

A Wild ChatGPT appeared!

Georgios Margaritis & Periklis Petridis

15.S60, 2024



Image by DALLE

Disclaimer

This module is included for the first time in this course

v

It's more light-hearted and discussion-oriented (not as technical)

Let us know in surveys what you would have liked to see instead!

(or don't want to see again)

Who Knows ChatGPT?

Who Uses ChatGPT?

Who Uses ChatGPT:

**For random stuff?
(book recommendations, etc.)**

Who Uses ChatGPT:

**For work?
(plots, emails, etc.)**

Who likes / ChatGPT?

Who dislikes ChatGPT?

Plan

Concepts

- A little bit of science
- Is ChatGPT useful (and when)?
- How to utilize it

Discussion (Non-technical)

- Ethics
- Responsible use
- Outlook of humanity

So what is ChatGPT?

Simply

- A commercialized/fine-tuned GPT3.5/4 Model

A

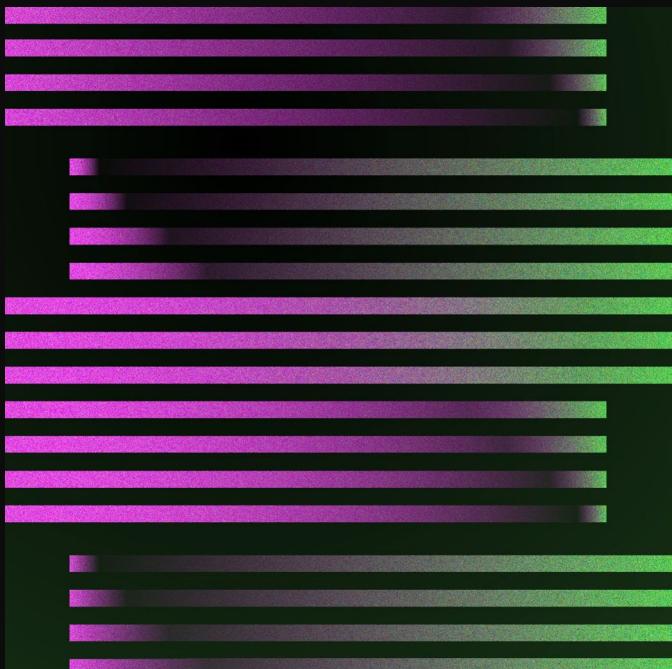
What is ChatGPT?



ChatGPT is an AI-powered chatbot developed by OpenAI. It is based on the GPT (Generative Pre-trained Transformer) architecture, specifically GPT-3.5. ChatGPT is designed to engage in text-based conversations with users, providing responses that are coherent, contextually relevant, and generated in a human-like manner.



The underlying technology behind ChatGPT is deep learning, specifically employing a transformer model. This architecture allows the model to process and understand the structure, context, and meaning of text input, enabling it to generate appropriate and informative responses.



So what is ChatGPT?

Simply

- A commercialized/fine-tuned GPT3.5/4 Model

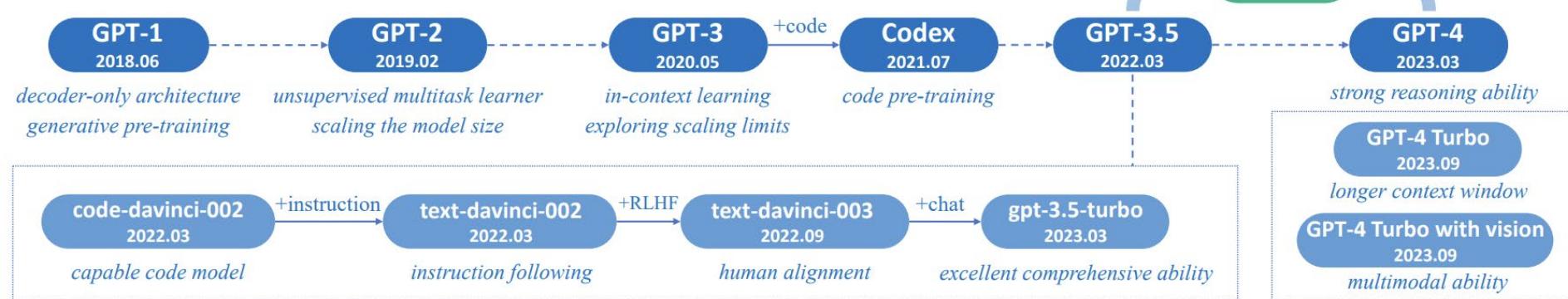


ChatGPT is an AI-powered chatbot developed by OpenAI. It is based on the GPT (Generative Pre-trained Transformer) architecture, specifically GPT-3.5. ChatGPT is designed to engage



How did we get here?

