

- Evolution -

• What is the potential for progress in the neuronal approach?
to natural language processing?

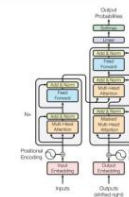
10b.1

1

Evolution: Transformer Families

• Codecs

• as in the initial proposal of the Transformer architecture
• Vaswani et al, 2017, Google



10b.2

2

Transformers Families 2/3 2/3

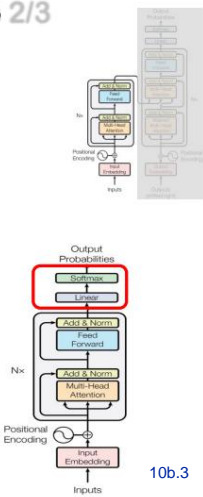
• Encoders

• only with the coding part
• dominate in research and industry in tasks such as text classification, named entity recognition, and question answering.

• in tuning they can receive an extra output layer where classification takes place

• in the literature, called **the classification head**

• NB: nothing to do with attention heads



10b.3

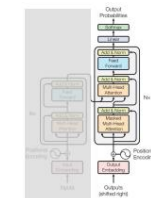
3

Transformers Families 3/3 3/3

• Decoders

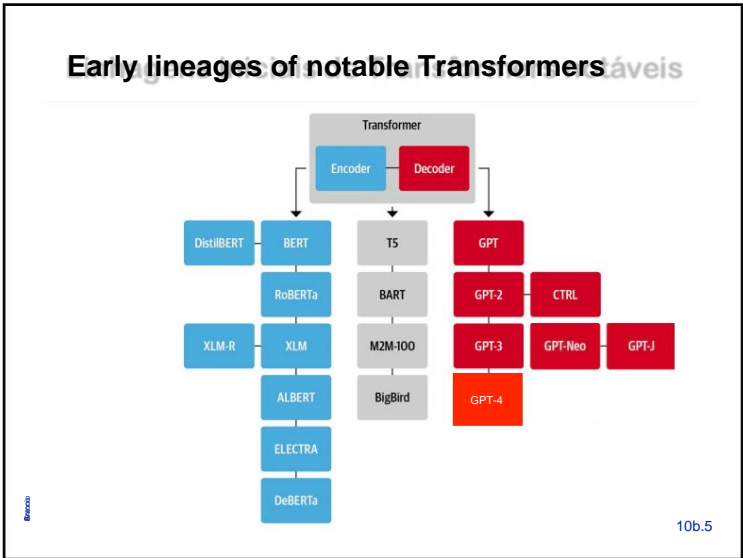
• only with the decoding part

• used mainly for text generation, and show their capabilities with colossal amounts of data



10b.4

4



5

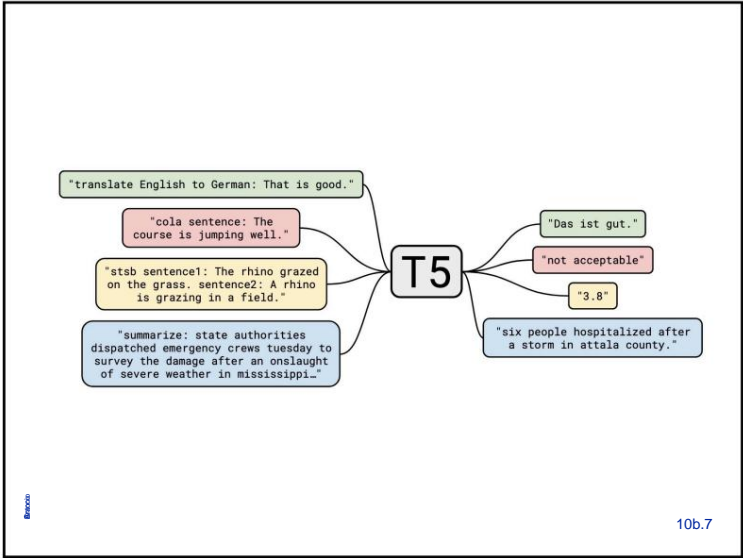
Encoder-decoder lineage 1/2

q T5

- Raffel et al, 2019, Google, 11 x 109 parameters
- All tasks, including classification, are reframed as transduction tasks
- input: segment to be processed together with indication of the task to be performed
- output: text with label designation (classification task), output text (transduction task)

