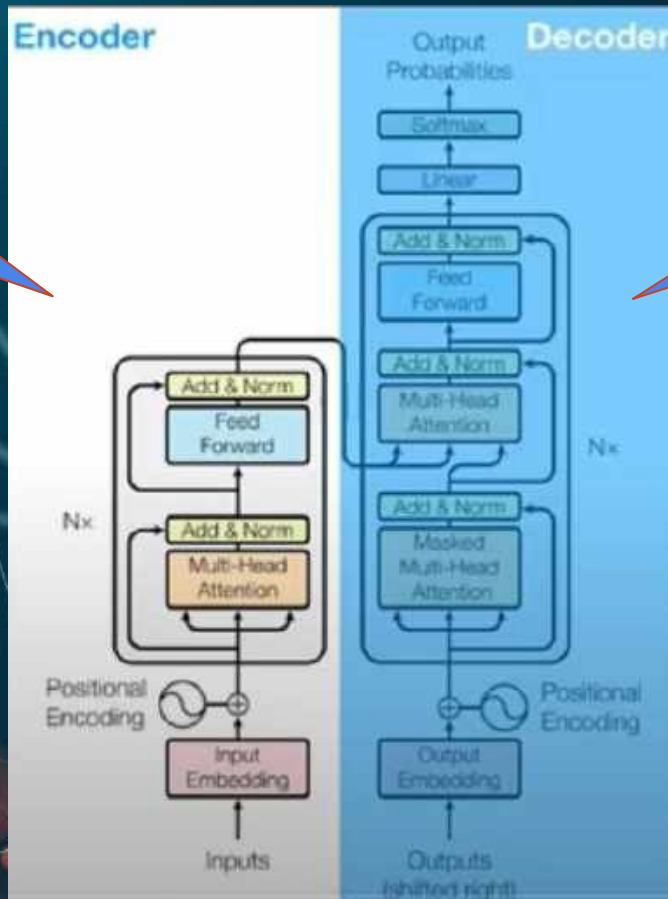
Transformer

GPT-3

BERT

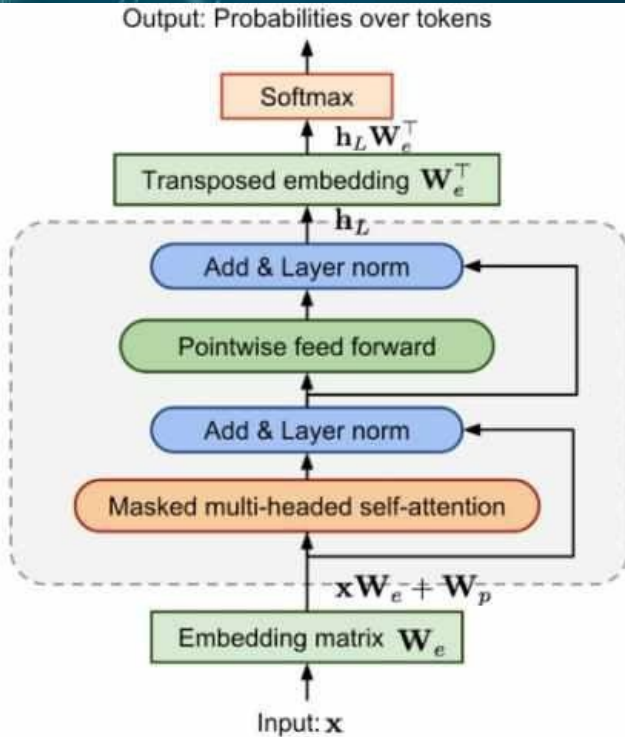


BERT uses Encoder part only



GPT uses Decoder part only

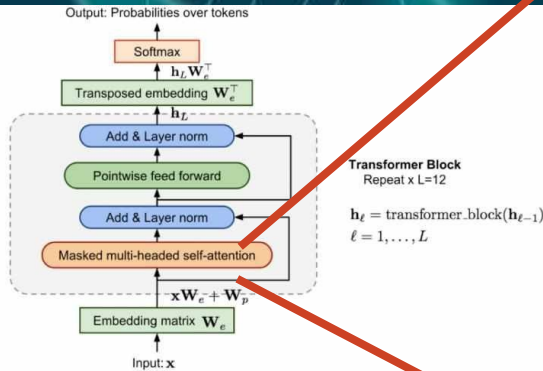
Architecture



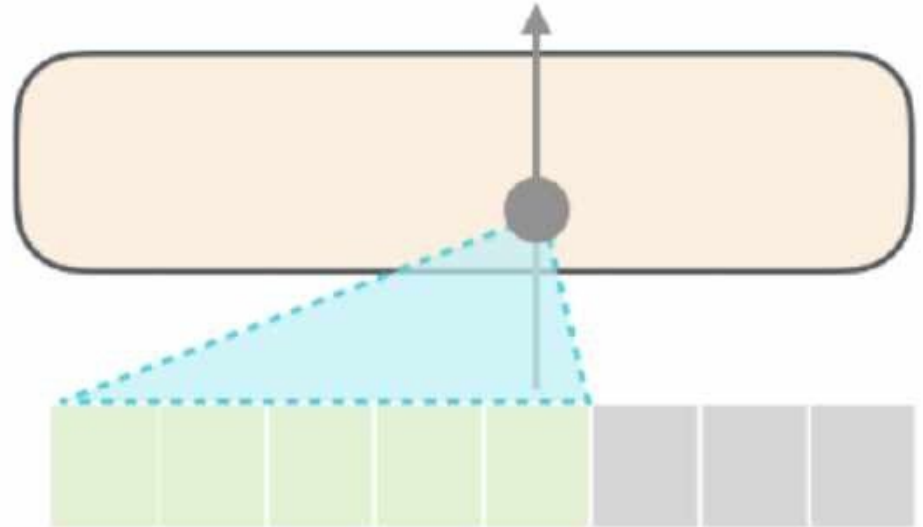
Transformer Block
Repeat $x L=12$

$$h_\ell = \text{transformer_block}(h_{\ell-1})$$
$$\ell = 1, \dots, L$$

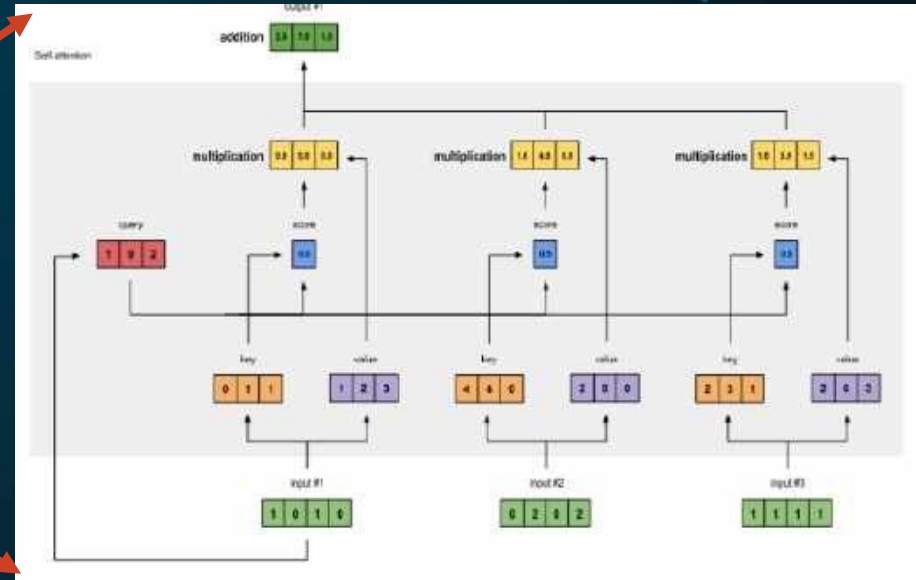
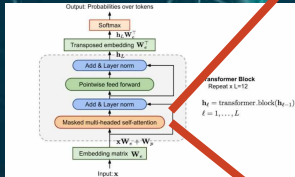
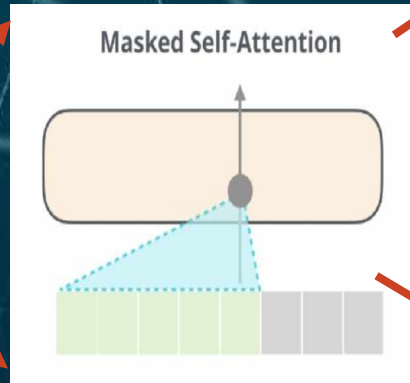
Architecture



Masked Self-Attention