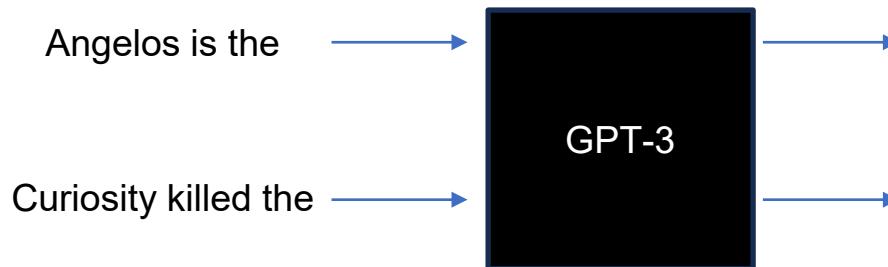
How did we get here?



Zhao, Wayne Xin, et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).

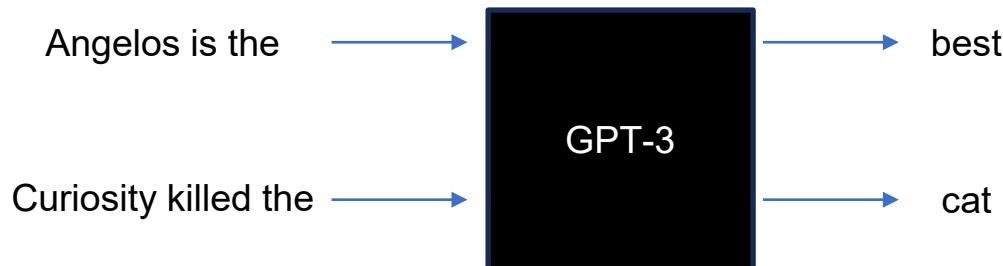
Brief explanation of GPT-3

- GPT stands for **Generative** Pre-trained Transformers
- GPT-3's job is to "predict the next" repeatedly
(this is called "Text **Generation**")



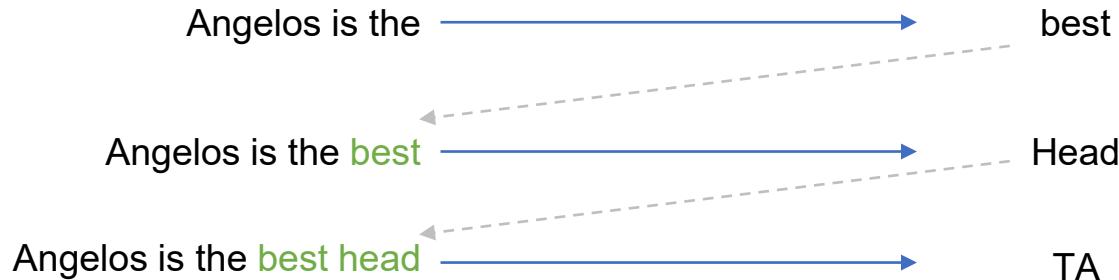
Brief explanation of GPT-3

- GPT stands for **Generative** Pre-trained Transformers
- GPT-3's job is to "predict the next" repeatedly
(this is called "Text **Generation**")