10b.6

6



7

Encoder-decoder lineage 2/2

q M2M-100

- Fan et al, 2020, Facebook, 12 x 109 parameters
- automatic translation between 100 languages
- common model
- Peers with fewer resources can benefit
- input is prefixed with indication of source language to target language
- e.g. to translate from Portuguese to English:
 - pt-en Cork is extracted from the cork oak tree.

10b.8

8

Coding lineage 1/3

q BERT

- Devlin et al, 2018, Google
- pre-trained with 2 objectives: predict tokens masked language modeling and determine whether a sequence/sentence should follow the previous one ("next sentence prediction", NSP)
- the first of this branch, of which the rest represent variants

10b.9

9

Coding lineage 2/3

q BERT

- Each entry is prefixed by the token [CLS]
- Forecast is formed from the representation of this token in the last layer

10b.10

10

Coders branch 3/3

q RoBERTa

- Liu, 2019, Facebook
- trained longer, with larger batches, and without NSP
- improves BERT performance

q DistilBERT

- Sahn, 2020, Hugging Face
- 97% of BERT performance with 40% less memory and 60% faster through knowledge distillation techniques (advanced topic)

10b.11

11

Bouquet of decoders

q GPT-3

- Brown et al, 2020, OpenAI/Msoft
- 175 x 109 parameters, 96 heads, 96 layers
- size per layer 12288
- training with 300 x 109 tokens
- available only via webservice

q GPT-Neo/J

- ongoing attempts to recreate opensource versions of GPT-3, by EleutherAI

q Bloom

- open source, multilingue (PT 5%), crowdsourced

10b.12

12

Prompting 1/3 (“prompting”) 1/3

q GPT-3

• instead of fine-tuning: just prompting:
textual input to be continued

q zero-shot

• provided textual instruction describing the task

q one-shot

• along with textual instruction describing the task, only one
demonstrative example is possible

q few-shot

• along with textual instruction describing the task, more than one demonstrative
example is possible

10b.13

13

Prompting 2/3 (“prompting”) 2/3

Zero-shot

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

```
1 Translate English to French:
2 cheese =>
```

task description

prompt

One-shot

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

```
1 Translate English to French:
2 sea otter => loutre de mer
3 cheese =>
```

task description

example

prompt

Few-shot

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

```
1 Translate English to French:
2 sea otter => loutre de mer
3 peppermint => menthe poivrée
4 plush girafe => girafe peluche
5 cheese =>
```

task description

examples

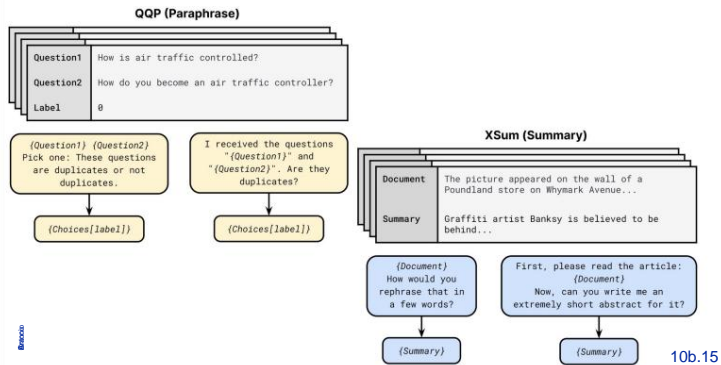
prompt

10b.14

14

Prompting 3/3 (“prompting”) 3/3

phrasing of prompting can improve performance;
exs (Ruder, 2021):



