

Transformers Families 2/3 2/3

q Encoders

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

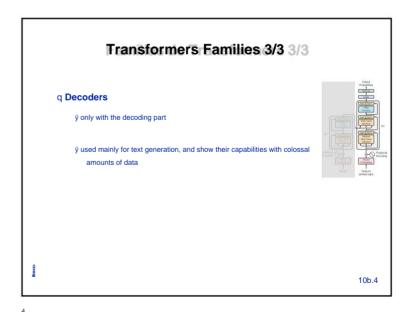
ÿ in tuning they can receive an extra output layer where classification takes place

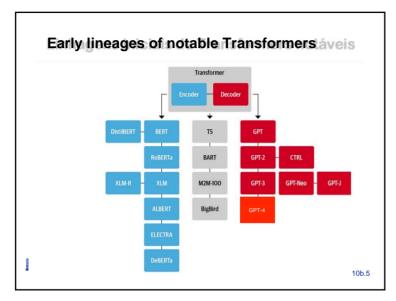
ÿ in the literature, called the classification head

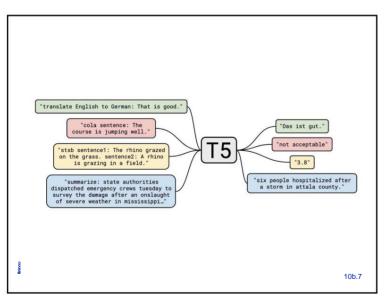
ÿ NB: nothing to do with attention heads

q Codecs
ÿ as in the initial proposal of the Transformer architecture
ÿ Vaswani et al, 2017, Google

2







g T5

ÿ Raffel et al, 2019, Google, 11 x 109 parameters

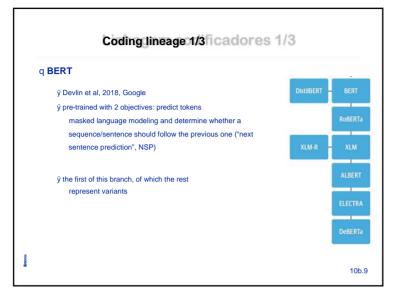
ÿ All tasks, including classification, are reframed as transduction tasks

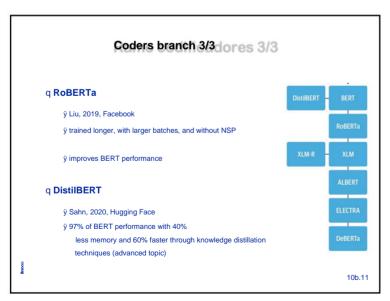
ÿ input: segment to be processed together with indication of the task to be performed

ÿ output: text with label designation (classification task), output text (transduction task)

6

# Encoder-decoder lineage 2/2 codifications 2/2 q M2M-100 ÿ Fan et al, 2020, Facebook, 12 x 109 parameters ÿ automatic translation between 100 languages ÿ common model ÿ Peers with fewer resources can benefit ÿ input is prefixed with indication of source language to target language ÿ e.g. to translate from Portuguese to English: • pt-en Cork is extracted from the cork oak tree.





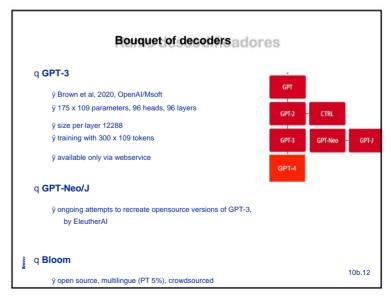
q BERT

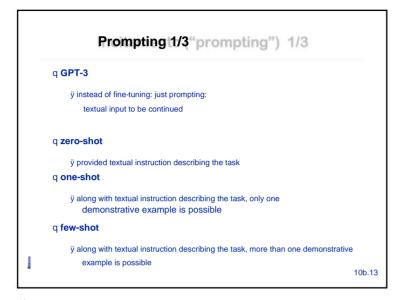
ÿ Each entry is prefixed by the token [CLS]
ÿ Forecast is formed from the representation of this token in the last layer

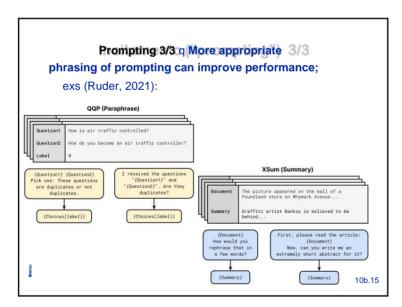
Output

Probabilities

Pro







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. Translate English to French: task descript. In addition to the task description, the model sees a single example of the task. No gradient updates are performed. Translate English to French: sea otter => loutre de mer example 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 peppermint => menthe poivrée plush girafe => girafe peluche 10b.14

14

16

# few-shot with GPT3 1/2 q Ex.: grammar correction task

15

Poor English input: I eated the purple berries. Good English output: I ate the purple berries.