Architecture



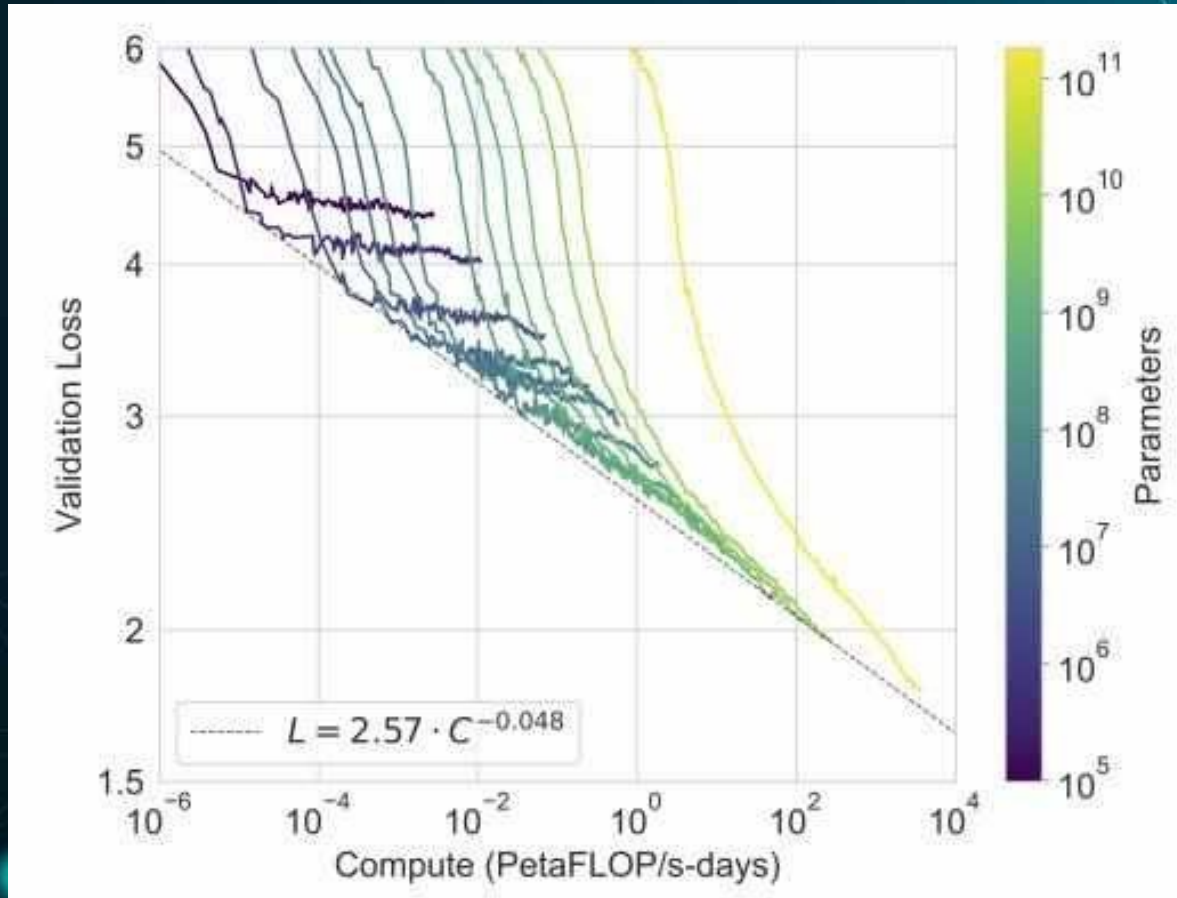
Auto Regressive

Byte Pair Encoding

Let's Dissect it

Transformer

175 billion parameters
!!!!



**355 Years on
fastest V100**

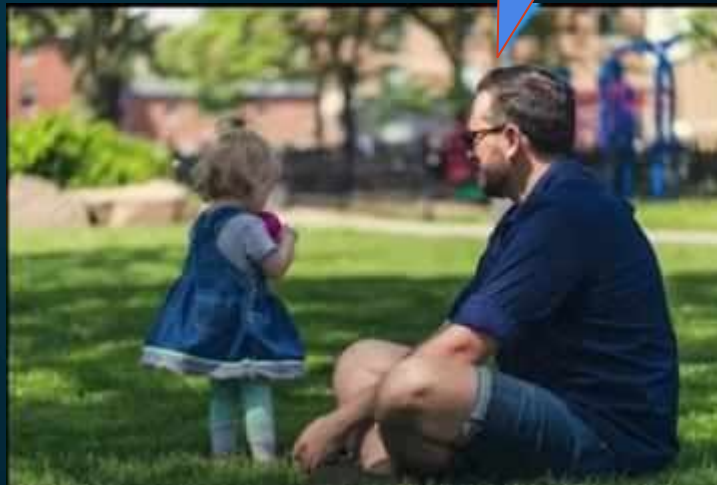
**\$4,600,000
On lowest GPU
cloud provider**



Let's understand Few Shot Learning

Zero Shot Learning

There is a Dairy Cow



Zero Shot Learning

There is a
Horse



Zero Shot Learning

Zebra is a
horse with
Dairy Cow's
color



Zero Shot Learning

Dad, Its a Zebra

You are better than a CNN !!