10b.15

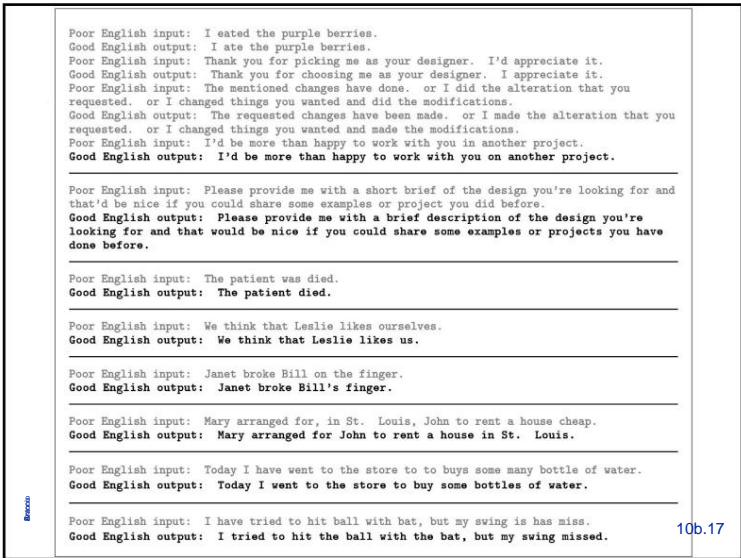
15

few-shot with GPT3 1/2

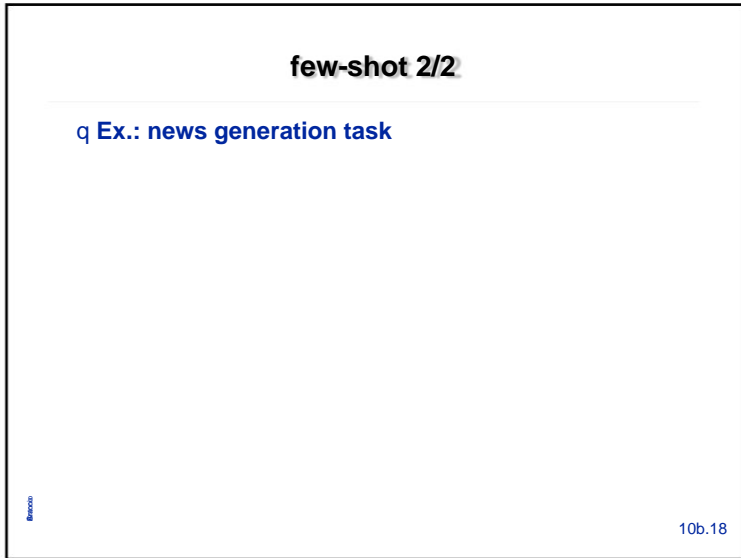
q Ex.: grammar correction task

10b.16

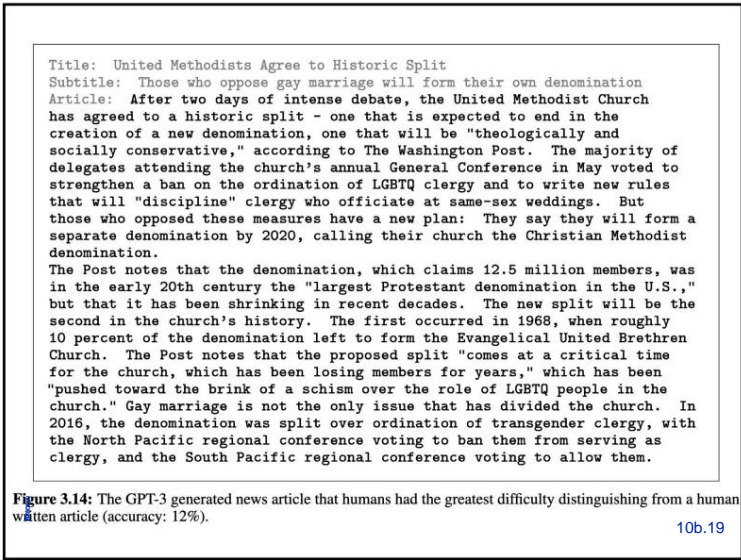
16



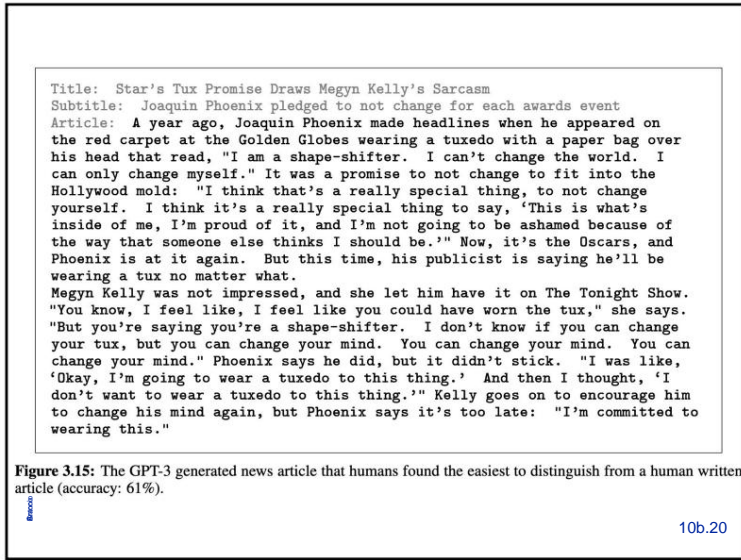
17



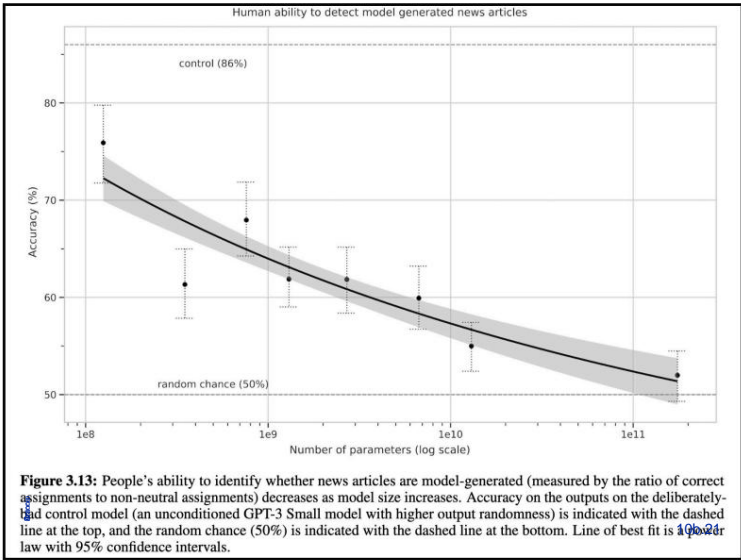
18



19



20



21

one-shot

q Ex.: task of using new words

10b.22

22

A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:
We were traveling in Africa and we saw these very cute whatpus.

To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:
One day when I was playing tag with my little sister, she got really excited and she started doing these crazy farduddles.

A "yalubalu" is a type of vegetable that looks like a big pumpkin. An example of a sentence that uses the word yalubalu is:
I was on a trip to Africa and I tried this yalubalu vegetable that was grown in a garden there. It was delicious.

A "Burringto" is a car with very fast acceleration. An example of a sentence that uses the word Burringto is:
In our garage we have a Burringto that my father drives to work every day.

A "Gigamuru" is a type of Japanese musical instrument. An example of a sentence that uses the word Gigamuru is:
I have a Gigamuru that my uncle gave me as a gift. I love to play it at home.

