

Where are Self-Supervised Models Headed?

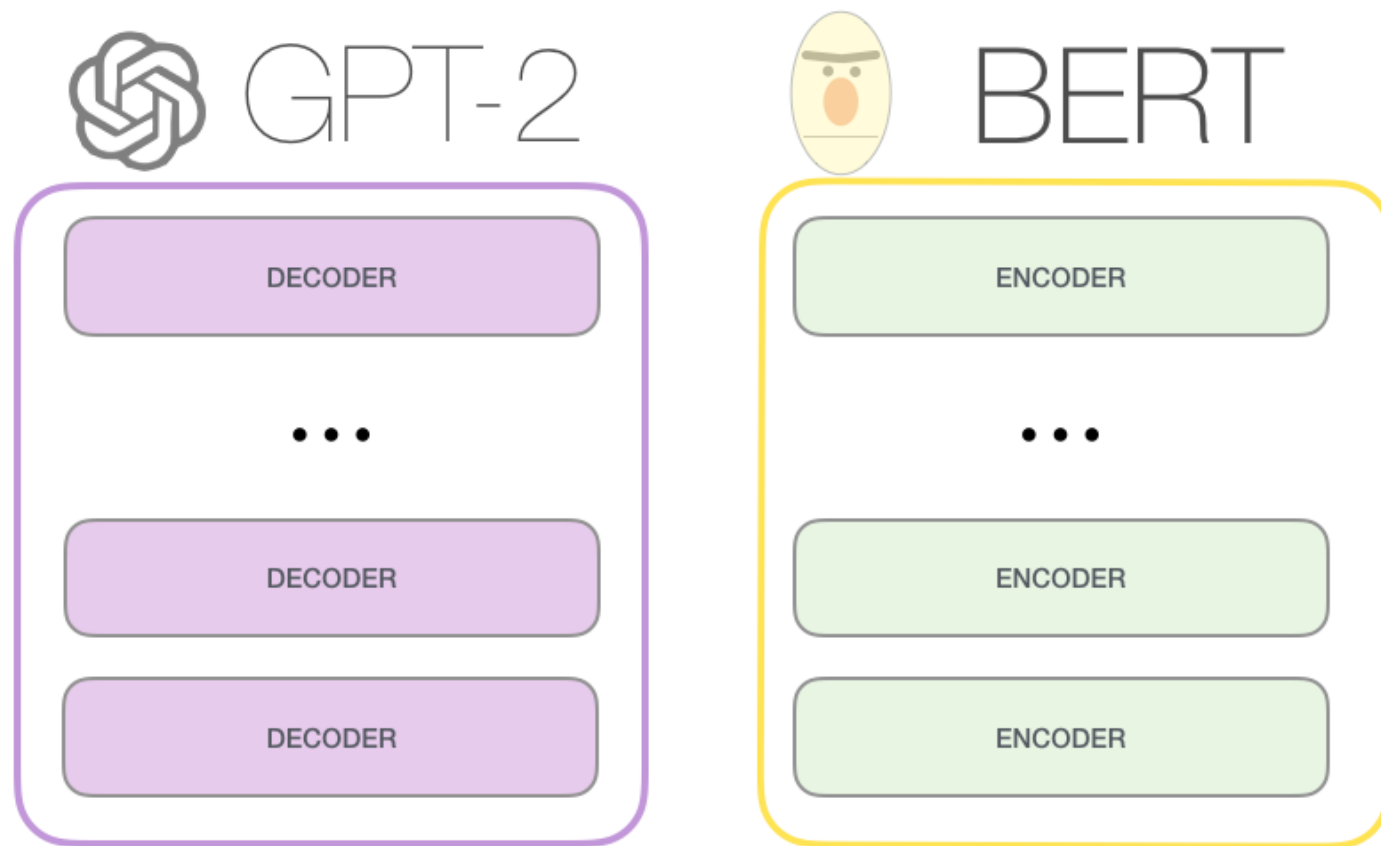
Daniel Khashabi



Self-Supervised Models

are *predictive models* of the world!

Large Language Models



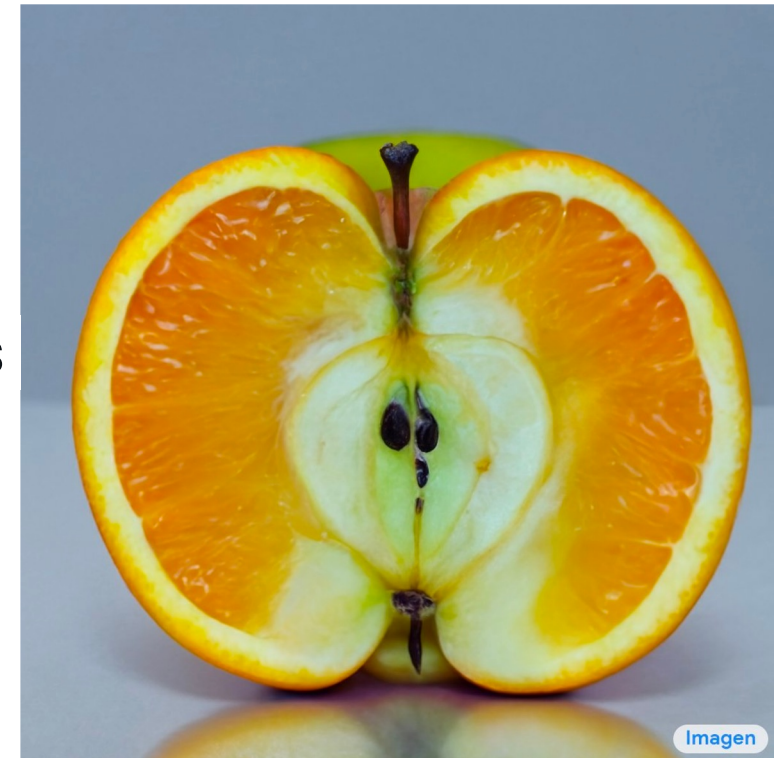
[Bengio et al. '04, Peters et al. '18, Raffel et al. '20, Brown et al. '20, many others]

