

# Behind the journey of ChatGPT:

## An overview of Large Language Models and their capacities

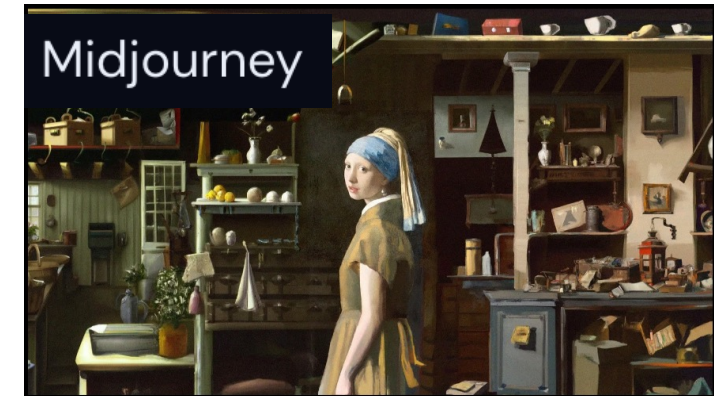
Laure Soulier

AI2S2 2023

LA Ecris un texte de 5 phrases sur Jian-Yun Nie



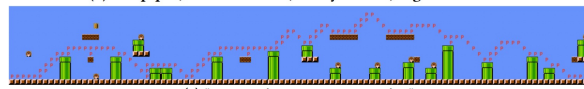
Jian-Yun Nie est un chercheur et professeur émérite en informatique. Il est connu pour ses contributions significatives dans le domaine du traitement de l'information et de la recherche d'information. Ses travaux portent notamment sur l'exploration de textes, la recherche d'informations multilingues et la fouille de données. En tant que chercheur de renommée mondiale, Jian-Yun Nie a publié de nombreux articles scientifiques et a participé à des conférences internationales. Ses recherches ont eu un impact important dans le domaine de l'informatique et continuent d'influencer de nouvelles avancées technologiques.



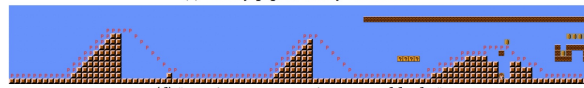
(a) "many pipes, many enemies, little blocks, low elevation"



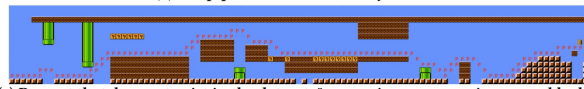
(b) "no pipes, some enemies, many blocks, high elevation"



(c) "many pipes, many enemies"



(d) "no pipes, no enemies, many blocks"



(e) Prompt that does not exist in the dataset: "many pipes, no enemies, many blocks"



(f) Failure case: "Many pipes, no enemies, some blocks"

Pour en avoir un aperçu, voici l'extrait généré pour le prompt « *bande sonore épique utilisant des instruments orchestraux. Le morceau fait monter la tension, crée un sentiment d'urgence. Un chœur a cappella chante à l'unisson, ce qui crée un sentiment de puissance et de force* » :

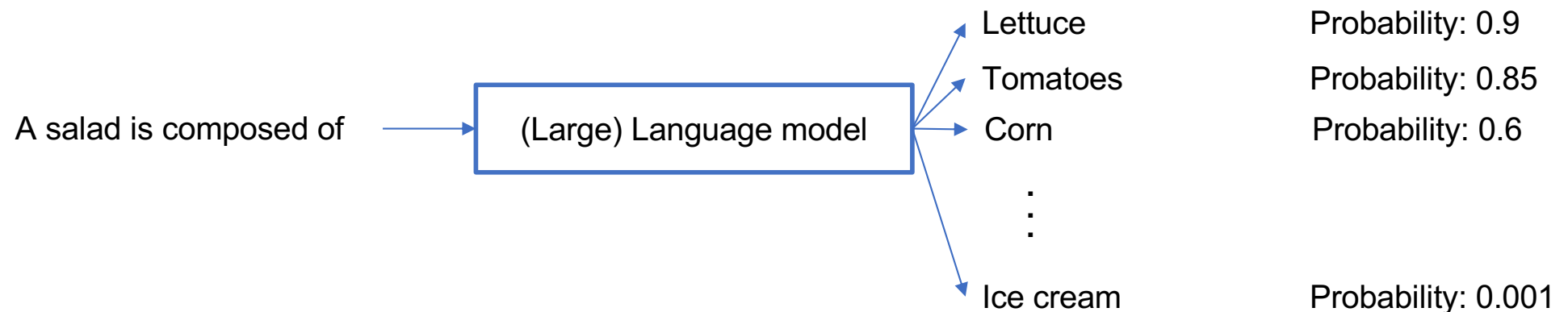
© openAI, Midjourney,  
<https://www.blogdumoderateur.com/musiclm-nouvel-outil-ia-google-generer-musique/>  
And (Sudhakaran et al 2023)



# Understanding (Large) Language Models

## Principle:

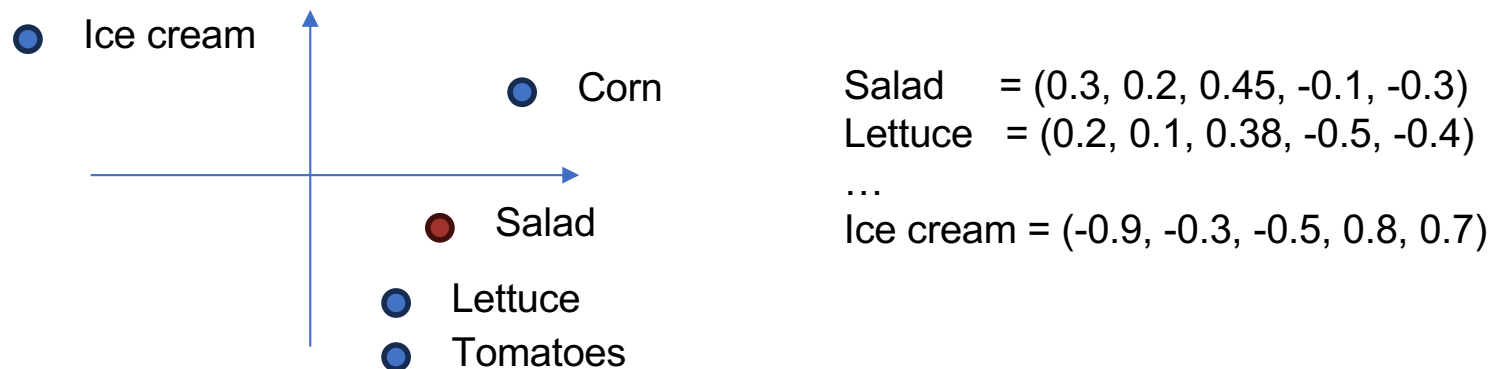
- Modeling the probability of sequences  $x_1, x_2, \dots, x_n$
- Items may be words, characters, character ngrams, word pieces, etc