Poor English input: Thank you for picking me as your designer. I'd appreciate it. Good English output: Thank you for choosing me as your designer. I appreciate it.
Poor English input: The mentioned changes have done. or I did the alteration that you requested. or I changed things you wanted and did the modifications. Good English output: The requested changes have been made, or I made the alteration that you requested. or I changed things you wanted and made the modifications. Poor English input: I'd be more than happy to work with you in another project.

Good English output: I'd be more than happy to work with you on another project. Poor English input: Please provide me with a short brief of the design you're looking for and that'd be nice if you could share some examples or project you did before Good English output: Please provide me with a brief description of the design you're looking for and that would be nice if you could share some examples or projects you have Poor English input: The patient was died. Good English output: The patient died. Poor English input: We think that Leslie likes ourselves. Good English output: We think that Leslie likes us. Poor English input: Janet broke Bill on the finger. Good English output: Janet broke Bill's finger. Poor English input: Mary arranged for, in St. Louis, John to rent a house cheap. Good English output: Mary arranged for John to rent a house in St. Louis. Poor English input: Today I have went to the store to to buys some many bottle of water. Good English output: Today I went to the store to buy some bottles of water. Poor English input: I have tried to hit ball with bat, but my swing is has miss. 10b.17 Good English output: I tried to hit the ball with the bat, but my swing missed.

17

19

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: After two days of intense debate, the United Methodist Church
has agreed to a historic split - one that is expected to end in the
creation of a new denomination, one that will be "theologically and
socially conservative," according to The Washington Post. The majority of
delegates attending the church's annual General Conference in May voted to
strengthen a ban on the ordination of LGBTQ clergy and to write new rules
that will "discipline" clergy who officiate at same-sex weddings. But
those who opposed these measures have a new plan: They say they will form a
separate denomination by 2020, calling their church the Christian Methodist

The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

Figure 3.14: The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a human watten article (accuracy: 12%).

10b.19

.....

### few-shot 2/2

q Ex.: news generation task

Mode

18

Title: Star's Tux Promise Draws Megyn Kelly's Sarcasm Subtitle: Joaquin Phoenix pledged to not change for each awards event Article: A year ago, Joaquin Phoenix made headlines when he appeared on the red carpet at the Golden Globes wearing a tuxedo with a paper bag over his head that read, "I am a shape-shifter. I can't change the world. I can only change myself." It was a promise to not change to fit into the Hollywood mold: "I think that's a really special thing, to not change yourself. I think it's a really special thing to say, 'This is what's inside of me, I'm proud of it, and I'm not going to be ashamed because of the way that someone else thinks I should be.'" Now, it's the Oscars, and Phoenix is at it again. But this time, his publicist is saying he'll be wearing a tux no matter what.

Megyn Kelly was not impressed, and she let him have it on The Tonight Show. "You know, I feel like, I feel like you could have worn the tux," she says. "But you're saying you're a shape-shifter. I don't know if you can change your tux, but you can change your mind. You can change your mind." Phoenix says he did, but it didn't stick. "I was like, '(Nay, I'm going to wear a tuxedo to this thing.' And then I thought, 'I don't want to wear a tuxedo to this thing.' Kelly goes on to encourage him to change his mind again, but Phoenix says it's too late: "I'm committed to wearing this."

Figure 3.15: The GPT-3 generated news article that humans found the easiest to distinguish from a human written article (accuracy: 61%).

10b.20

10b.18

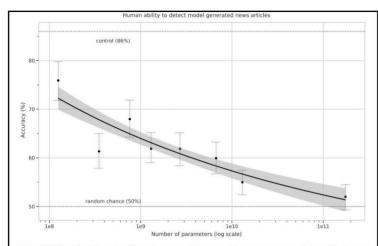


Figure 3.13: People's ability to identify whether news articles are model-generated (measured by the ratio of correct assignments to non-neutral assignments) decreases as model size increases. Accuracy on the outputs on the deliberately bad control model (an unconditioned GPT-3 Small model with higher output randomness) is indicated with the dashed line at the top, and the random chance (50%) is indicated with the dashed line at the bottom. Line of best fit is â@w@ar law with 95% confidence intervals.

23

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 valubalu 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 "Burringo" is a car with very fast acceleration. An example of a sentence that uses the word Burringo is: In our garage we have a Burringo 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 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 We screeghed at each other for several minutes and then we went outside and ate ice cream.

q Ex.: task of using new words

22

24

# zero-shot

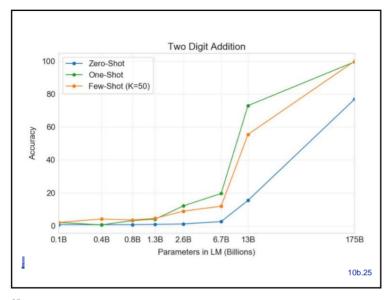
q Arithmetic operation task

q ... incitement in a few-shot variant

ÿ format of the incitement

 $\begin{array}{ccc} {\sf Context} \to & {\sf Q: \ What \ is \ 98 \ plus \ 45?} \\ & & {\sf A:} \end{array}$  Target Completion  $\to & 143$ 

10b.24



## zero-shot with GPT4

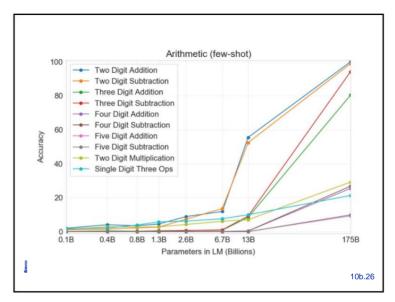
Bubeck et al., 2023, Sparks of Artificial Intelligence: Early experiments with GPT-4, MSFT, April 12, 2023, arXiv.