To "screeg" something is to swing a sword at it. An example of a sentence that uses the word screeg is:
We screeghed at each other for several minutes and then we went outside and ate ice cream.

10b.23

23

zero-shot

q Arithmetic operation task

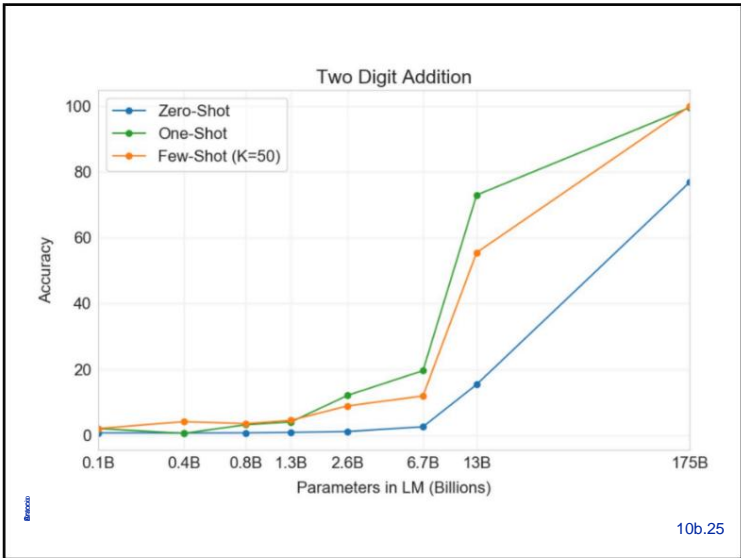
q ... incitement in a few-shot variant

y format of the incitement

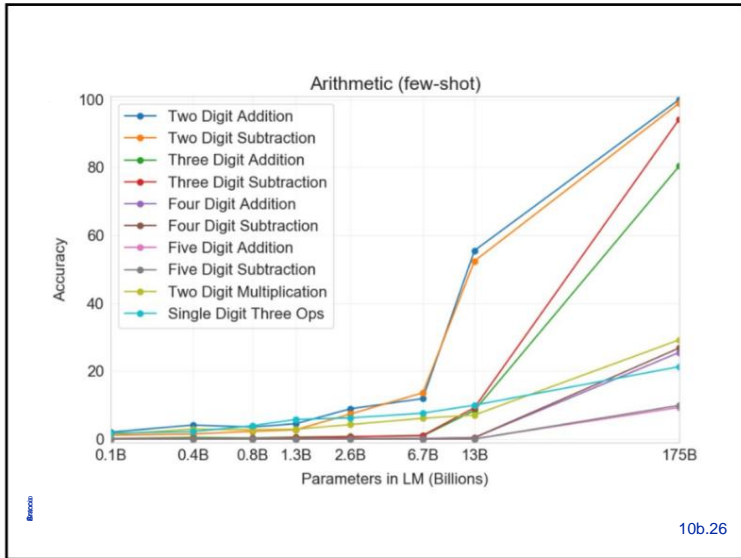
Context →	Q: What is 98 plus 45?
	A:
Target Completion →	143

10b.24

24



25



26

zero-shot with GPT4

Bubeck et al., 2023, Sparks of Artificial Intelligence: Early experiments with GPT-4, MSFT, April 12, 2023, arXiv.

Prompt: Can you write a proof that there are infinitely many primes, with every line that rhymes?

GPT-4:

Yes, I think I can,
Though it might take a clever plan.
I'll start by noting Euclid's proof,
Which shows that primes aren't just aloof.
Assume we have a finite list,
Of primes, and that none have been missed.
Multiply them all together,
And add one, just to be clever.
The result will be a number,
That has no prime factors, oh wonder!
But every number has a prime,
That divides it, it's just a matter of time.
So we've found a contradiction,
And our finite list needs eviction.
There must be infinitely many primes,
And that's the end of my rhyming lines.s

10b.27

27

GPT-4 vs ChatGPT

Prompt: Write a proof of the fact that there are infinitely many primes; do it in the style of a Shakespeare play through a dialogue between two parties arguing over the proof.