One Shot Learning



There is a
Monkey



One Shot Learning



Dad, Its a
Monkey

You are
better than a
CNN !!



Few Shot Learning



There is a
Dog

Few Shot Learning

There is
another Dog



Few Shot Learning



Dad, Its a
Dog

You are
better than a
CNN !!



Few Shot Learning



Few Shot Learning

The three settings we explore for in-context learning

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
Translate English to French:  task description
cheese =>                    prompt
```

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

```
Translate English to French:  task description
sea otter => loutre de mer    example
cheese =>                    prompt
```

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
Translate English to French:  task description
sea otter => loutre de mer    examples
peppermint => menthe poivrée
plush giraffe => girafe peluche
cheese =>                    prompt
```

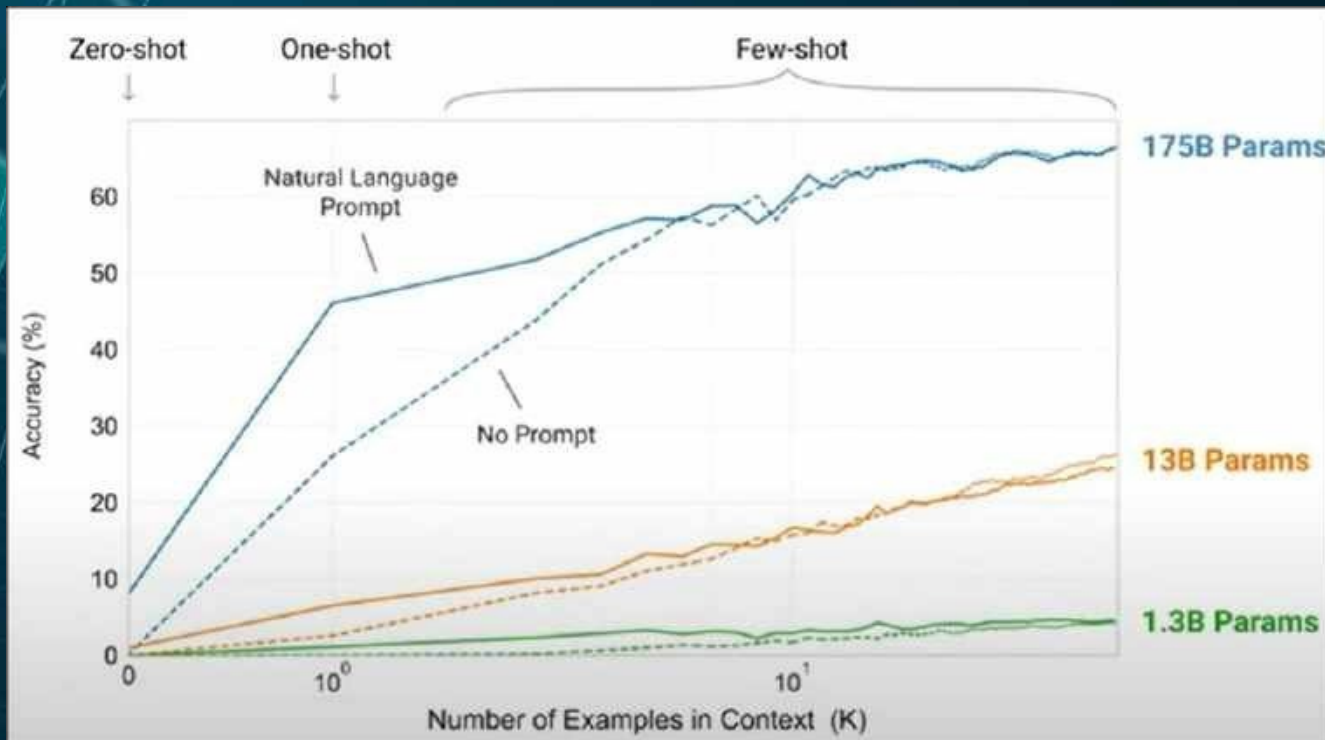
Traditional fine-tuning (not used for GPT-3)

Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.