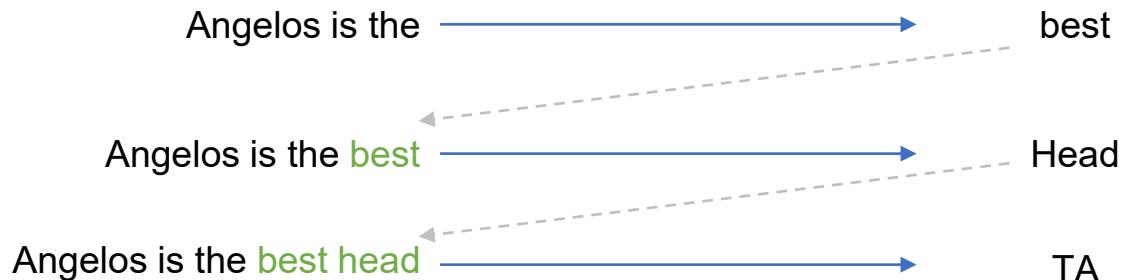
Brief explanation of GPT-3

- GPT stands for **Generative** Pre-trained Transformers
- GPT-3's job is to "predict the next" repeatedly
(this is called "Text **Generation**")

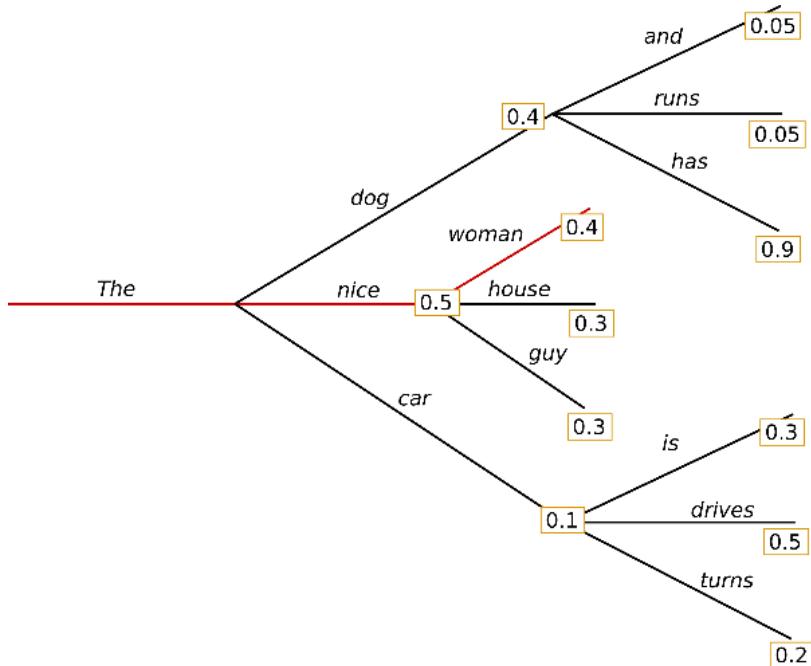


Brief explanation of GPT-3

- GPT stands for **Generative** Pre-trained Transformers
- GPT-3's job is to "predict the next" repeatedly
(this is called "Text **Generation**")