**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



26

### 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:

28

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?

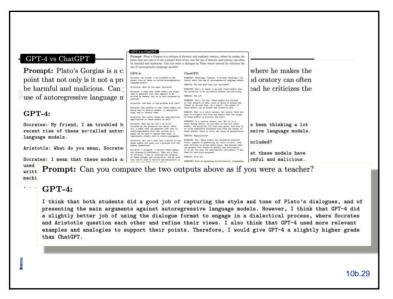
Romeo: Forsooth, my lady, I do it to prove a point. Thou see, I am trying to prove that there are infinitely many primes.

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

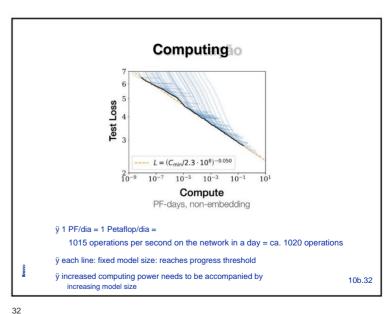
27

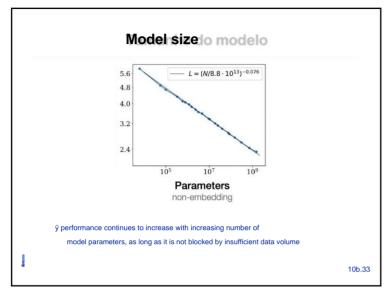


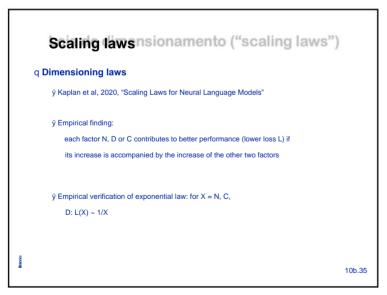
# PEvolution: Performance q Dimensioning laws ÿ Kaplan et al, 2020, "Scaling Laws for Neural Language Models" ÿ Family: decoder ÿ Task: language modeling ÿ Loss metric L: cross-entropy, weighted every 1024 tokens ÿ Vocabulary: 50257, tokenized via byte-pair encoding ÿ Performance/loss vs. Model size (N), excluding embeddings ÿ Performance/loss vs. Data volume (D) ÿ Performance/loss vs. Computation (C)

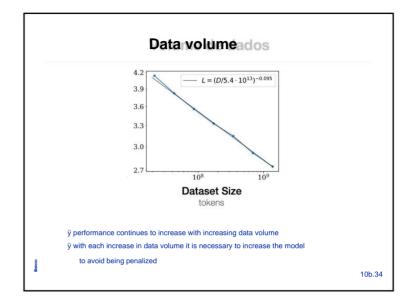
**▷** Evolution: Size of models modelos q Tunstall et al., 2022, p.346 Switch-C GShard GPT-3 E 10 Turing-NLG 1010 Megatron GPT-2 BERT GPT 2018-01 2018-07 2019-01 2020-01 2020-07 2019-07 q zero- and few-shot capabilities from the 10-100x109 parameter range q performance continues to increase with increasing model size.. 10b.30

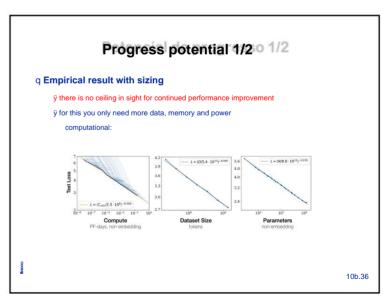
30

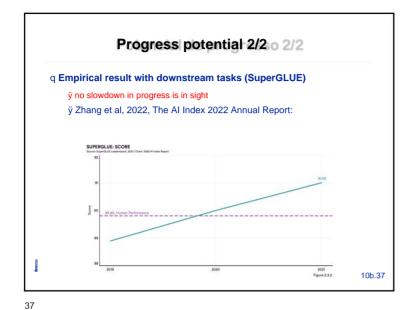












- Conclusion 
q Index

ÿ Transformer families: encoders-decoders,
encoders, decoders

ÿ Notable lineages of Transformers: T5, BERT, GPT, etc.

ÿ Incitement: zero-, one- and few-shot

ÿ 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?