Image-Text Models



an invisible man, wearing glasses and sitting at a desk in front of a computer"

one piece of fruit that's apple on the outside, orange texture on the inside, cut in half



<https://imagen.research.google/>

<https://openai.com/dall-e-2/>

Scaling Self-Supervised Models

- Larger pre-training datasets

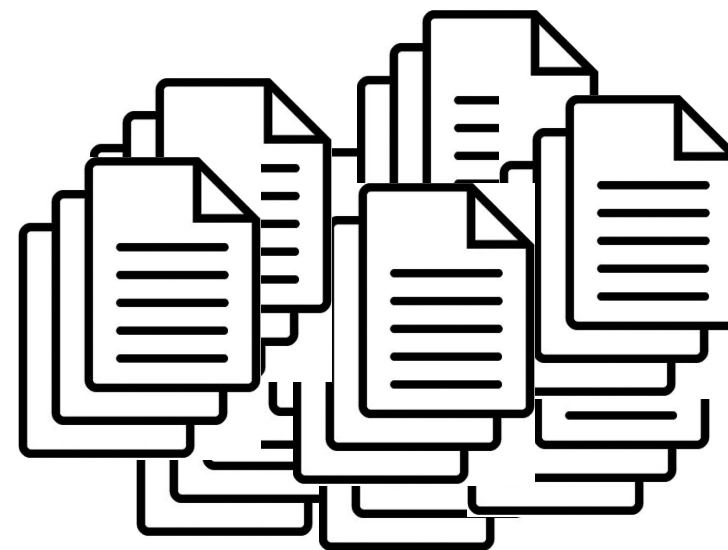
*cheap freely-
available text*



1GB



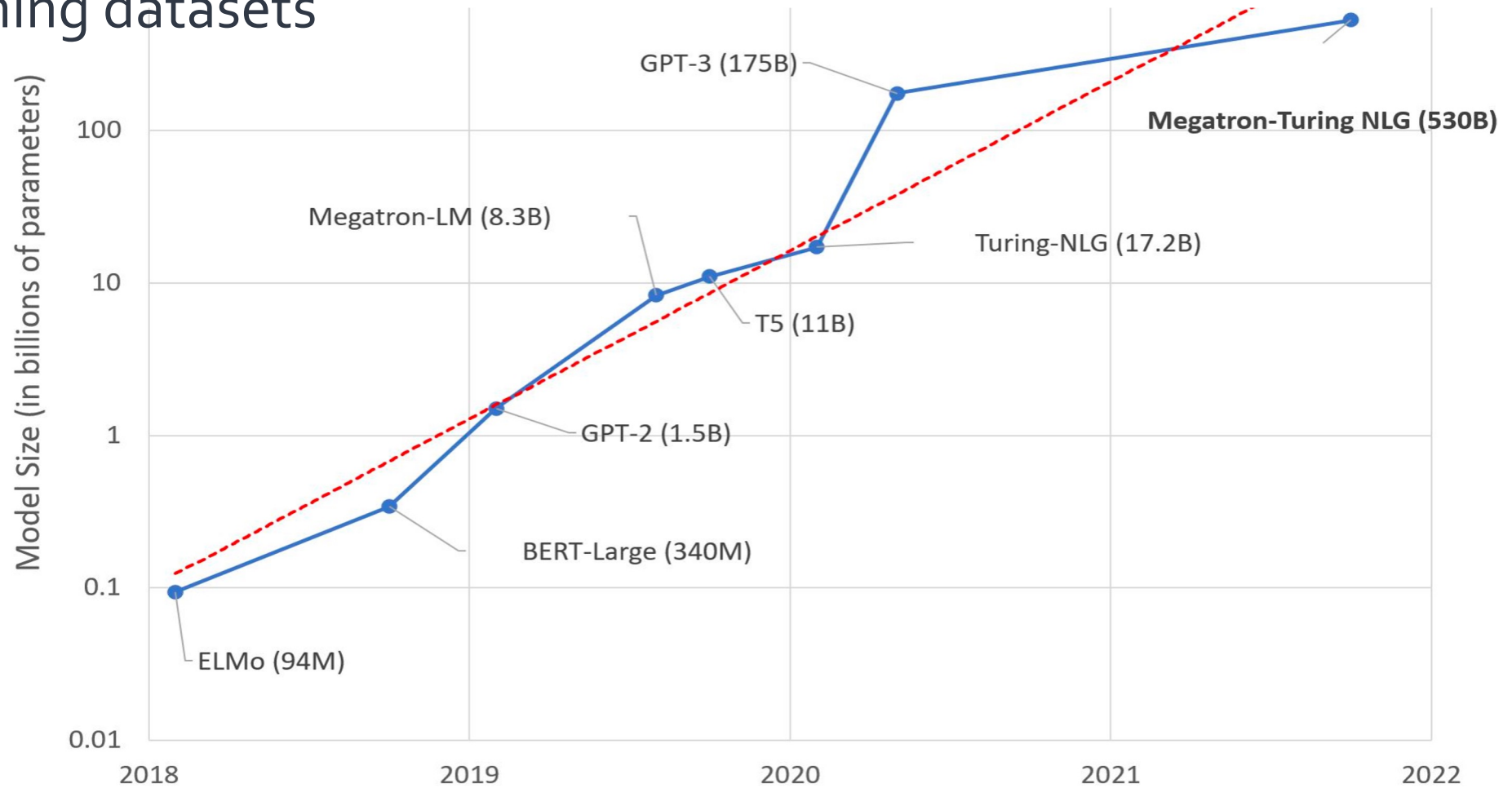
10GB



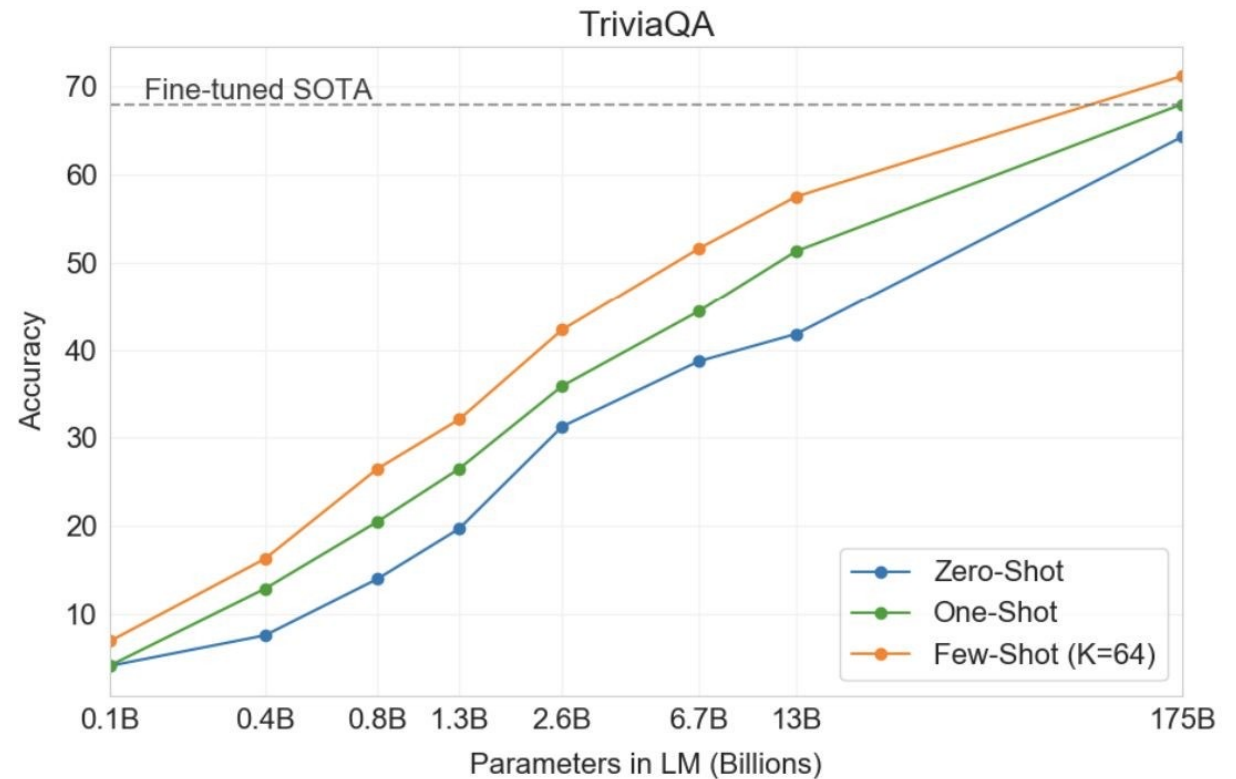
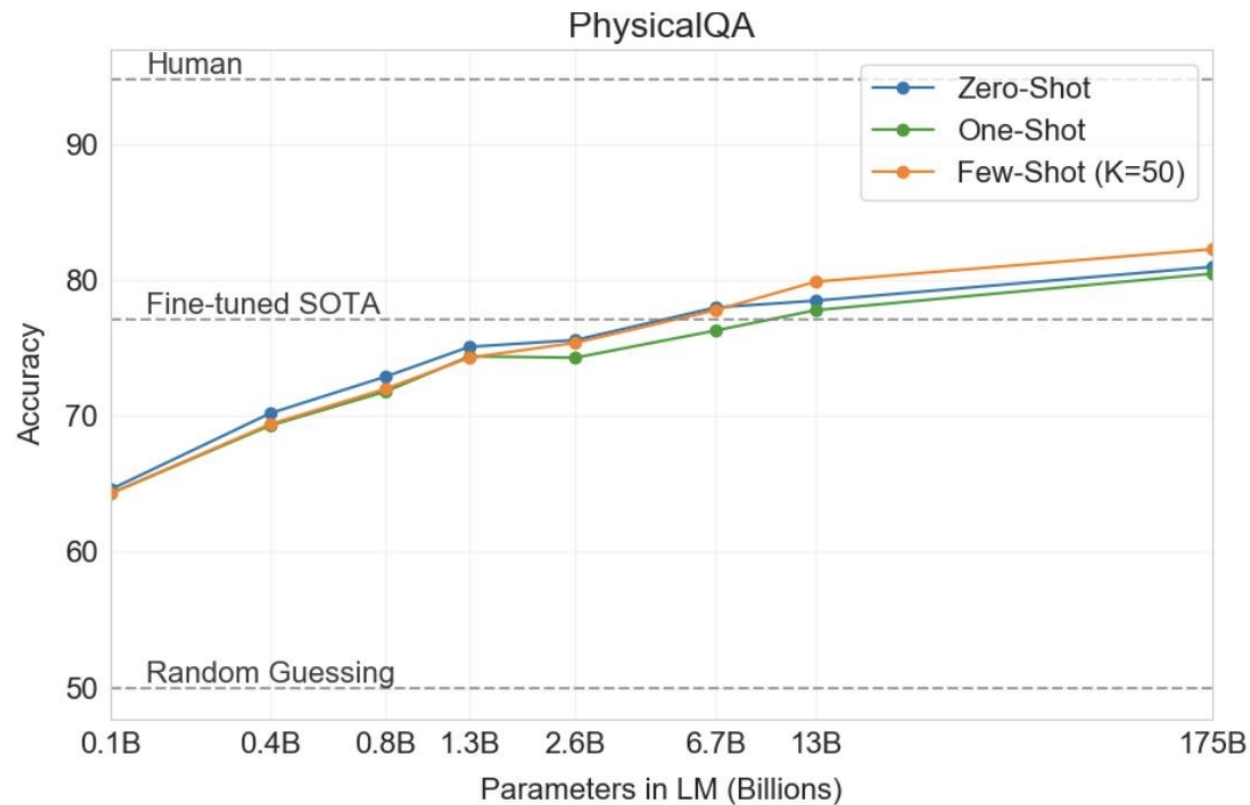
10TB