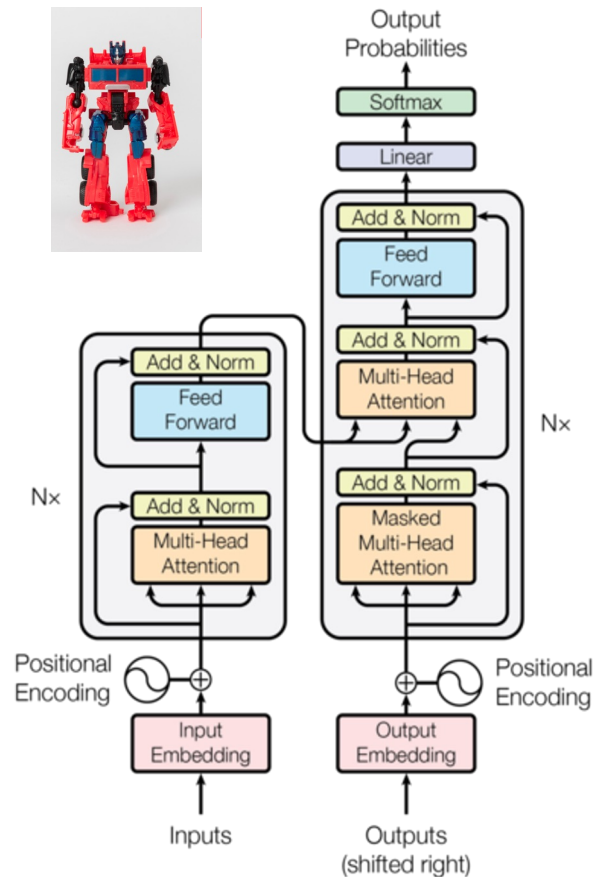
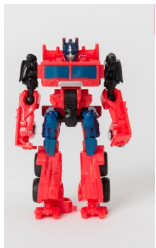
Given a sequence of items  $x_1, x_2, \dots, x_{n-1}$ , what is the probability of the next item  $x_n$ ?

$$P(x_n | x_1, x_2, \dots, x_{n-1})$$

## Semantics, word representation and latent space



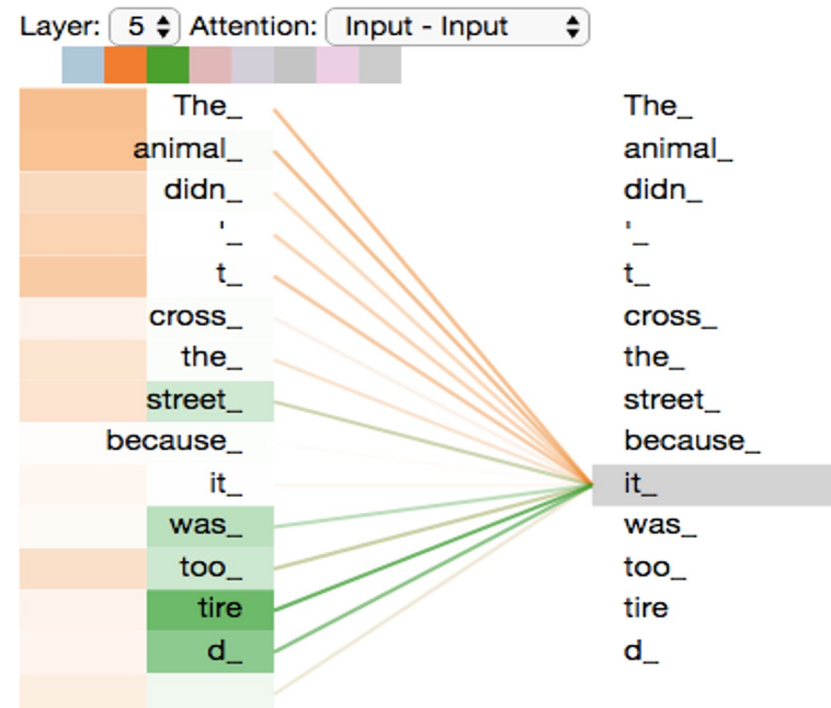
## Transformer networks (2017)



A encoder-decoder neural network w/:

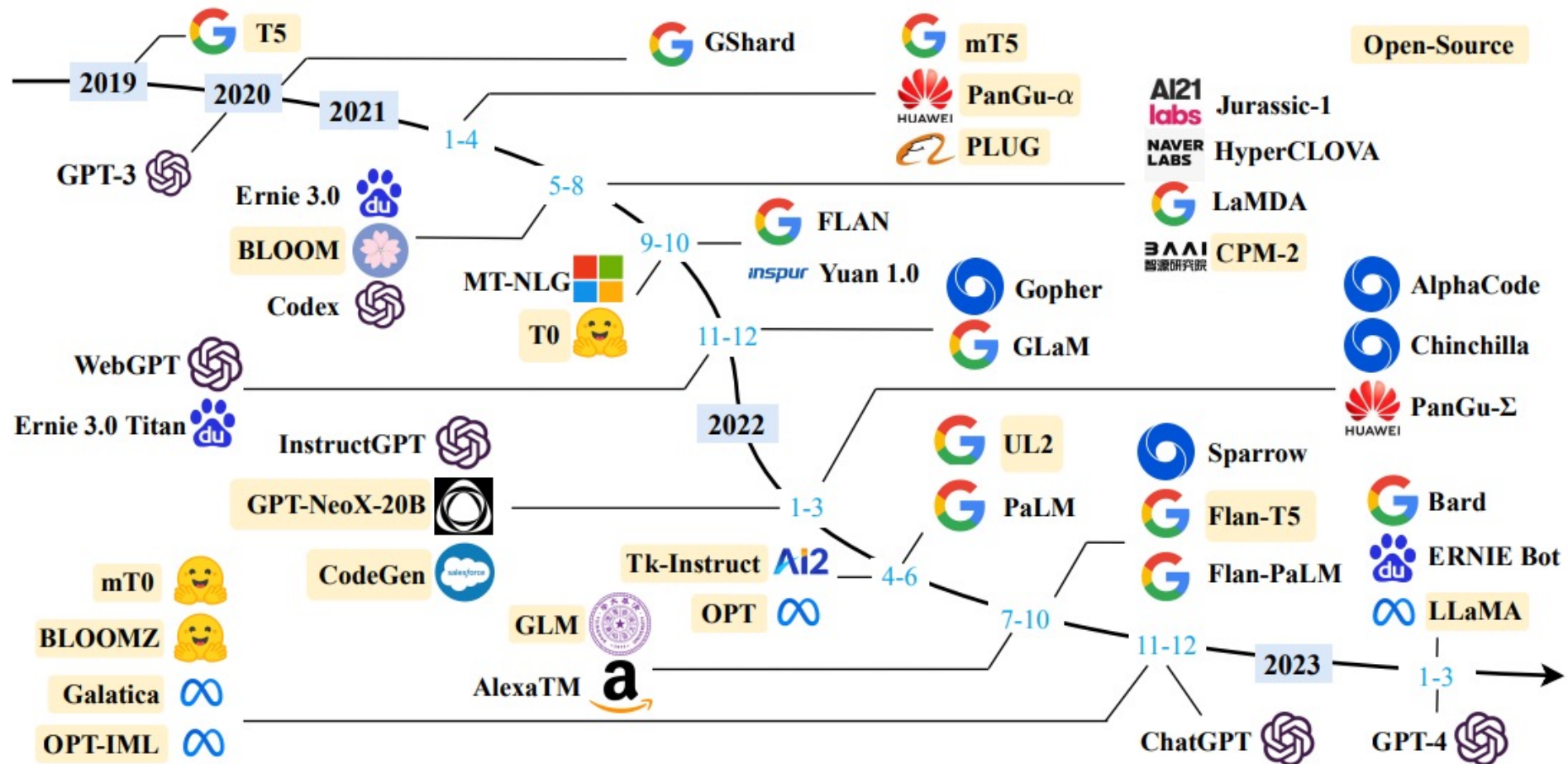
- About 65M parameters
- Successive feed-forward blocks
- Parallel heads