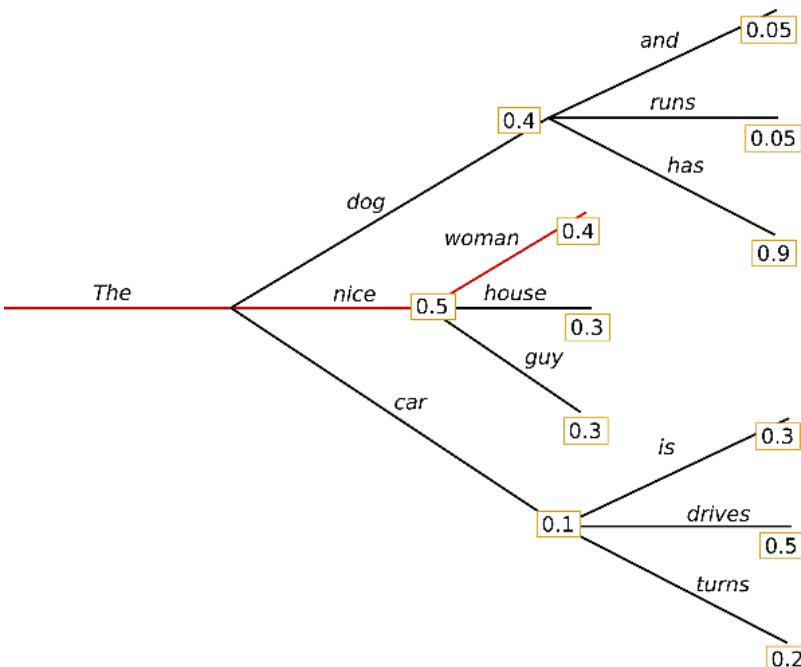


GPT-3 is a little smarter when generating text

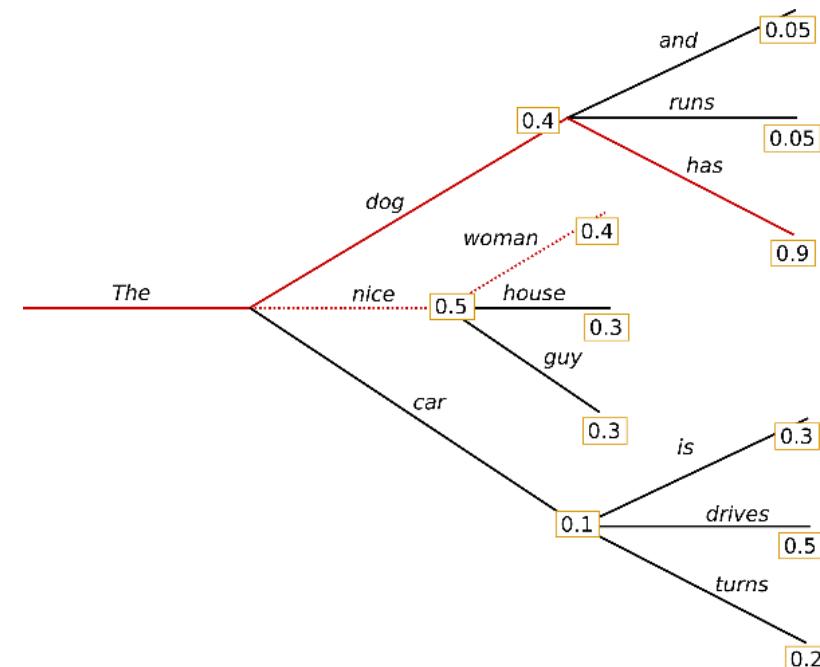


Greedy (select most probable next word)

GPT-3 is a little smarter when generating text



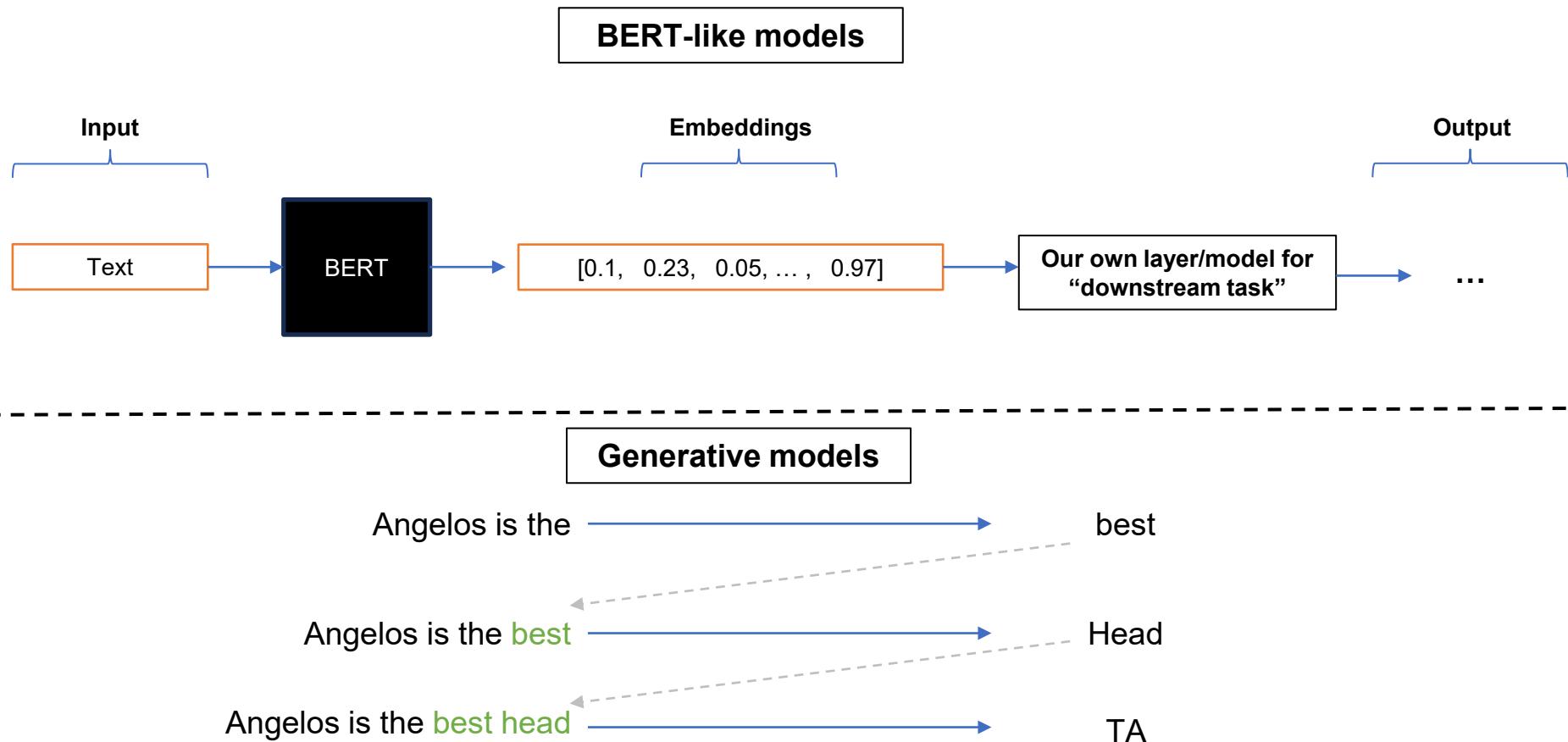
Greedy (select most probable next word)