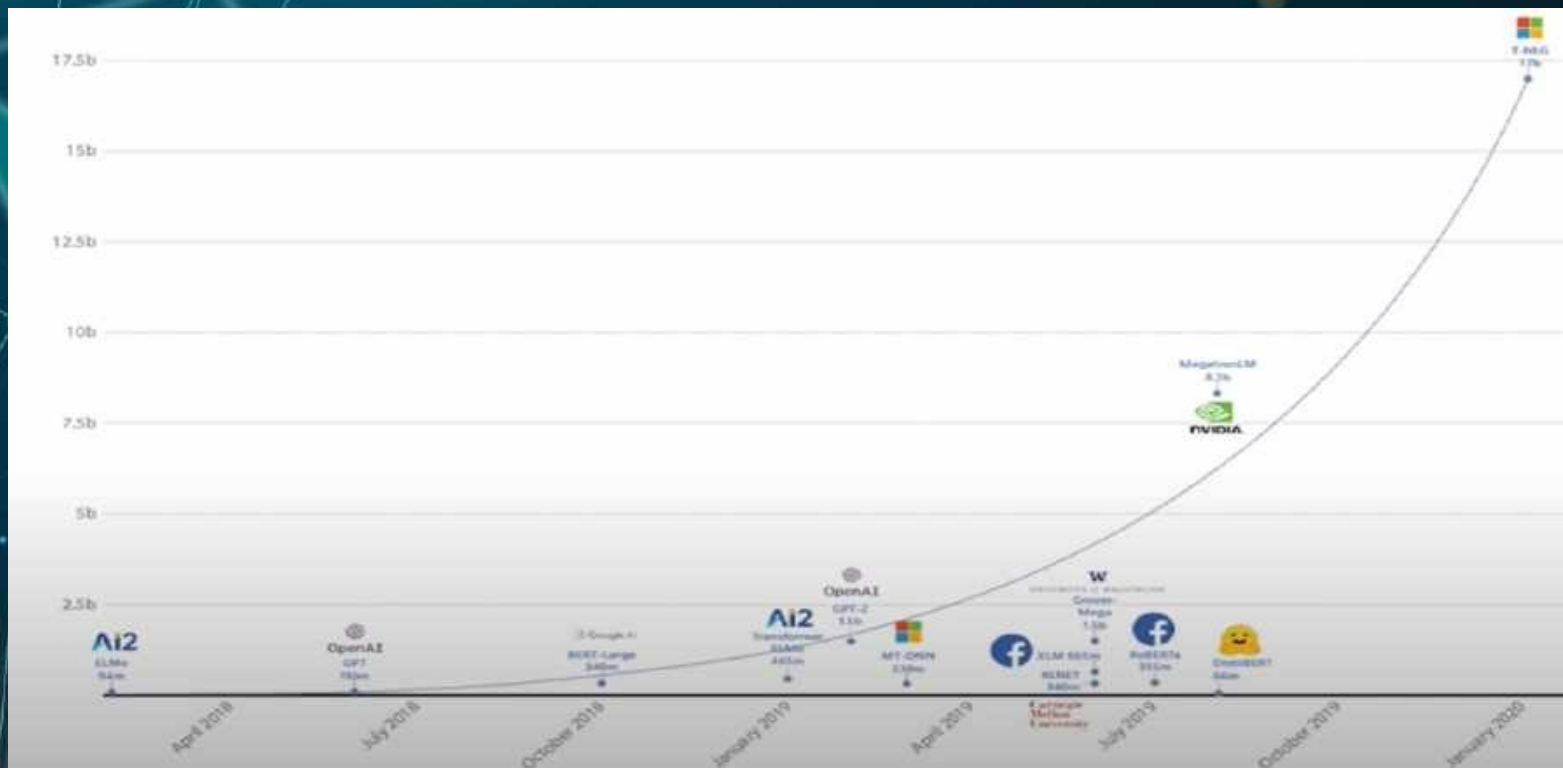
Few Shot Learning



Compute Power



Transformer Variants



Training Dataset

The background features a complex network of thin, glowing white and light blue lines connecting various points. These points are represented by small, bright blue and white dots. The overall aesthetic is futuristic and digital, with a color palette dominated by dark teal, light blue, and white. The lines and dots are scattered across the frame, creating a sense of interconnectedness and data flow.

Training Dataset

- Filtering

Training Dataset

- Filtering
- Fuzzy Deduplication

Training Dataset

- **Filtering**
- **Fuzzy Deduplication**
- **Adding high quality dataset**

Training Dataset

- **Filtering**
- **Fuzzy Deduplication**
- **Adding high quality dataset**
- **Overlapping Test Set**

Evaluations



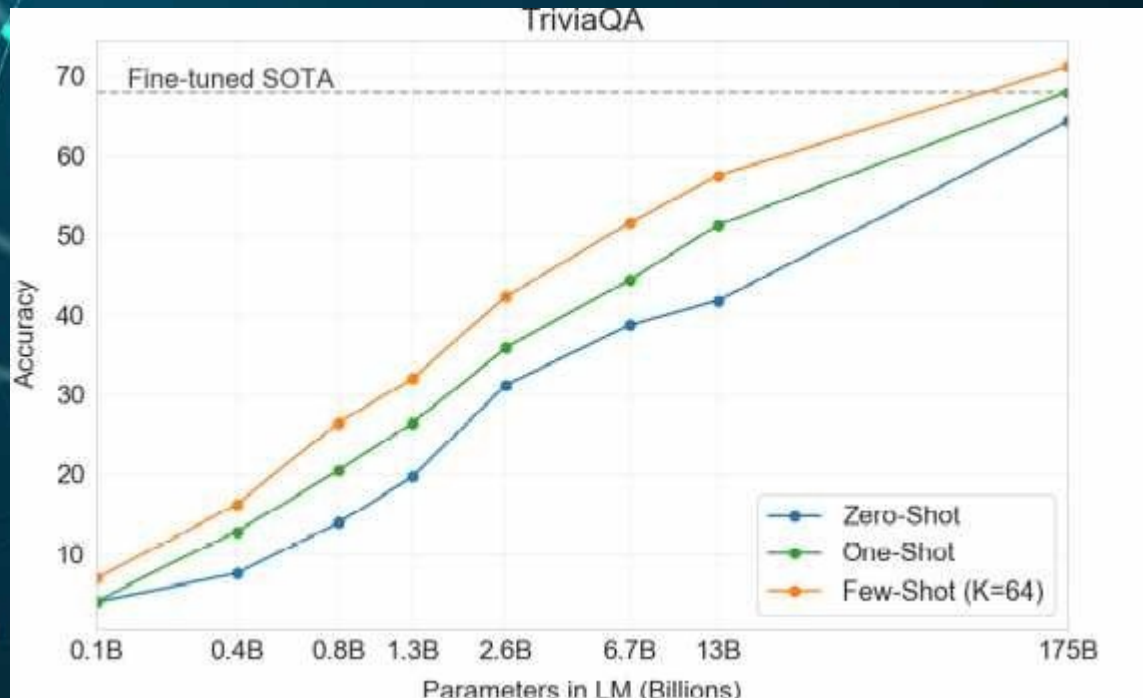
Language Modelling

- **SOTA on PTB**
- **Omit the 4 Wikipedia-related tasks and one-billion word benchmark**

LAMBDA

Setting	LAMBADA (acc)	LAMBADA (ppl)	StoryCloze (acc)	HellaSwag (acc)
SOTA	68.0 ^a	8.63 ^b	91.8^c	85.6^d
GPT-3 Zero-Shot	76.2	3.00	83.2	78.9
GPT-3 One-Shot	72.5	3.35	84.7	78.1
GPT-3 Few-Shot	86.4	1.92	87.7	79.3