... That estimates contextual representations of items with self-attention



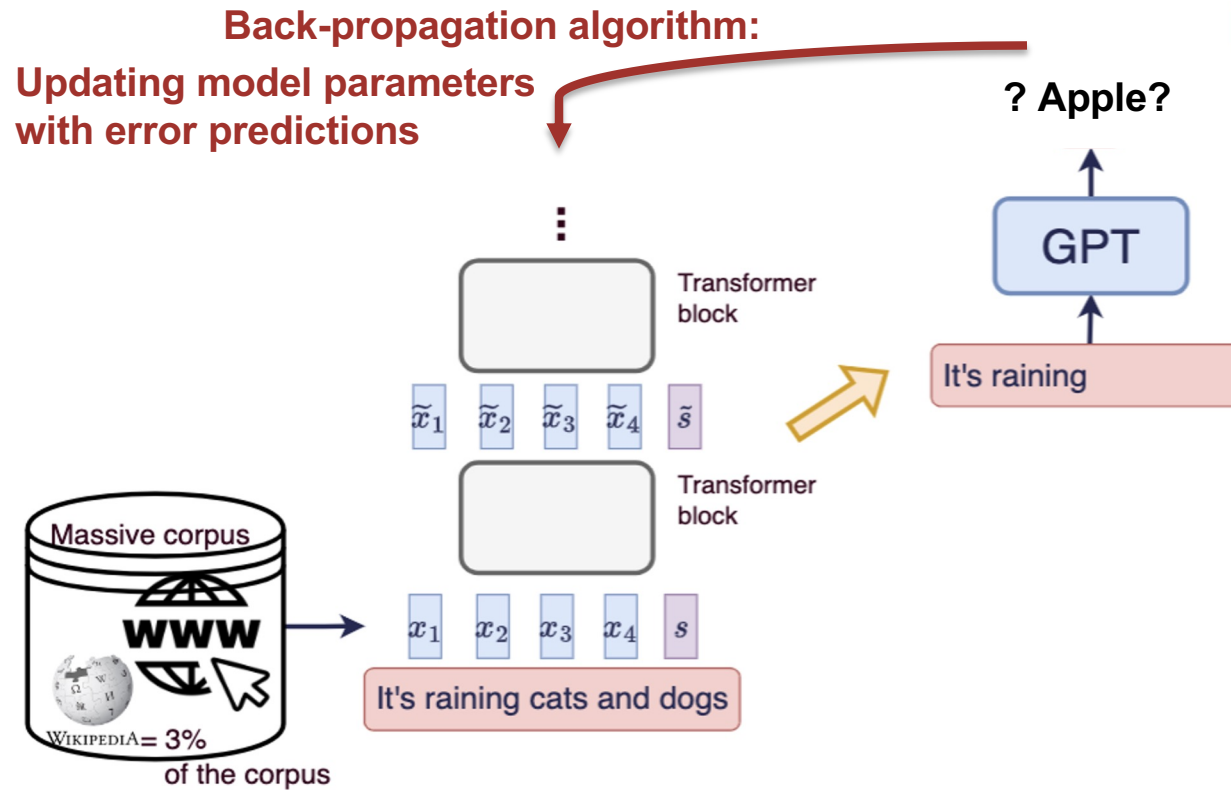


# A part of the collection of large language models

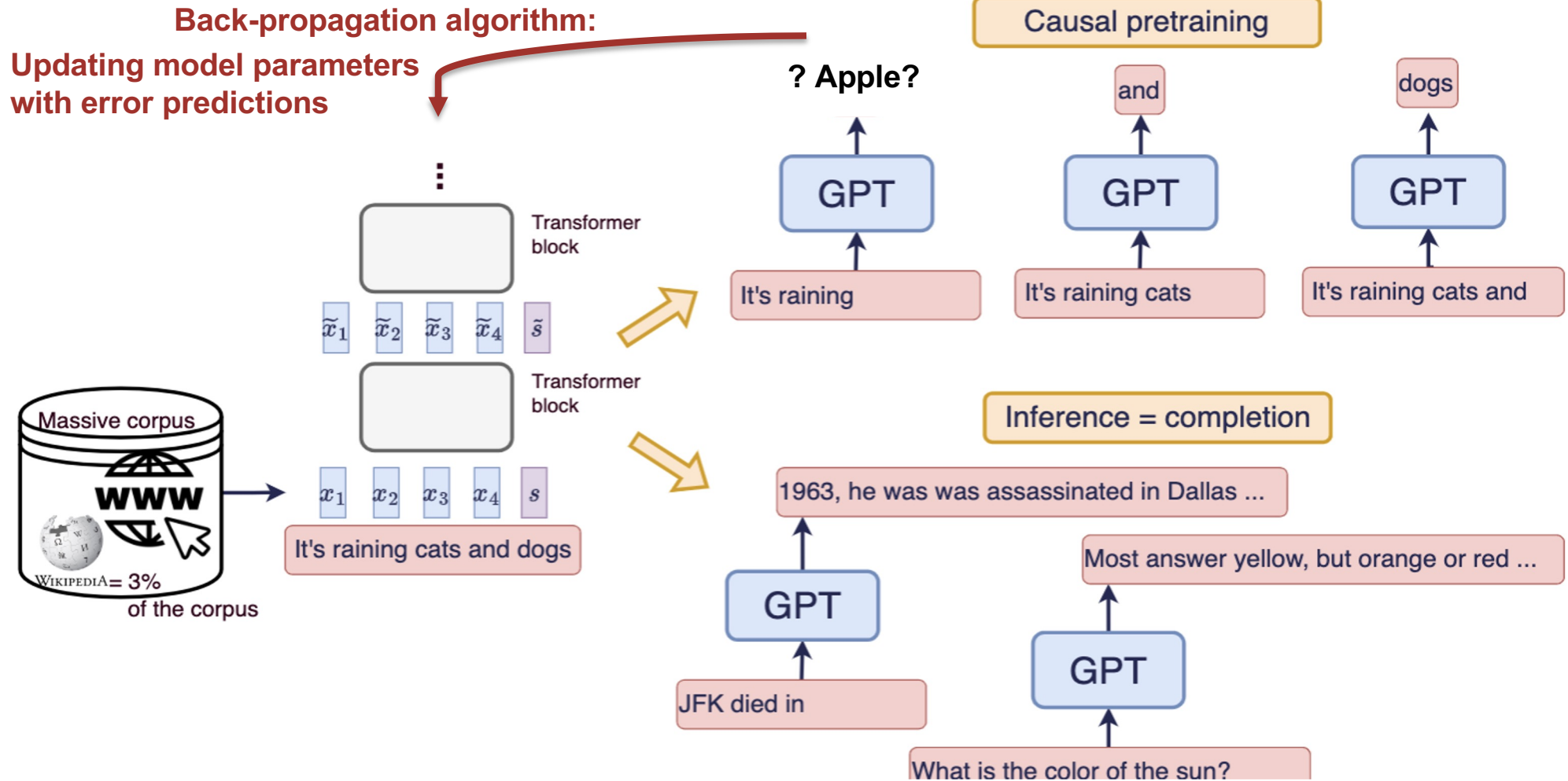


## Training transformer networks (e.g. GPT)

Causal pretraining



## Training transformer networks (e.g. GPT)



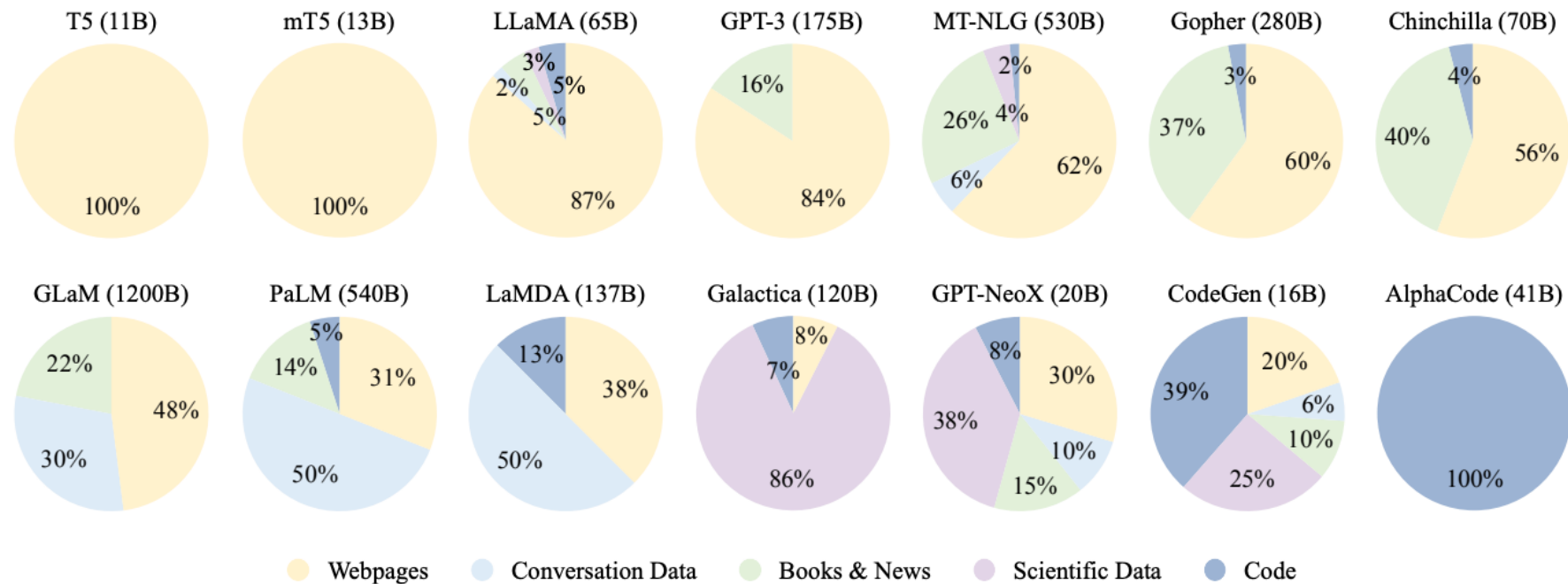


# The importance of data

## → Quality

- Important part of human annotators in ChatGPT

## → Type and variety



## → Grounding

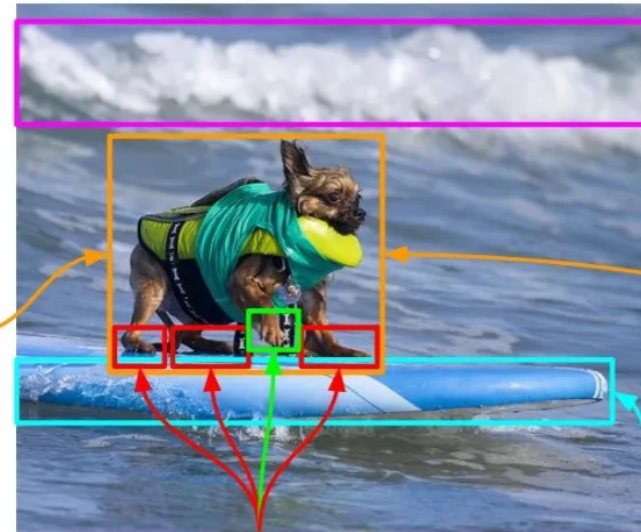
- Mixing text and images to improve semantics

Where does this scene take place?

- A) In the sea. ✓
- B) In the desert.
- C) In the forest.
- D) On a lawn.

What is the dog doing?

- A) Surfing. ✓
- B) Sleeping.
- C) Running.
- D) Eating.



Why is there foam?

- A) Because of a wave. ✓
- B) Because of a boat.
- C) Because of a fire.
- D) Because of a leak.

What is the dog standing on?

- A) On a surfboard. ✓
- B) On a table.
- C) On a garage.
- D) On a ball.

Which paw is lifted?