Beam Search (select most probable sentence*)

Difference between Generative models and BERT

Difference between Generative models and BERT



Difference between Generative models and BERT

Focus:

- BERT – Contextual understanding
- GPT – Generative Capabilities

Difference between Generative models and BERT

Focus:

- BERT – Contextual understanding
- GPT – Generative Capabilities

Applications:

- BERT – Useful for tasks where **contextual understanding** is necessary, such as **Text Classification**
- GPT - Useful for **generative tasks**, such as **Text Completion, Summarization, Translation**

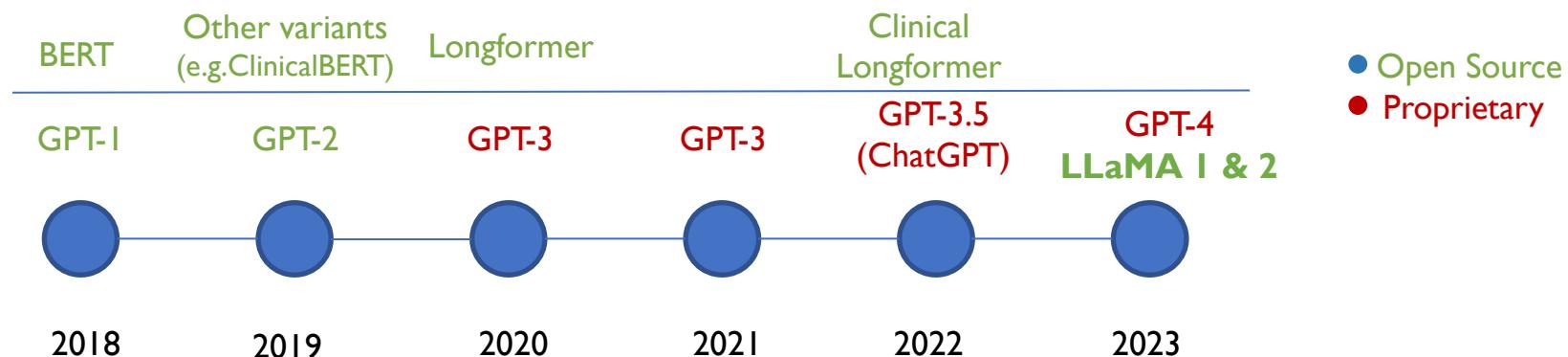
Difference between Generative models and BERT

Focus:

- BERT – Contextual understanding
- GPT – Generative Capabilities

Applications:

- BERT – Useful for tasks where **contextual understanding** is necessary, such as **Text Classification**
- GPT - Useful for **generative tasks**, such as **Text Completion, Summarization, Translation**



But wait...