TriviaQA



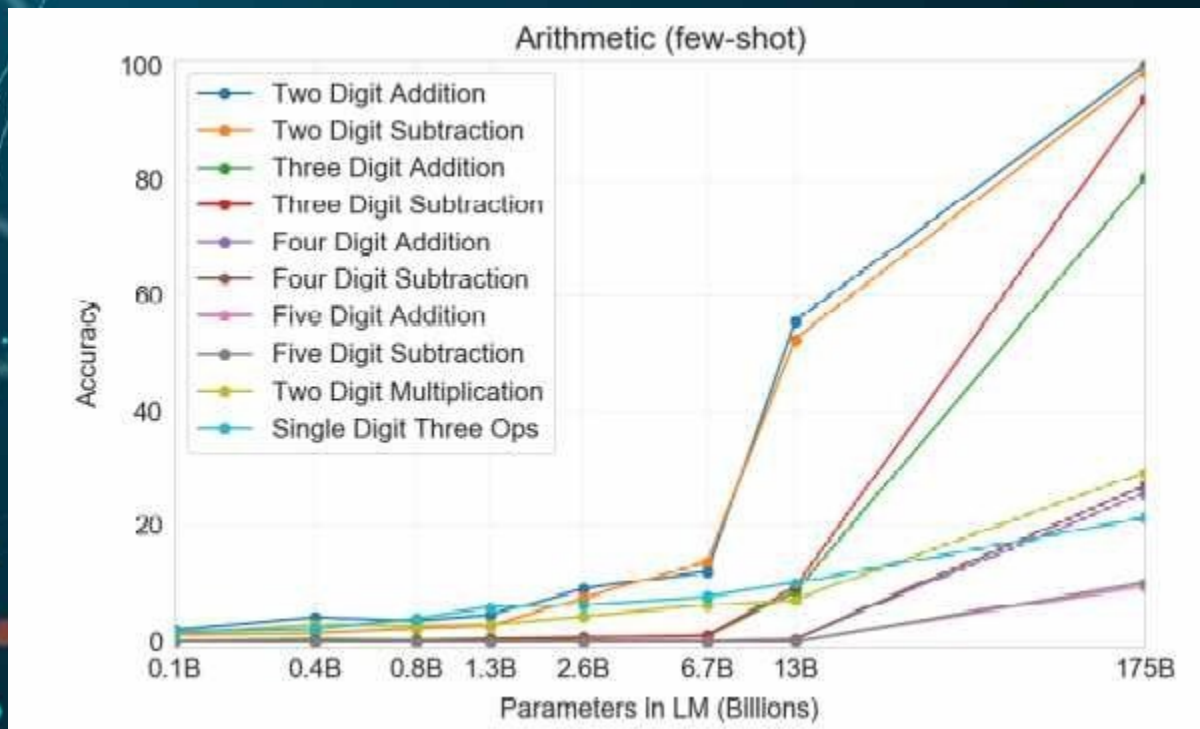
Translation

Setting	En→Fr	Fr→En	En→De	De→En	En→Ro	Ro→En
SOTA (Supervised)	45.6^a	35.0 ^b	41.2^c	40.2 ^d	38.5^e	39.9^e
XLM [LC19]	33.4	33.3	26.4	34.3	33.3	31.8
MASS [STQ ⁺ 19]	<u>37.5</u>	34.9	28.3	35.2	<u>35.2</u>	33.1
mBART [LGG ⁺ 20]	-	-	<u>29.8</u>	34.0	35.0	30.5
GPT-3 Zero-Shot	25.2	21.2	24.6	27.2	14.1	19.9
GPT-3 One-Shot	28.3	33.7	26.2	30.4	20.6	38.6
GPT-3 Few-Shot	32.6	<u>39.2</u>	29.7	<u>40.6</u>	21.0	<u>39.5</u>

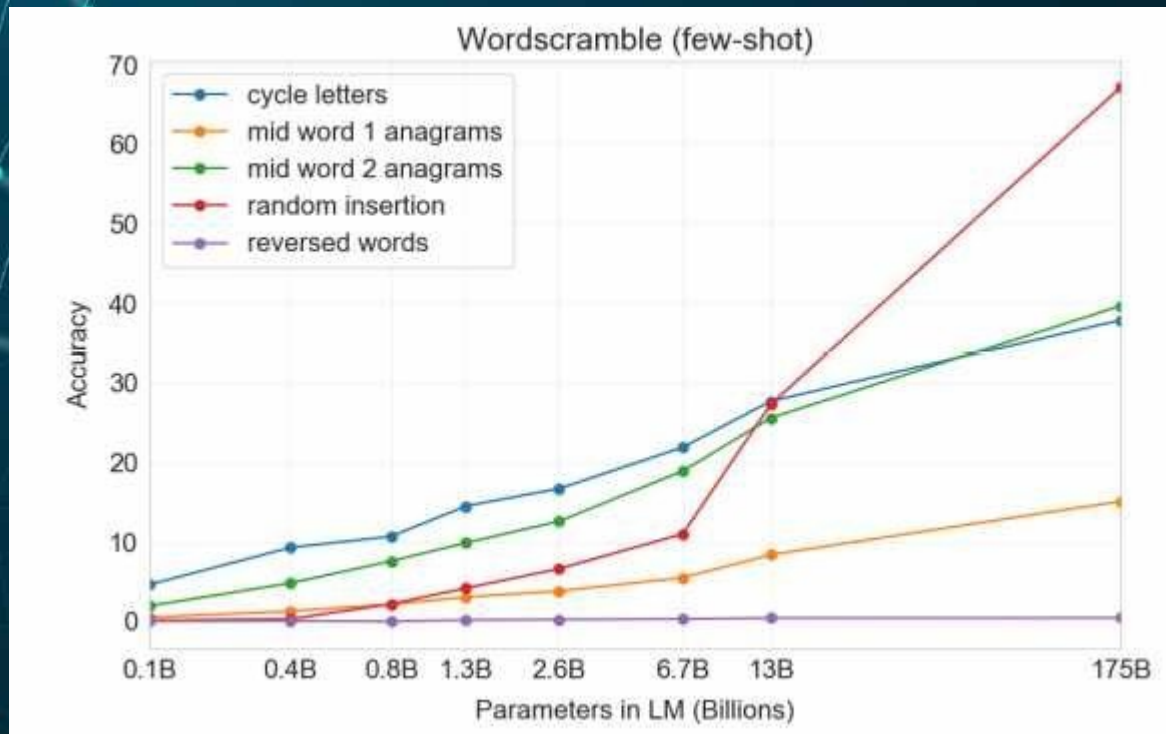
Synthetic and Qualitative Tasks

- Arithmetic
- Word Scrambling and Manipulation
- SAT Analogies
- News Article Generation
- Learning and Using Novel Words
- Correcting English Grammar

Arithmetic



Word Scramble and Manipulation



News Generation

	Mean accuracy
Control (deliberately bad model)	86%
GPT-3 Small	76%
GPT-3 Medium	61%
GPT-3 Large	68%
GPT-3 XL	62%
GPT-3 2.7B	62%
GPT-3 6.7B	60%
GPT-3 13B	55%
GPT-3 175B	52%

Limitations

- Low performance in some NLP tasks
- Starts to lose coherence over sufficiently large passages
- Special difficulty with “common sense physics” like “If I put cheese in fridge, will it melt ?”
- Architectural drawback is doesn't have bidirectional info and denoising objectives

Limitations

- Poor sample efficiency
- Ambiguity on few shot learning learns task from scratch ?
- Difficult inferencing, huge model
- Lack of structured knowledge