Scaling Self-Supervised Models

- Larger pre-training datasets
- Larger models



Scaling Self-Supervised Models



ChatGPT: Optimizing Language Models for Dialogue

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.



Do you think a magnet could attract a penny?



Yes, a magnet could potentially attract a penny if the penny is made of a magnetic material. However, pennies made in the United States after 1982 are made primarily of zinc, which is not magnetic. Therefore, a magnet would not be able to attract a penny made after 1982.



What if it's made of copper?



Copper is not a magnetic material, so a magnet would not be able to attract a penny made of copper. Copper is a good conductor of electricity, which is why it is often used in electrical wiring. However, it does not have magnetic properties.





To factor the polynomial $x^2 + 35x - 7$, we need to find two numbers that multiply to -7 and add to 35 . The factors of -7 are -1 and 7 , so we can write the polynomial as:

$$x^2 + 35x - 7 = (x - 1)(x + 7)$$

We can verify that this is the correct solution by multiplying the two factors:

$$(x - 1)(x + 7) = x^2 + 7x - x - 7 = x^2 + 35x - 7$$

Therefore, the factored form of the polynomial is $(x - 1)(x + 7)$.

"Write George Washington's farewell address as if it was delivered by a pirate"



Ahoy mateys! George Washington here, former captain of the good ship America. I come before ye today to bid ye farewell, as I step down from me post as captain of this great nation.

As ye know, I have been at the helm of this ship for many years, steering it through rough waters and guiding it towards a bright future. But now, it be time for me to hand over the reins to a new captain and retire to me cabin.

But before I do, I wanted to share with ye some wisdom that I have gathered over the years, in the hopes that it will help ye navigate the treacherous seas ahead.

First and foremost, I urge ye to always be vigilant in the defense of our great nation. We have fought hard to earn our freedom and independence, and we must never let it slip through our fingers. Keep a watchful eye on those who would seek to do us harm, and be ready to defend our shores at a moment's notice.

Second, I implore ye to always be mindful of the welfare of our fellow shipmates. We are stronger when we work together, and we must always strive to ensure that all members of our crew are treated fairly and with respect.


VI

Can you fix the bug in this code snippet?

```
# check if two strings are anagrams
def is_anagram(a,b):
    for char in a:
        if char not in b:
            return False
    return True
```



There are a few problems with the code snippet you provided. First, it only checks if all the characters in the first string `a` are also in the second string `b`, but it does not check if the strings have the same length or if the number of occurrences of each character is the same in both strings. To fix these issues, we can use the following updated code:

 Copy code

```
# check if two strings are anagrams
def is_anagram(a,b):
    # check if the strings have the same length
    if len(a) != len(b):
        return False

    # create dictionaries to store the number of occurrences of each
    # character in the strings
    a_count = {}
    b_count = {}

    # count the number of occurrences of each character in the first
```

```
        a_count[char] = 1

    # count the number of occurrences of each character in the second
    # string
    for char in b:
        if char in b_count:
            b_count[char] += 1
        else:
            b_count[char] = 1

    # check if the number of occurrences of each character is the same in
    # both strings
    for char in a_count:
        if char not in b_count or a_count[char] != b_count[char]:
            return False

    # if all checks pass, the strings are anagrams
    return True
```

This updated code first checks if the two strings have the same length. If not, the strings cannot be anagrams and the function returns `False`. Next, it creates two dictionaries to store the number of occurrences of each character in the two strings. It then counts the number of occurrences of each character in each string, and finally checks if the number of occurrences of each character is the same in both strings. If all these checks pass, the function returns `True`, indicating that the two strings are anagrams.