GPT-4:

Enter FIRST PARTY and SECOND PARTY.

FIRST PARTY:

Thou dost not understand, my friend,
There simply is no upper end
To primes, those numbers indivisible,
No matter if it seems impossible.

SECOND PARTY:

But how dost thou so surely know?
To prove this truth, what canst thou

ChatGPT:

O Romeo, Romeo! Wherefore art thou, Romeo?

Juliet: Why do you speak in such strange verse, my love?

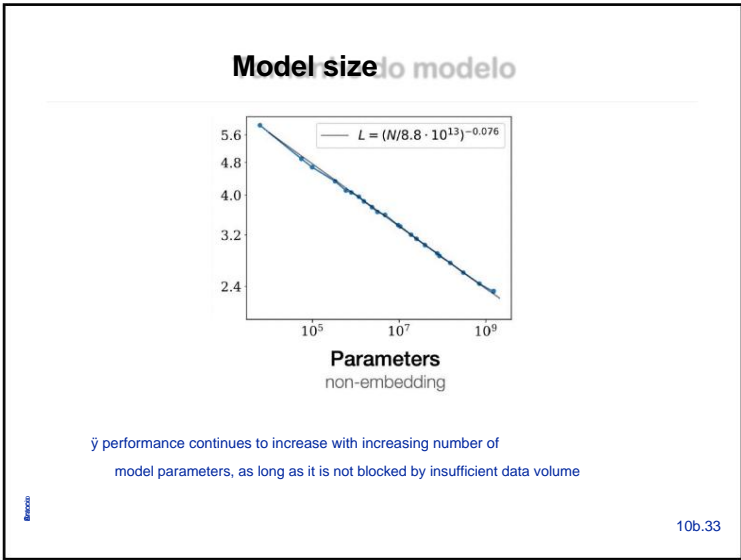
Romeo: Forsooth, my lady, I do it to prove a point. Thou see, I am trying to prove that there are infinitely many primes.

Juliet: Infinitely many primes? How can you prove such a thing?

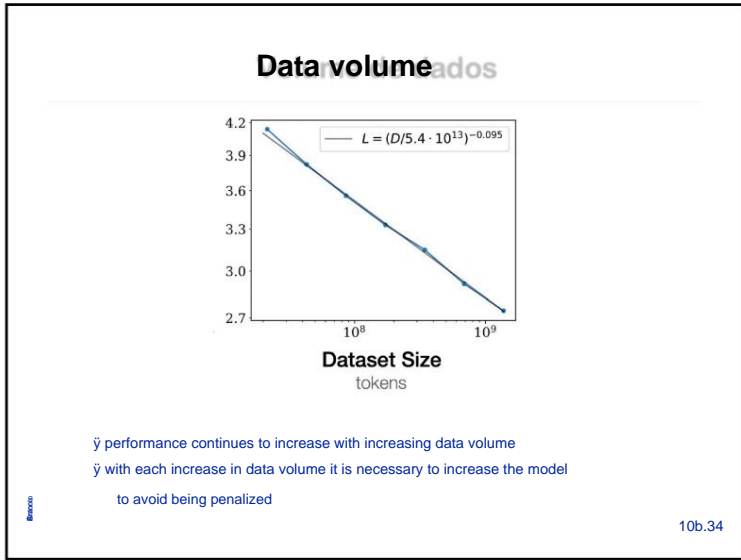
Romeo: Ah, well let me explain. Thou knowest that a prime is a number that is only divisible by one and itself, correct?