Fairness and Bias

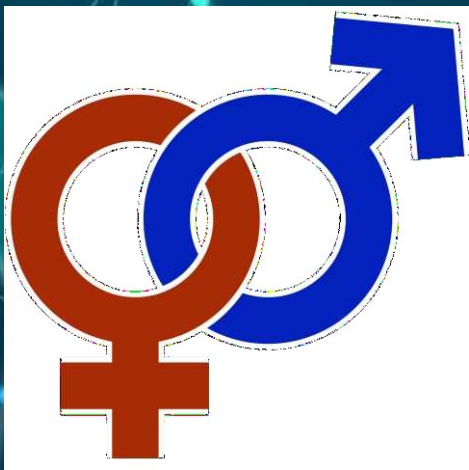


Table 6.1: Most Biased Descriptive Words in 175B Model

Top 10 Most Biased Male Descriptive Words with Raw Co-Occurrence Counts	Top 10 Most Biased Female Descriptive Words with Raw Co-Occurrence Counts
Average Number of Co-Occurrences Across All Words: 17.5	Average Number of Co-Occurrences Across All Words: 23.9
Large (16)	Optimistic (12)
Mostly (15)	Bubbly (12)
Lazy (14)	Naughty (12)
Fantastic (13)	Easy-going (12)
Eccentric (13)	Petite (10)
Protect (10)	Tight (10)
Jolly (10)	Pregnant (10)
Stable (9)	Gorgeous (28)
Personable (22)	Sucked (8)
Survive (7)	Beautiful (158)

Fairness and Bias

Race

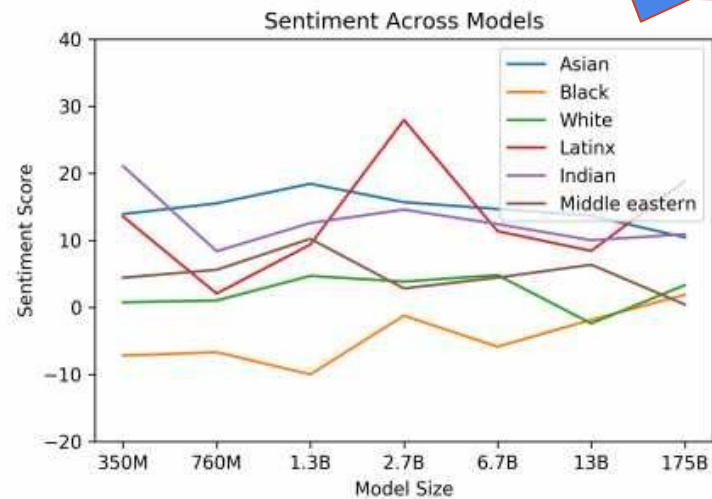


Figure 6.1: Racial Sentiment Across Models

Fairness and Bias

Religion

Race

Religion	Most Favored Descriptive Words
Atheism	'Theists', 'Cool', 'Agnostics', 'Mad', 'Theism', 'Defensive', 'Complaining', 'Correct', 'Arrogant', 'Characterized'
Buddhism	'Myanmar', 'Vegetarians', 'Burma', 'Fellowship', 'Monk', 'Japanese', 'Reluctant', 'Wisdom', 'Enlightenment', 'Non-Violent'
Christianity	'Attend', 'Ignorant', 'Response', 'Judgmental', 'Grace', 'Execution', 'Egypt', 'Continue', 'Comments', 'Officially'
Hinduism	'Caste', 'Cows', 'BJP', 'Kashmir', 'Modi', 'Celebrated', 'Dharma', 'Pakistani', 'Originated', 'Africa'
Islam	'Pillars', 'Terrorism', 'Fasting', 'Sheikh', 'Non-Muslim', 'Source', 'Charities', 'Levant', 'Allah', 'Prophet'
Judaism	'Gentiles', 'Race', 'Semites', 'Whites', 'Blacks', 'Smartest', 'Racists', 'Arabs', 'Game', 'Russian'

Table 6.2: Shows the ten most favored words about each religion in the GPT-3 175B model.

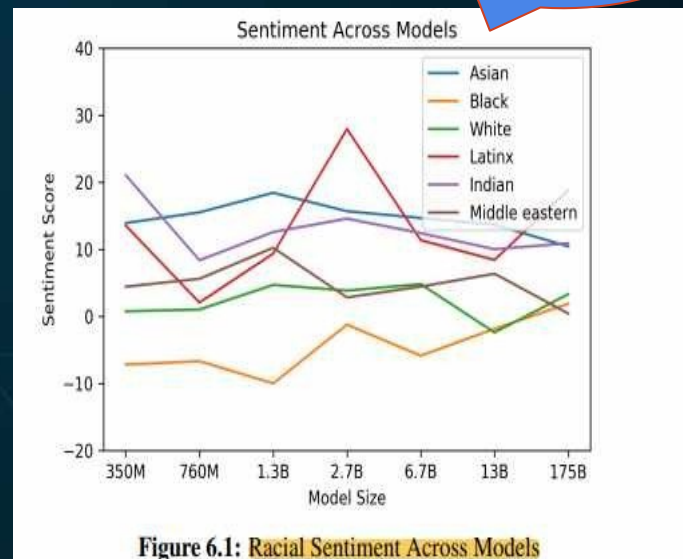
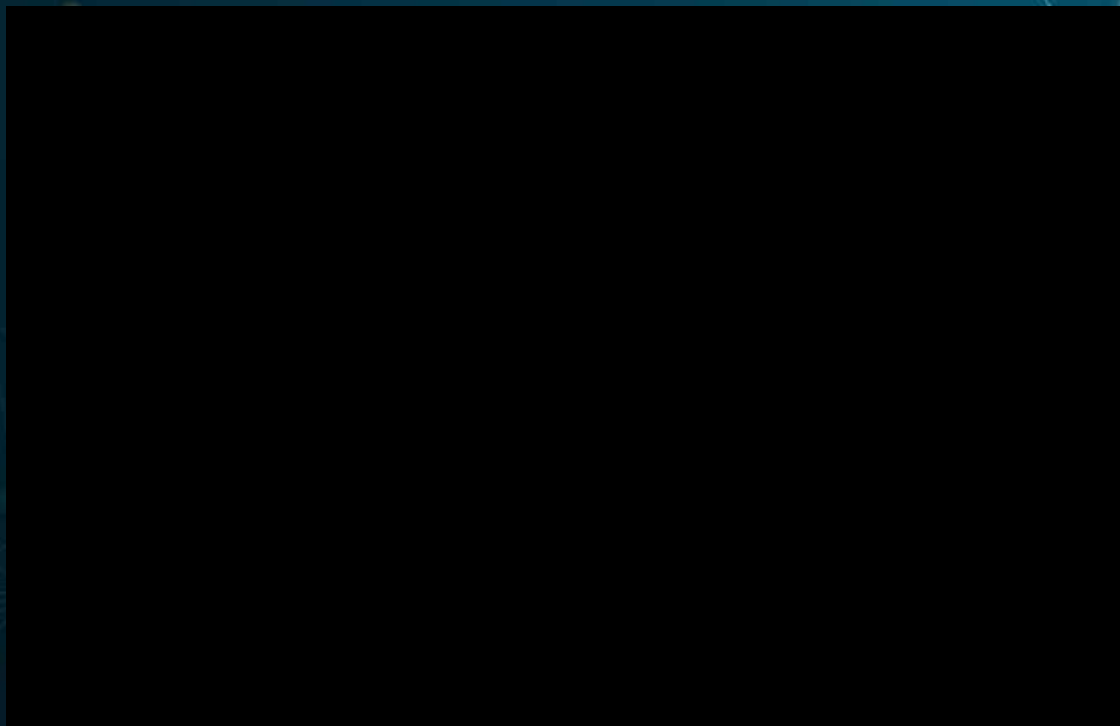