Is Scale All We Need?

1. Is scale the/a right “hill to climb”?
2. Even if it is a right “hill” is it feasible/practical to climb this hill? (there might be other “hills” too).



Nando de Freitas 🇧🇷
@NandoDF



Someone's opinion article. My opinion: It's all about scale now! The Game is Over! It's about making these models bigger, safer, compute efficient, faster at sampling, smarter memory, more modalities, INNOVATIVE DATA, on/offline, ... 1/N



thenextweb.com

DeepMind's new Gato AI makes me fear humans will never achieve AGI
DeepMind just unveiled a new AI system called Gato that makes OpenAI's GPT-3 look like a child's toy. But are we any closer to AGI?

4:46 AM · May 14, 2022 · Twitter for iPhone

233 Retweets 114 Quote Tweets 885 Likes



Nando de Freitas 🇧🇷
@NandoDF



Solving these scaling challenges is what will deliver AGI. Research focused on these problems, eg S4 for greater memory, is needed. Philosophy about symbols isn't. Symbols are tools in the world and big nets have no issue creating them and manipulating them 2/n

4:50 AM · May 14, 2022 · Twitter for iPhone

23 Retweets 5 Quote Tweets 153 Likes

Let's do a poll!

What is “Scale”

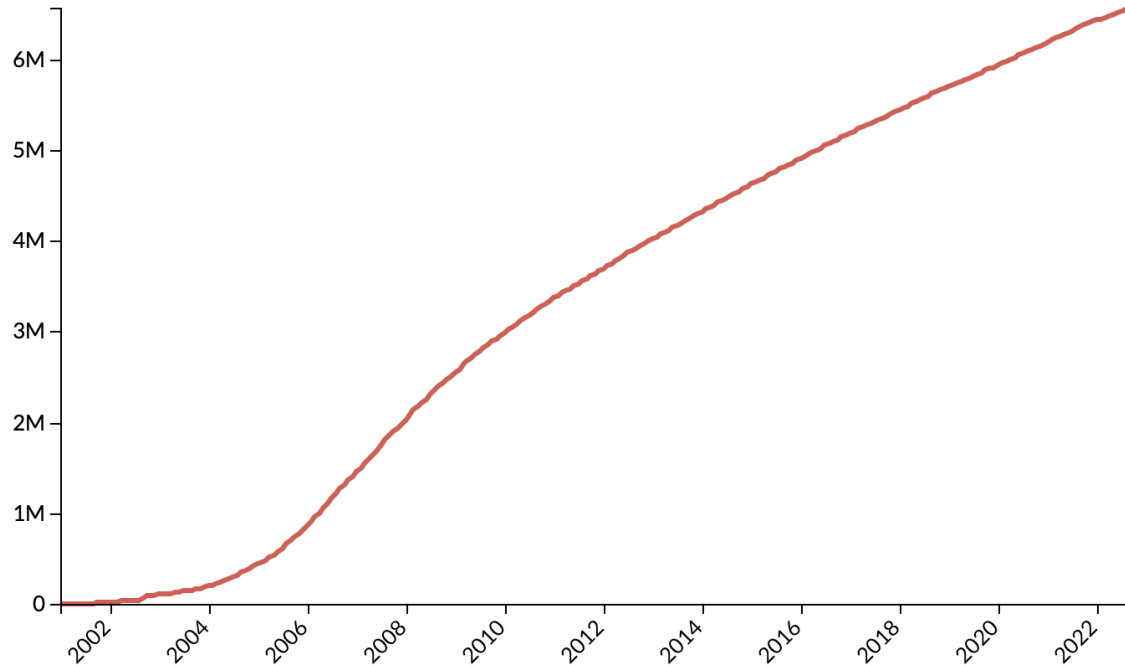
- Data? Compute? Information?

Argument for “scale is **not** all you need”

- Hoffmann et al: To be compute-optimal, model size and training data must be scaled equally.
- It shows that most LLMs are severely starved of data and under-trained.
 - Given the new scaling law, even if you pump a quadrillion parameters into a model, the gains will not compensate for 4x more training tokens
- There is simply not enough [language] data. [QED]

Counter Argument for Not Enough Data (1)

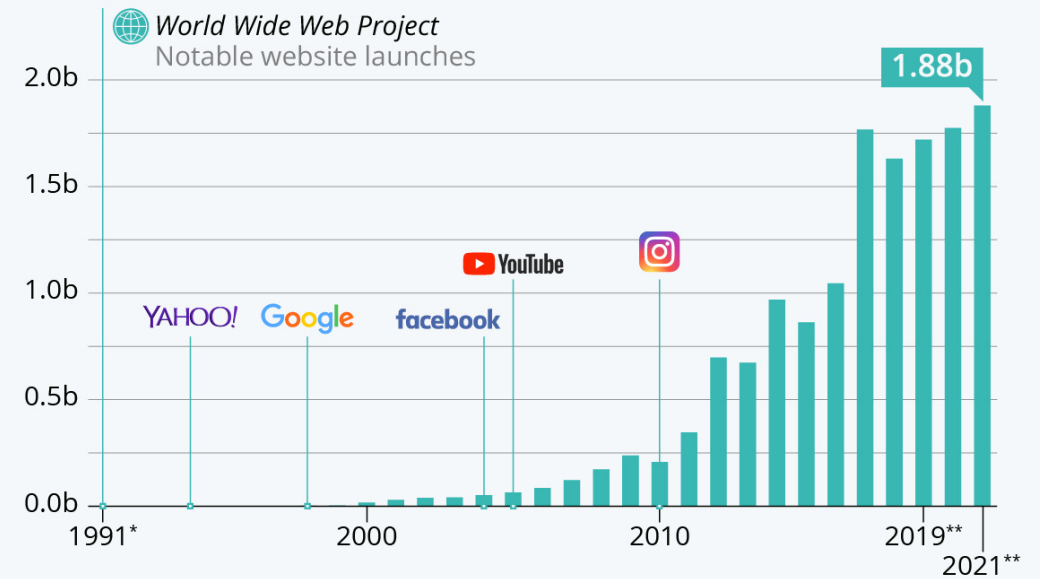
- Data is growing exponentially (?)