10b.28

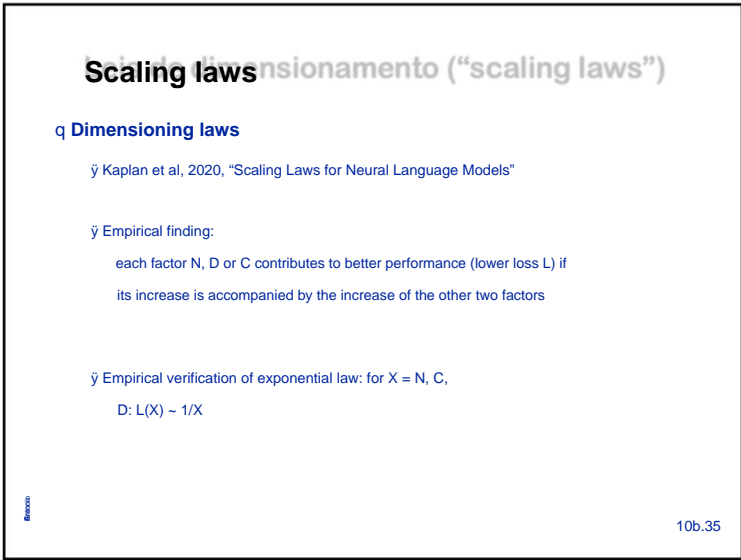
28



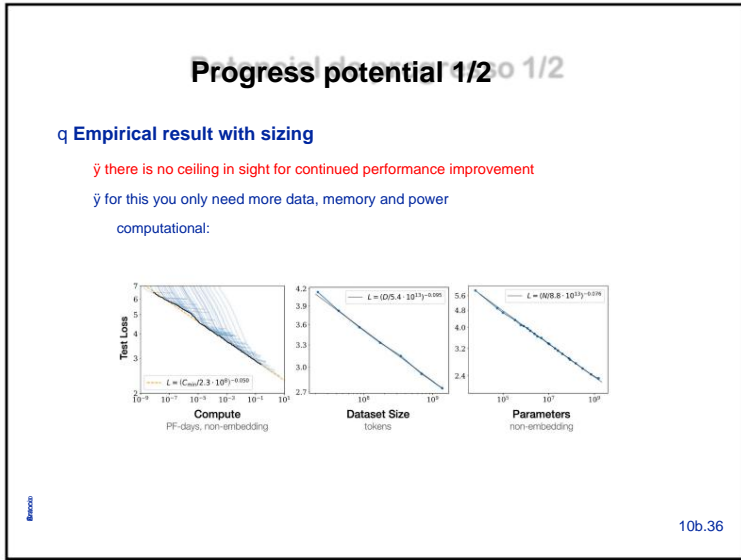
33



34



35

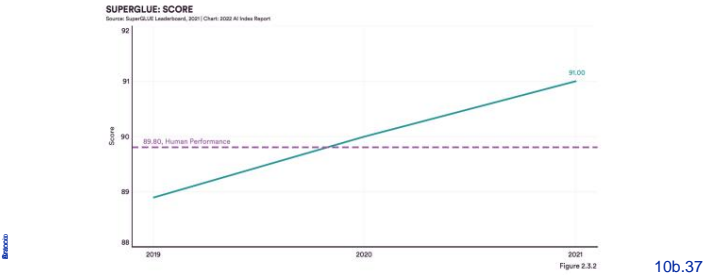


36

Progress potential 2/2

q Empirical result with downstream tasks (SuperGLUE)

- no slowdown in progress is in sight
- Zhang et al, 2022, The AI Index 2022 Annual Report:



37

- Conclusion -

q Index

- Transformer families: encoders-decoders, encoders, decoders
- Notable lineages of Transformers: T5, BERT, GPT, etc.
- Incitement: zero-, one- and few-shot
- Exponential Scaling Law

We have seen that there is great potential for progress in natural language processing based on Transformers

What are the most prominent challenges facing the scientific and technological progress in this area?

38