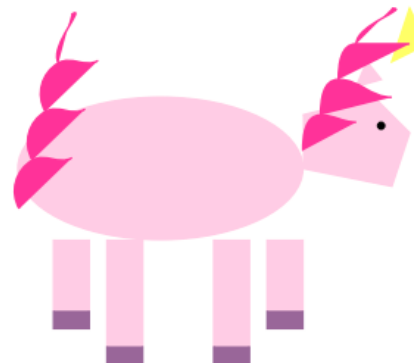
(Zhu et al, 2016)

**Prompt:** Draw a unicorn in TikZ.

**ChatGPT:** [Produces L<sup>A</sup>T<sub>E</sub>X compiling to following picture.]

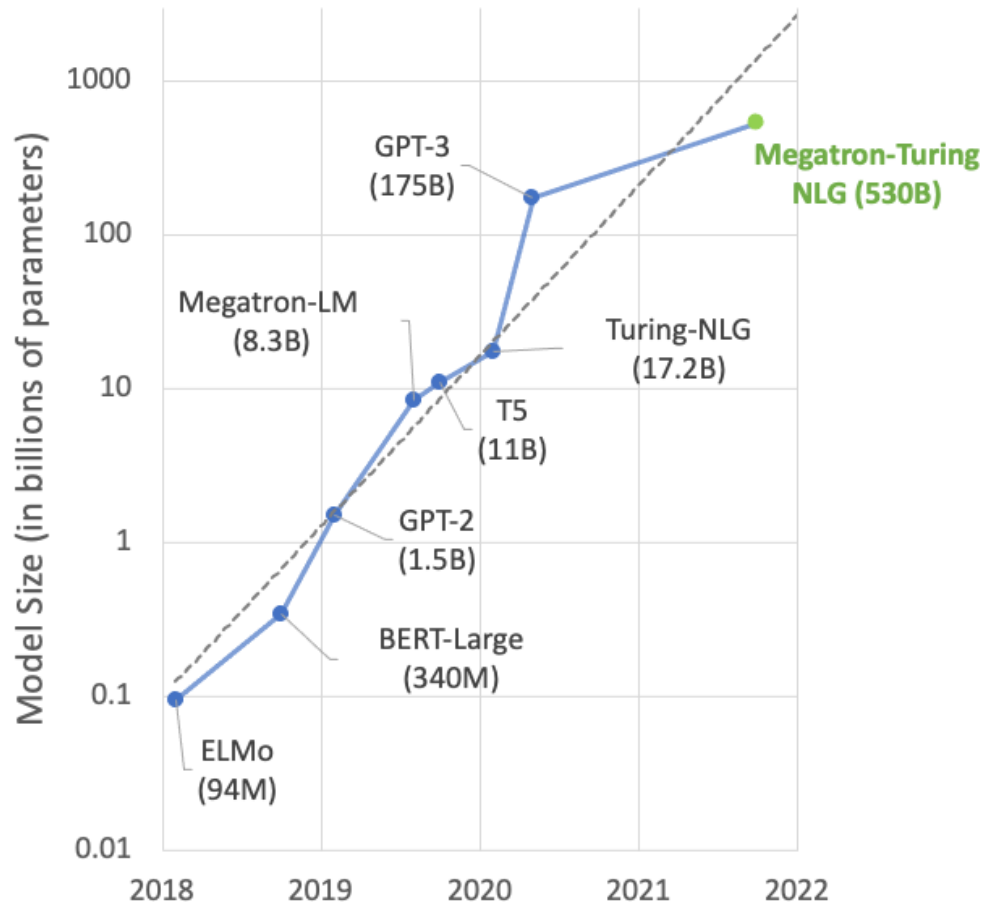


**GPT-4:**

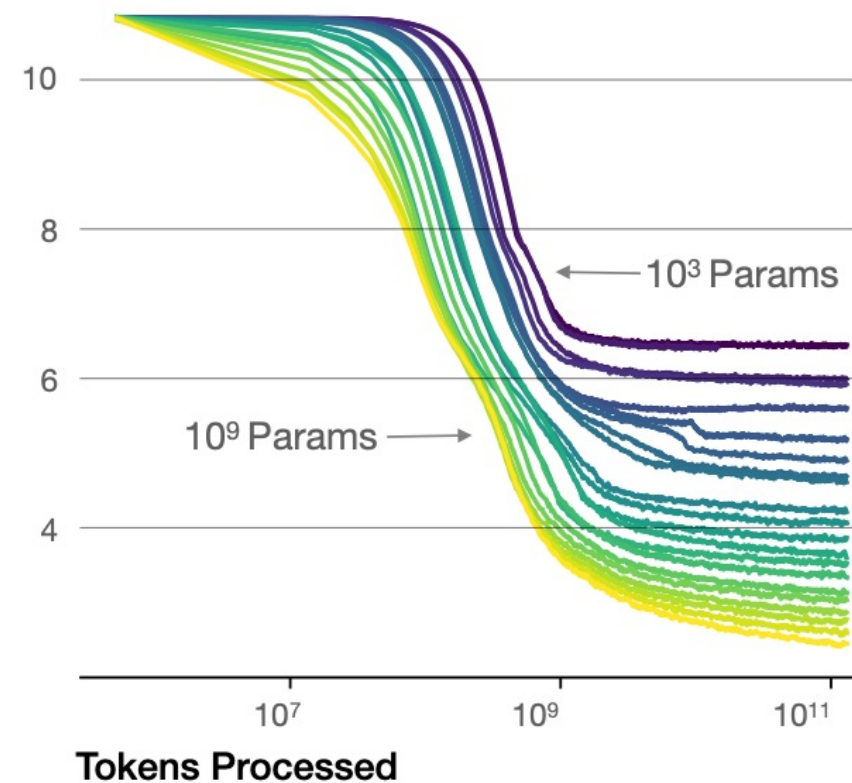


# Large Language Models: interesting properties

## → Scaling law



Larger models require **fewer samples** to reach the same performance



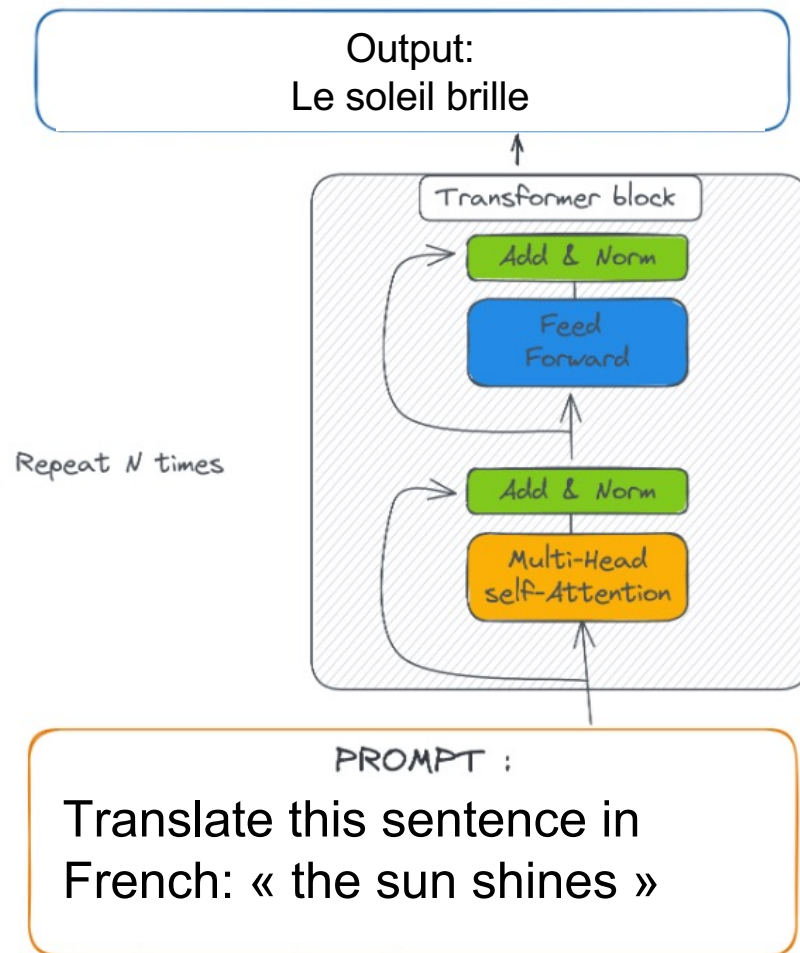