But wait... there's more...

But wait... there's more...

The Open LLM Leaderboard aims to track, rank and evaluate open LLMs and chatbots.

Submit a model for automated evaluation on the GPU cluster on the "Submit" page! The leaderboard's backend runs the great EleutherAI Language Model Evaluation Harness - read more details in the "About" page!

LLM Benchmark Metrics through time About Submit here!

Search for your model (separate multiple queries with `;` and press ENTER...)

Select columns to show

Average	ARC	HellaSwag	MMLU	TruthfulQA	Winogrande	GSM8K	Type
<input checked="" type="checkbox"/>	<input type="checkbox"/>						
<input type="checkbox"/>	<input type="checkbox"/>						

Model types

pretrained	fine-tuned	instruction-tuned	RL-tuned	?
<input checked="" type="checkbox"/>				
<input type="checkbox"/>				

Precision

float16	bfloat16	8bit	4bit	GPTQ	?
<input checked="" type="checkbox"/>					
<input type="checkbox"/>					

Model sizes (in billions of parameters)

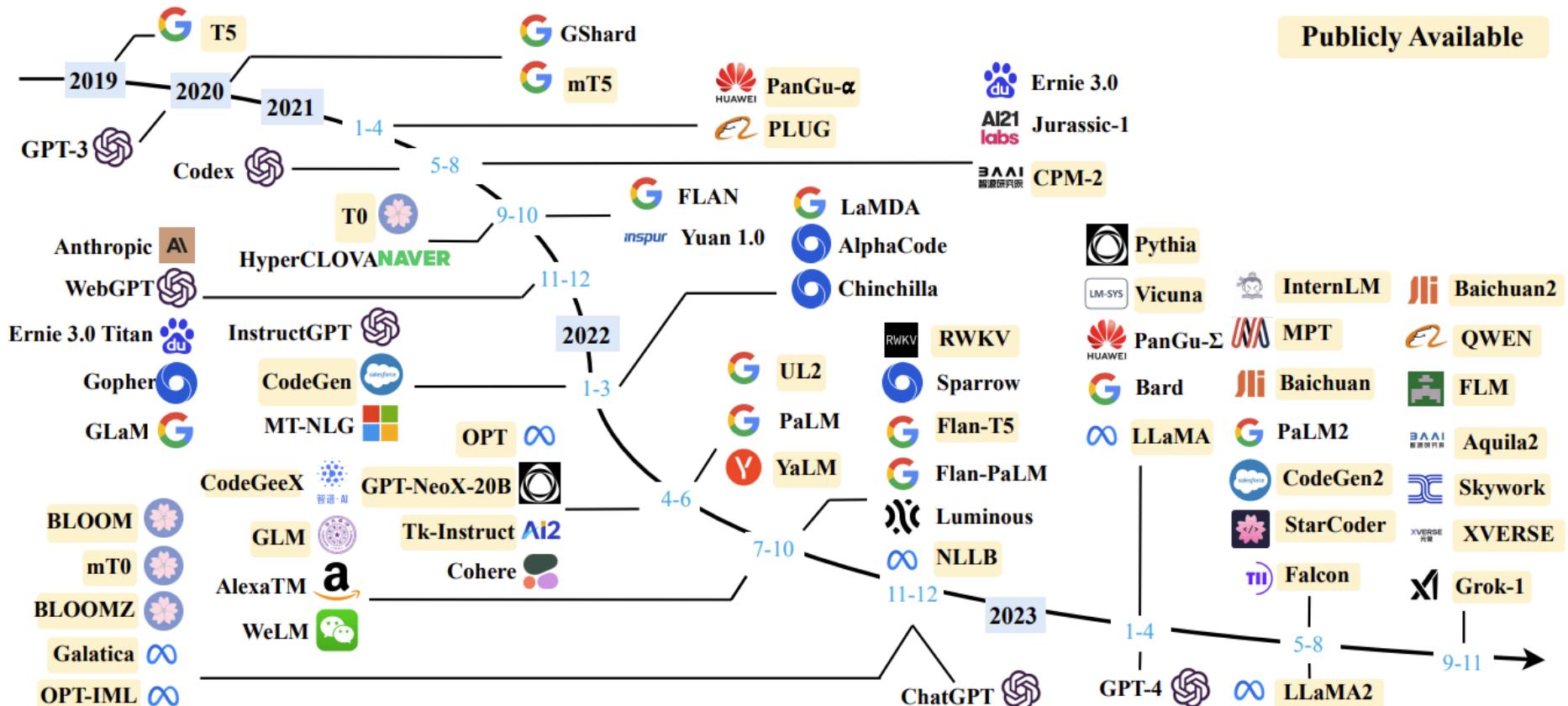
?	~1.5	~3	~7	~13	~35	~60	~70+
<input checked="" type="checkbox"/>							
<input type="checkbox"/>							