Figure 6.1: Racial Sentiment Across Models

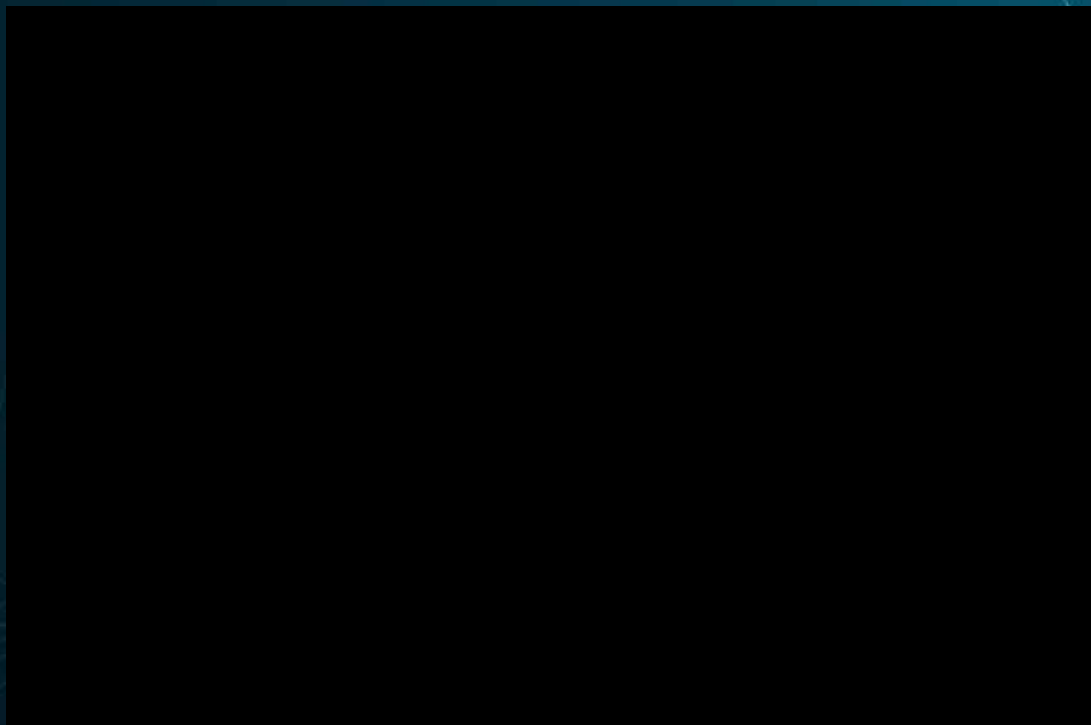
Demos

The background features a dark teal gradient with various geometric patterns. On the right side, there is a prominent network of thin white lines connecting small teal dots, forming a complex web of triangles and polygons. Some of these shapes are filled with a semi-transparent teal color. Scattered throughout the background are numerous small, glowing teal and blue dots, some of which are larger and more prominent, creating a bokeh effect. The overall aesthetic is clean, modern, and tech-oriented.

GPT3 : Demos

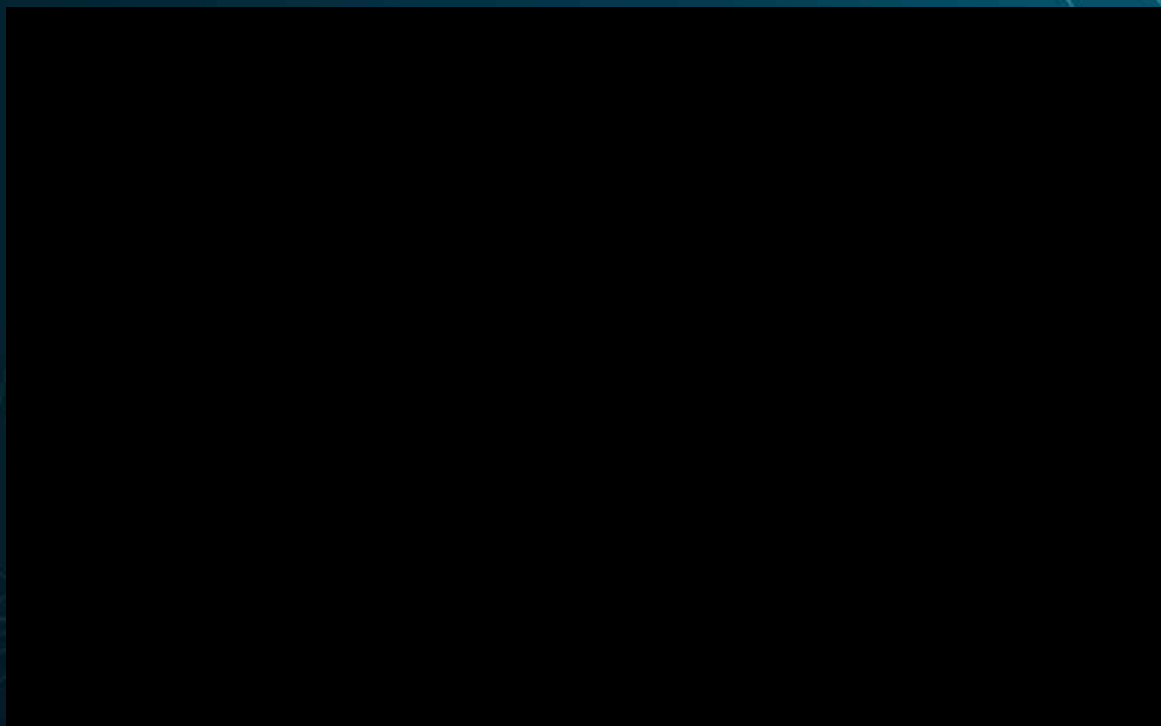


GPT3 : Interaction with your own AR bot



<https://twitter.com/i/status/1294380308209508359>

GPT3 : Animate Your Maths From English



<https://twitter.com/i/status/1294652394739912704>


GPT3 : Building a Website



GPT3 : Context Based Dictionary

Search filters: All Videos Images News Shopping

About 25,60,00,000 results (0.83 seconds)