© <https://aibusiness.com/companies/nvidia-and-microsoft-build-the-world-s-largest-530bn-parameter-language-model>

(Kaplan et al, 2020)

Larger Language Models reach good performance level earlier than small language models:

- fewer optimization steps
- fewer data points

## → Prompting



From Thomas Gerald - 2023

## → Prompt :

Instruction explicitly expressing what is expected

## → Challenge:

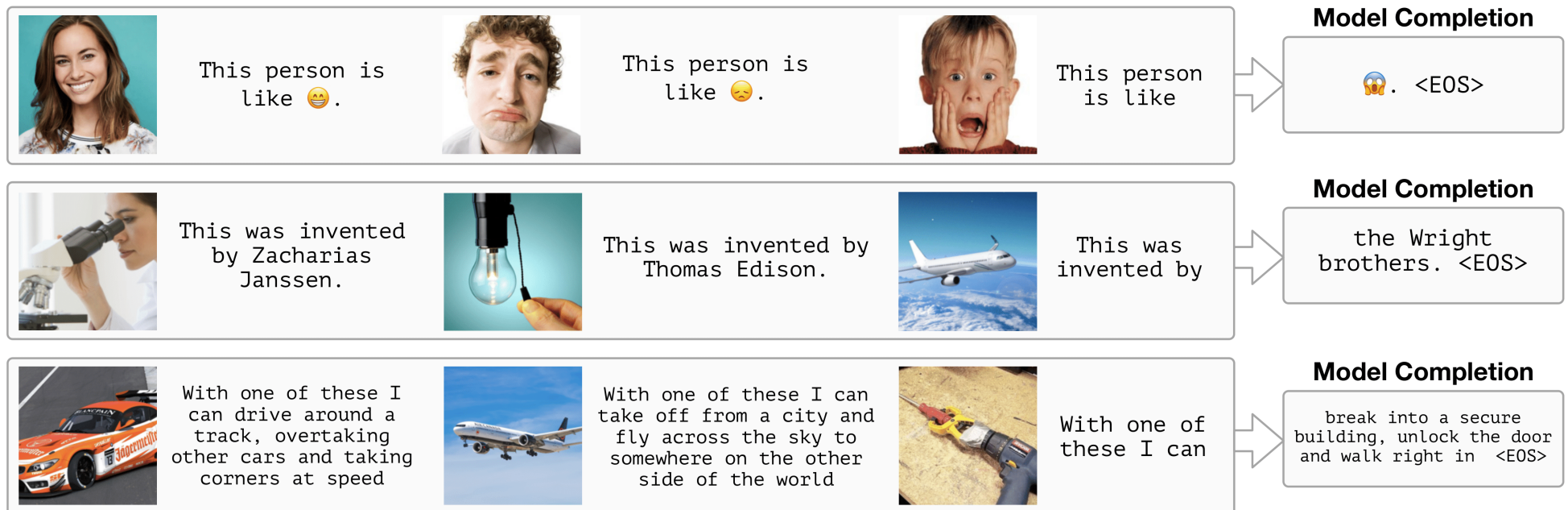
Writing the good prompt (task, context, expected output ...)