Wikipedia size

How Many Websites Are There?

Number of websites online from 1991 to 2021



* As of August 1, 1991.

** Latest available data for 2019: October 28, for 2020: June 2, for 2021: August 6.

Source: Internet Live Stats

Counter Argument for Not Enough Data (2)

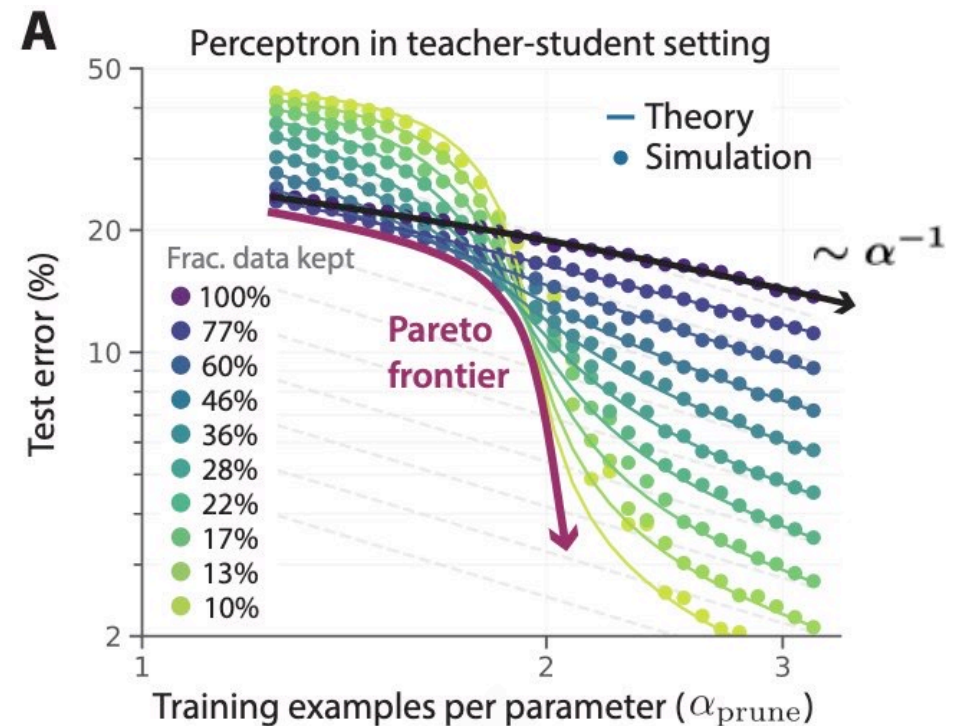
- You can harness data from other modalities.
 - If you want more text, build a solid speech data that converts speech to text.

SKYQUEST

Global Online Video Platforms Market Drives over 80% of Total Internet Traffic |
Skyquest Technology

Counter Argument for Not Enough Data (3)

- You can use data more effectively.
- Sorscher et al. lays out recipes to achieve *exponential* scaling instead through statistical mechanics theory.
- Carefully curating a small subset goes a long way!



Argument for “scale is **not** all you need”