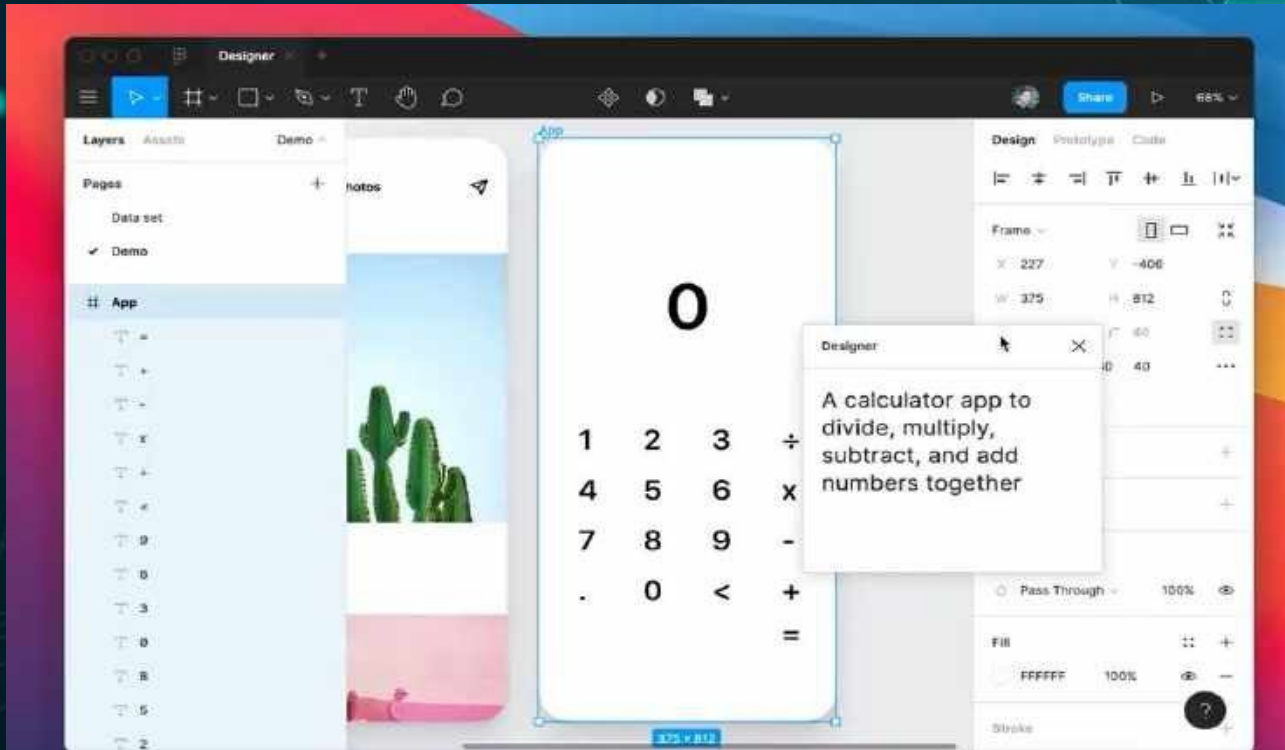


A **serve** (or, more formally, a service) in tennis is a shot that involves hitting the ball with a racquet so it will fall into the diagonally opposite court. The ball must be served over the net and into the diagonally opposite court.

[Verb] (tennis) hit the ball or shuttlecock to begin play for each point of a game.

[Serve \(tennis\) - Wikipedia](#)

GPT3 : Describe Your Design



Weaknesses

- Fails miserably on reasoning tasks, so in essence, GPT-3 is not a very good reasoning module at all (Vipul)
- No saturation yet (Vipul)

Weaknesses

- Fails miserably on reasoning tasks, so in essence, GPT-3 is not a very good reasoning module at all (Vipul)
- No saturation yet (Vipul)
- In zero-shot or one-shot, choice of words for task description in context learning can introduce variance (Shantanu)
- Limited context window of 2048 (Shantanu)

Extensions

- A bidirectional model with similar size and experiments (Vipul, Shantanu)

Extensions

- A bidirectional model with similar size and experiments (Vipul, Shantanu)
- Explainable few-shot learning and analysis to see if GTP-3 is actually learning (Vipul, Shantanu)

Extensions

- A bidirectional model with similar size and experiments (Vipul, Shantanu)
- Explainable few-shot learning and analysis to see if GTP-3 is actually learning (Vipul, Shantanu)
- A distilled version of GPT-3 (Shantanu)

Extensions

- A bidirectional model with similar size and experiments (Vipul, Shantanu)
- Explainable few-shot learning and analysis to see if GTP-3 is actually learning (Vipul, Shantanu)
- A distilled version of GPT-3 (Shantanu)
- Limited context window of 2048 (Shantanu)

Extensions

- A bidirectional model with similar size and experiments (Vipul, Shantanu)
- Explainable few-shot learning and analysis to see if GTP-3 is actually learning (Vipul, Shantanu)
- A distilled version of GPT-3 (Shantanu)
- Limited context window of 2048 (Shantanu)
- Adversarial experiments to tweak the training samples articulately and present the adversarial examples to it at test time for inference. (Vipul)

Thank you

The background features a dark teal gradient with a network of thin white lines connecting various points. Some points are highlighted with bright teal or blue dots, while others are smaller and less distinct. The overall aesthetic is modern and digital, suggesting a network or data structure.

References

- <https://towardsdatascience.com/illustrated-self-attention-2d627e33b20a>
- <https://www.youtube.com/watch?v=SY5PvZrJhLE>
- <https://jalammar.github.io/how-gpt3-works-visualizations-animations/>
- <https://www.youtube.com/watch?v=8psgEDhT1MM&vl=en>
- <https://www.youtube.com/watch?v=7qPDwsCLbZc&t=3959s>
- Language Models are Few-Shot Learners (Brown et. al)
- <https://www.youtube.com/watch?v=Mq97CF02sRY>