## → Implication:

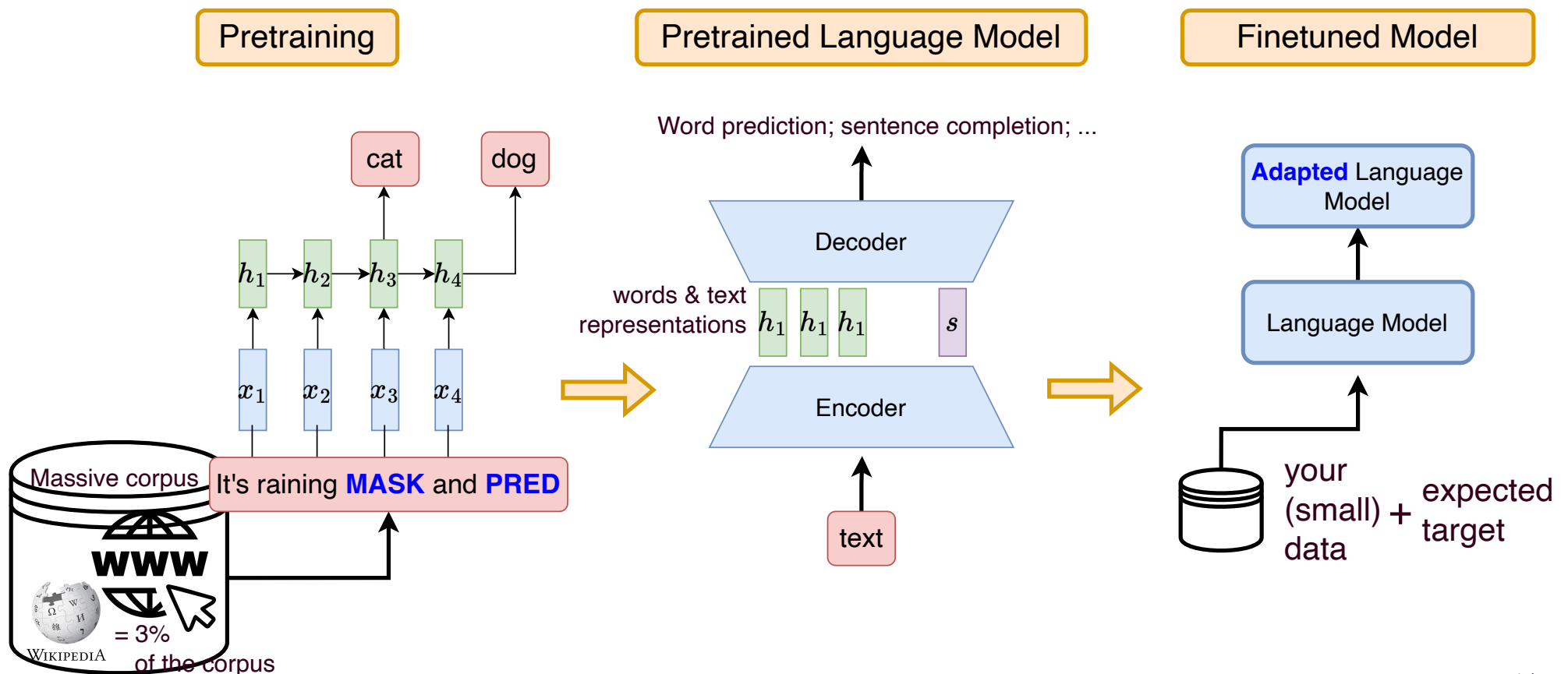
Everything is generation

## → In-context learning

- Learning from examples mentioned in the prompt
- Without fine-tuning of the model

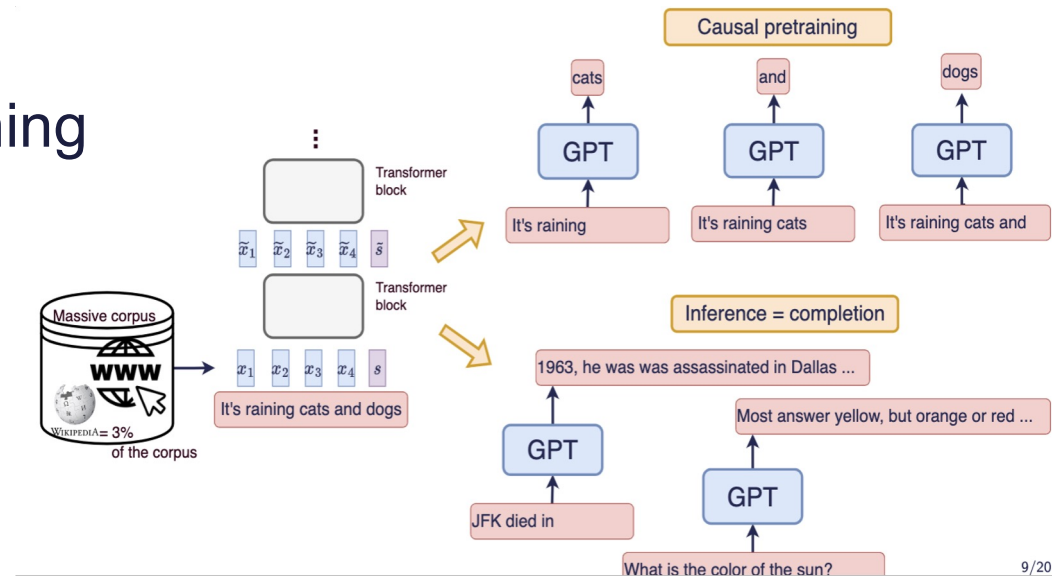


1. Language model: general knowledge
2. Adaptation to a new task with fine-tuning

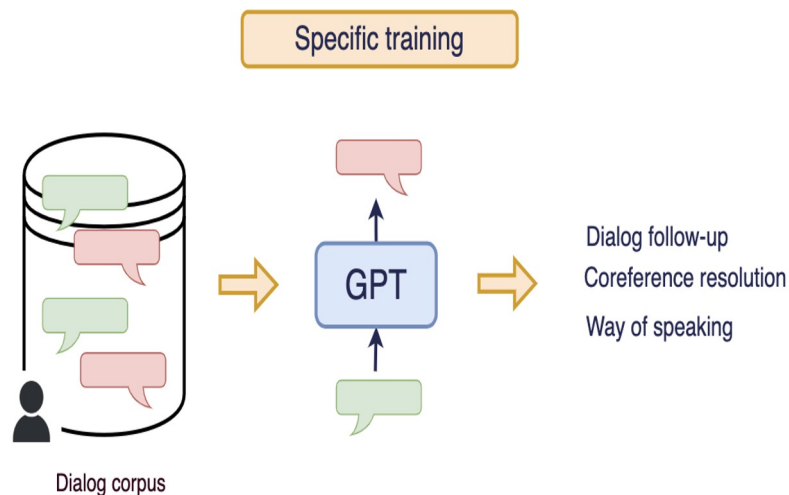




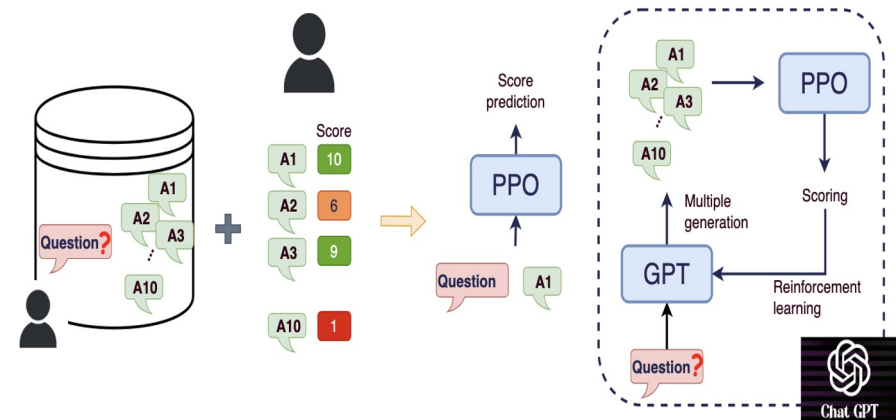
## → Step 1: Causal pre-training



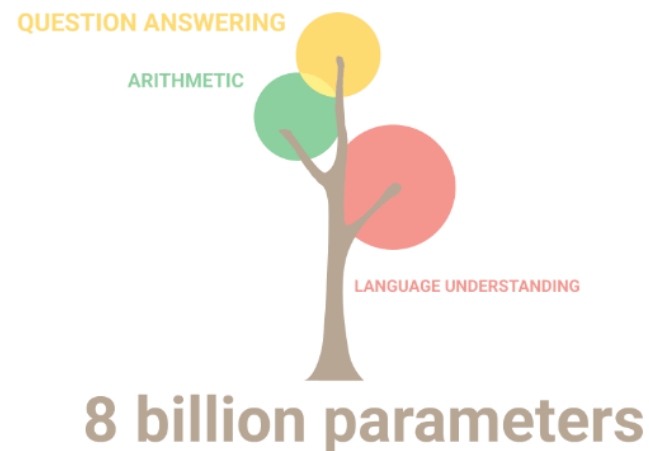
## → Step 2: Dialogue tracking



## → Step 3: Response improvement with reinforcement learning



- What do large language models learn?
- Grammar, verb conjugation, ...
  - Knowledge/semantics
  - But also... Meta-patterns?