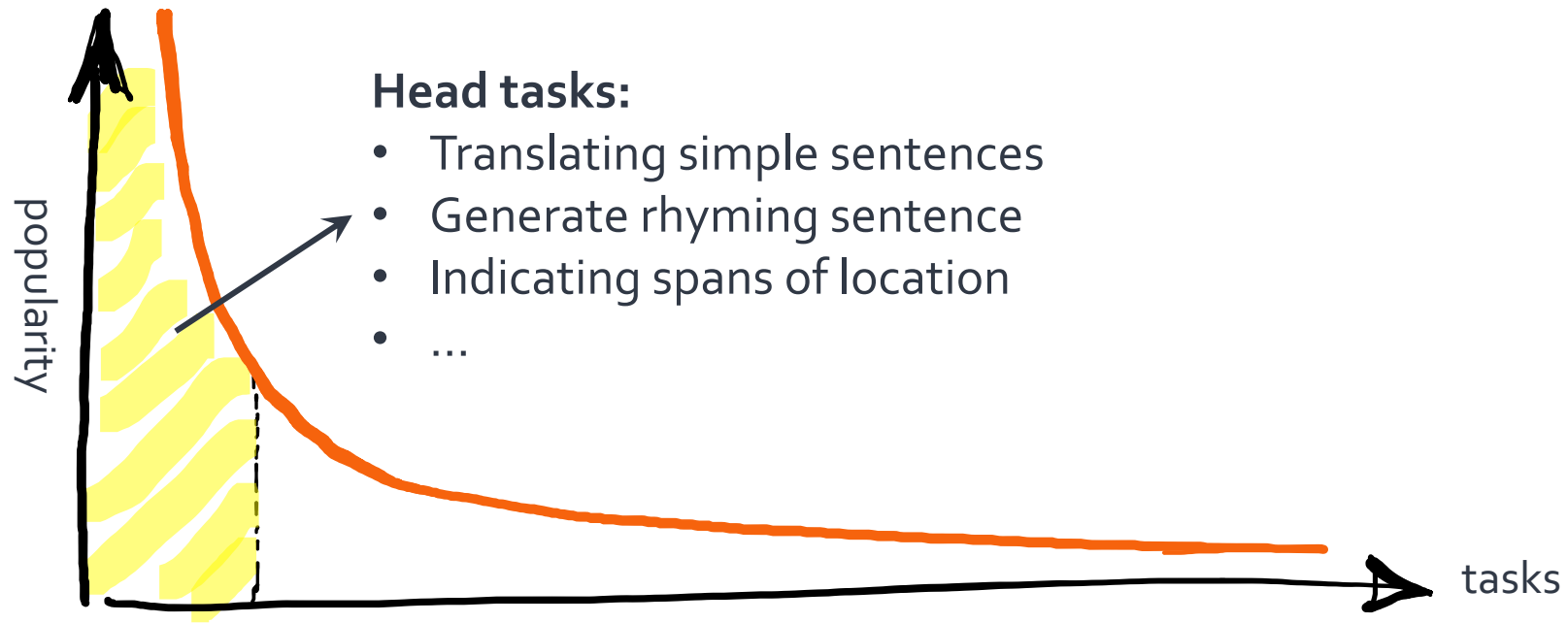
- Argument: There is simply not enough compute available.
- Argument: Scaling is quite costly.
- Counter-argument: large models might be net-negative.

If “Scale” is All You Need”, is “good prediction” all needed for “general intelligence”?

* self-supervised models == predictive models of the world.

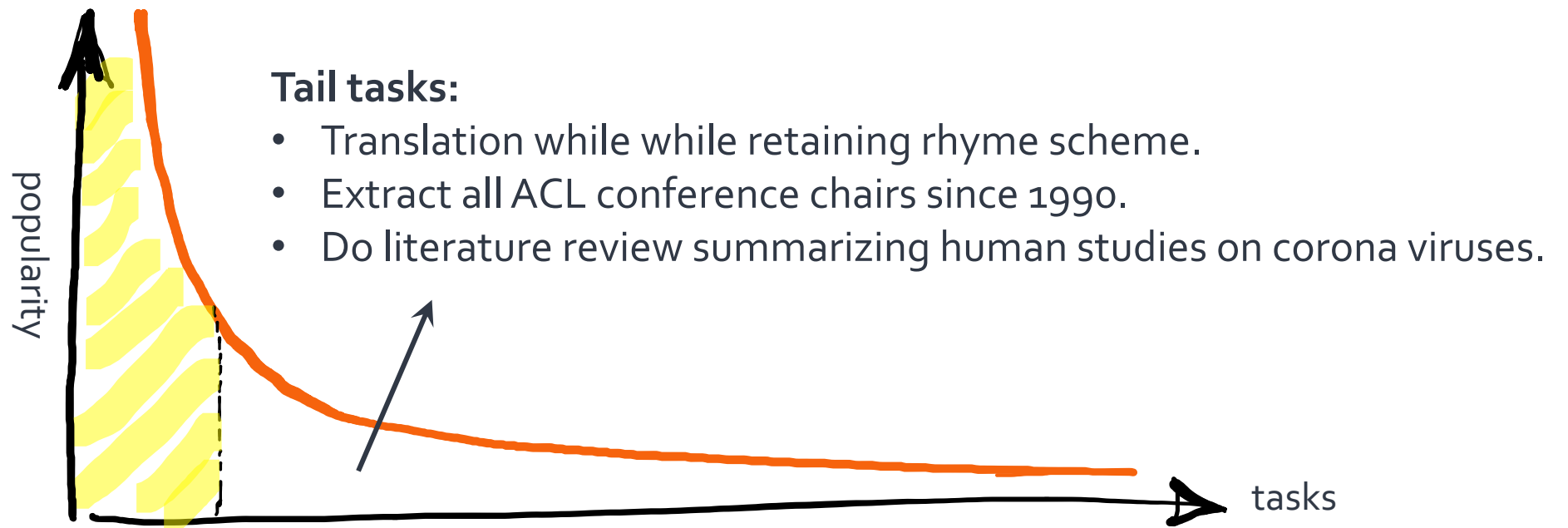
Argument for “scale is **not** all you need”

- Tail phenomena will never go away!