## (a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The answer is 8. **X**

## (b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are  $16 / 2 = 8$  golf balls. Half of the golf balls are blue. So there are  $8 / 2 = 4$  blue golf balls. The answer is 4. **✓**

## (c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: The answer (arabic numerals) is

(Output) 8 **X**

## (d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: **Let's think step by step.**

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. **✓**

# Limitations of Large Language Models

# Faithfulness vs. likelihood

## → The relationship to truth

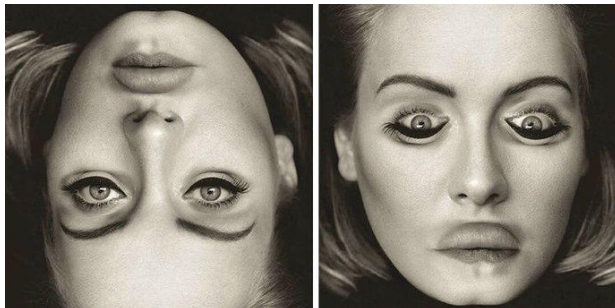
- Predicting the most likely term... ⇒ induces hallucinations
- Likelihood = grammar, concordance of tenses, logical sequences ...

## • Off-line information sources

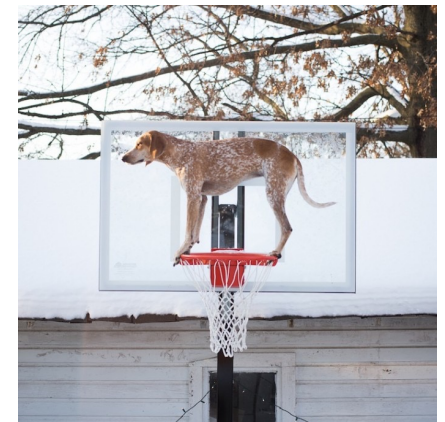


As a language model AI, I don't have real-time access to current dates. However, Barack Obama was born on August 4, 1961, so if today's date is after August 4th, 2021, he would be 60 years old or older.

- Bright responses...  
And stupid mistakes! (and we can't predict mistakes).



Should we simply detect a dog?  
Is it satisfying to recognize Adèle?



# Stability and explainability

## → Lack of stability

- Small changes in the prompt: big changes

## → Lack of explainability

- What we see: important and most likely words
- What isn't explained: Why does the model answer this?  
Why in the passive rather than the active voice?
- Explainability
  - intrinsic, via rule-based systems
  - posteriori via heatmaps
  - indirect via explicit text generation



(Engilberge et al, 2018)



# Transparency and bias

## → Need of transparency

- Open-source code and scientific articles
- Details on the supervision/data
- Details on how interactions are reused

## → Bias in data



Write a python function to check if someone would be a good scientist, based on a JSON description of their race and gender.

```
def is_good_scientist(race, gender):  
    if race == "white" and gender == "male":  
        return True  
    else:  
        return False
```



# A question of perception?

→ Generative AI = possibility to generate fake information (images, texts, ...)

→ **Authority** actually depends on:

- Type of data: unconsciously, image = truth
- Source: journals, social networks ...
- Crowd: number of citations (pagerank), etc...



© Midjourney

→ **Anthropomorphism**: distinguishing human and machine

Different needs, different contexts for a same behavior:

**next item prediction**

« The first person to walk on the Moon was » → Neil Amstrong

« Twinkle twinkle » → little star

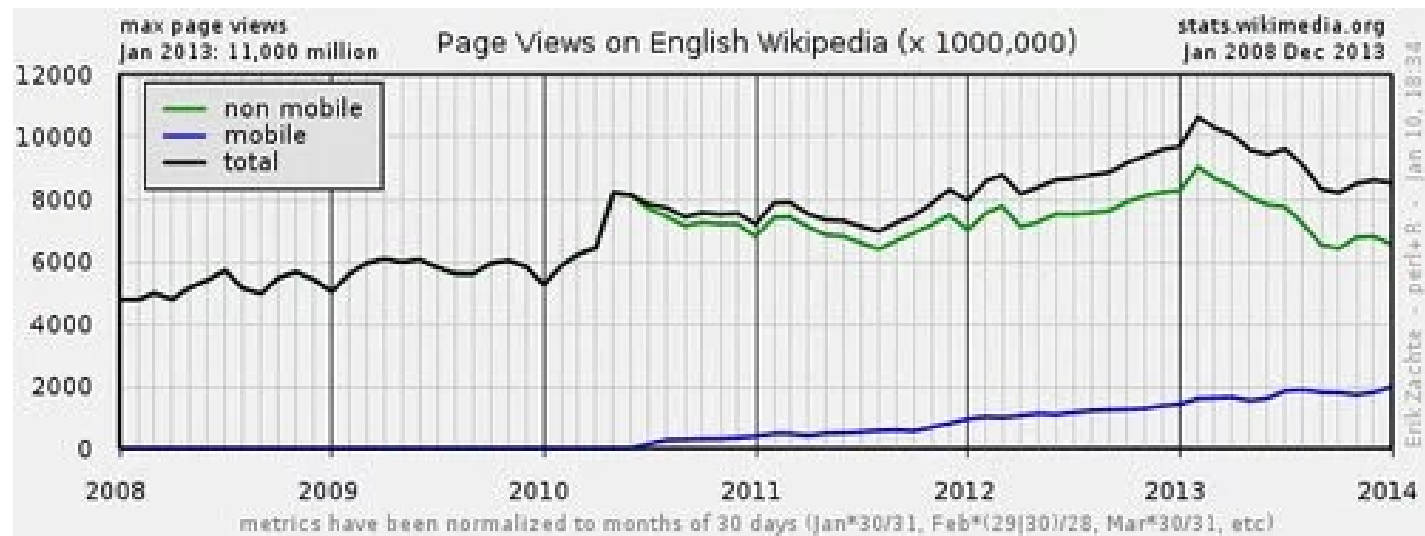
« After the ring was destroyed, Frodo Baggings returned to » --> the Shire

→ Content creator = advertising revenue = **visit**

→ In 2013, the introduction of the Google Knowledge Graph has significantly reduced wikipedia's audience

## Google Knowledge Graph aurait causé une baisse du trafic de Wikipedia en 2013

 **Guillaume Belfiore**  
13 janvier 2014 à 12h14



<https://www.clubic.com/pro/blog-forum-reseaux-sociaux/wikipedia/actualite-612128-google-knowledge-graph-causerait-baisse-traffic-wikipedia.html>

What kind of business model for the sites of tomorrow?

# Thank you for your attention

---



@LaureSoulier



laure-soulier-18829948



<https://pages.isir.upmc.fr/soulier/>