Argument for “scale is **not** all you need”

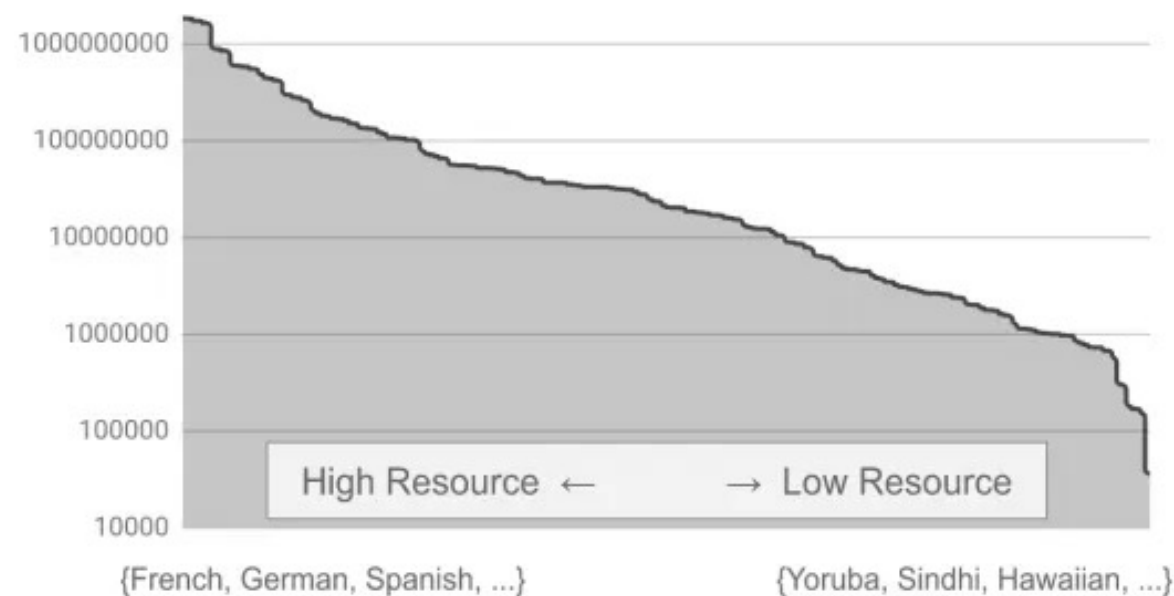
- Tail phenomena will never go away!



Argument for “scale is **not** all you need”

- Tail phenomena will never go away!

Data distribution over language pairs



Example: Google Translate

“The number of parallel sentences [...] ranges from around tens of thousands to almost 2 billion.”

Argument for “scale is **not** all you need”

- Tail phenomena will never go away!
- Will result in brittleness to small changes

“Tesla's Autopilot system confusing horse-drawn carriage for truck”





which number is bigger, 999934 or 32323232332?



The number 32323232332 is larger than 999934. To compare the sizes of two numbers, we need to look at the number of digits in each number. In this case, the number 32323232332 has 10 digits, while the number 999934 has 6 digits. Because the number with more digits is always larger, we can conclude that 32323232332 is the larger number in this case.



which number is bigger, -999934 or -32323232332?



The number -32323232332 is larger than -999934. When comparing negative numbers, we follow the same rules as when comparing positive numbers. In this case, the number -32323232332 has 10 digits, while the number -999934 has 6 digits. Because the number with more digits is always larger, we can conclude that -32323232332 is the larger number in this case.



It is worth noting that negative numbers are always less than positive numbers, so even though -32323232332 is larger than -999934, it is still smaller than 999934. This is because a negative number represents a value that is less than zero, while a positive number represents a value that is greater than zero.

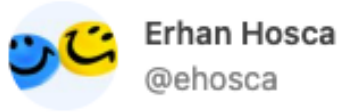
TU

how many more legs does a spider have than an ant?



A spider has eight more legs than an ant. A spider typically has eight legs, while an ant typically has six legs. The number of legs can vary depending on the species of spider or ant, but in general, spiders have eight more legs than ants. The legs of a spider are used for walking, running, climbing, and spinning webs, and they enable the spider to move quickly and efficiently in different environments and terrains. On the other hand, the legs of an ant are used for walking, running, climbing, and digging, and they enable the ant to move quickly and efficiently in different environments and terrains. The additional legs of a spider allow it to have a wider range of movements and behaviors, and to adapt to different situations and challenges.

But does it really "understand"?



Erhan Hosca
@ehosca

@GaryMarcus Isn't this basic probability? Enough monkeys banging on typewriters sort of thing...

3:18 PM · Dec 1, 2022



@emilybender@dair-community.social on Mastodon
@emilybender

That "Limitations" section has it wrong though. ChatGPT generates strings based on combinations of words from its training data. When it sometimes appears to say things that are correct and sensible when a human makes sense of them, that's only by chance.



Mike Ma
@MMikeMMa

@plibin @GaryMarcus So, I'm really struggling here.

These machines don't *understand* anything. As I understand, they are MASSIVE correlation engines.

Yet, they simultaneously:

- hilariously fail
- breathtakingly win

WHY/HOW are both things true? What Qs do the engine win/fail on and why?



3:03 PM · Dec 1, 2022

But does it really “understand”?

- What is “meaning”?
- What is “understanding”?
-

**Climbing towards NLU:
On Meaning, Form, and Understanding in the Age of Data**

Emily M. Bender
University of Washington
Department of Linguistics
ebender@uw.edu

Alexander Koller
Saarland University
Dept. of Language Science and Technology
koller@coli.uni-saarland.de

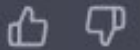
Models are not Grounded in World

k

which is heavier, 10kg iron or 10kg cotton?



10kg iron is heavier than 10kg cotton. Iron is typically heavier than cotton, so the answer is
10kg iron is heavier. Answer: `\boxed{iron}`.



Argument against (1)

- You can make models grounded in world by training them on grounded data.

Argument against (2)

- There are many “intelligent” behaviors in the world that consist of non-intelligent atomic elements.
 - Every neuron in our body
 - Every step of evolution

Back full circle:

Is Scale All We Need?

How does the future look like to you?

Which future will we have?

1. One very large model
2. Few very large models
3. Many very large models