T	Model	Average	ARC	HellaSwag	MMLU	TruthfulQA	Winogrande	GSM8K
●	cloudyu/Yi-34Bx2-MoE-60B	76.72	71.08	85.23	77.47	66.19	84.85	75.51
●	cloudyu/Mixtral_34Bx2_MoE_60B	76.66	71.33	85.25	77.34	66.59	84.85	74.6
●	cloudyu/Mixtral_34Bx2_MoE_60B	76.63	71.25	85.36	77.28	66.61	84.69	74.6
■	moreh/MoMo-70B-lora-1.8.4-DPO	76.23	69.62	85.35	77.33	64.64	84.14	76.27
◆	cloudyu/Yi-34Bx3-MoE-90B	76.18	70.9	85.33	77.41	66.31	84.29	72.86
■	moreh/MoMo-70B-lora-1.8.5-DPO	76.14	69.54	85.6	77.49	65.79	84.14	74.3
◆	TomGrc/FusionNet_7Bx2_MoE_14B	75.91	73.55	88.84	64.68	69.6	88.16	70.66
◆	one-man-army/UNA-34Beagles-32K-bf16-v1	75.41	73.55	85.93	76.45	73.55	82.95	60.05
■	mLabonne/NeuralPeggle14-7B	74.74	72.95	88.34	64.55	69.93	82.4	70.28
◆	jondurbin/nontoxic_bage1-34b-v0.2	74.69	72.44	85.64	76.41	72.7	82.48	58.45
●	jondurbin/bagel-dpo-34b-v0.2	74.69	71.93	85.25	76.58	70.05	83.35	60.96
◆	moreh/MoMo-70B-LoRA-V1.4	74.67	69.2	85.07	77.12	62.66	83.74	70.2

Citation

https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

But wait... there's more... loads of more

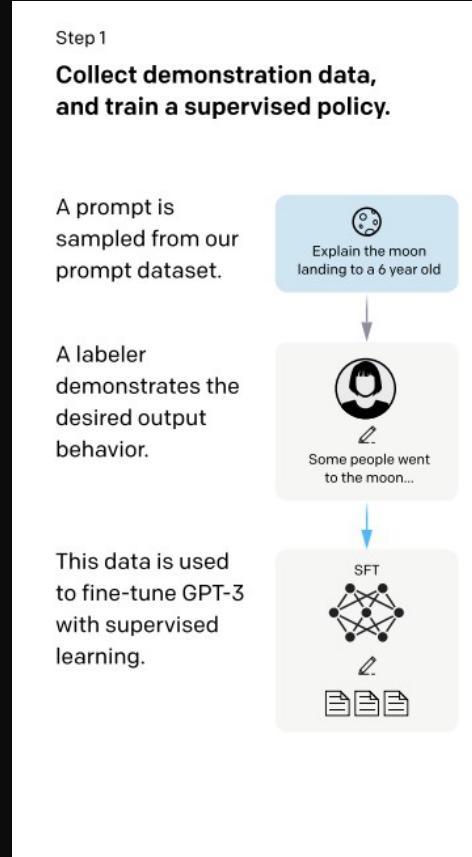


Back to GPT

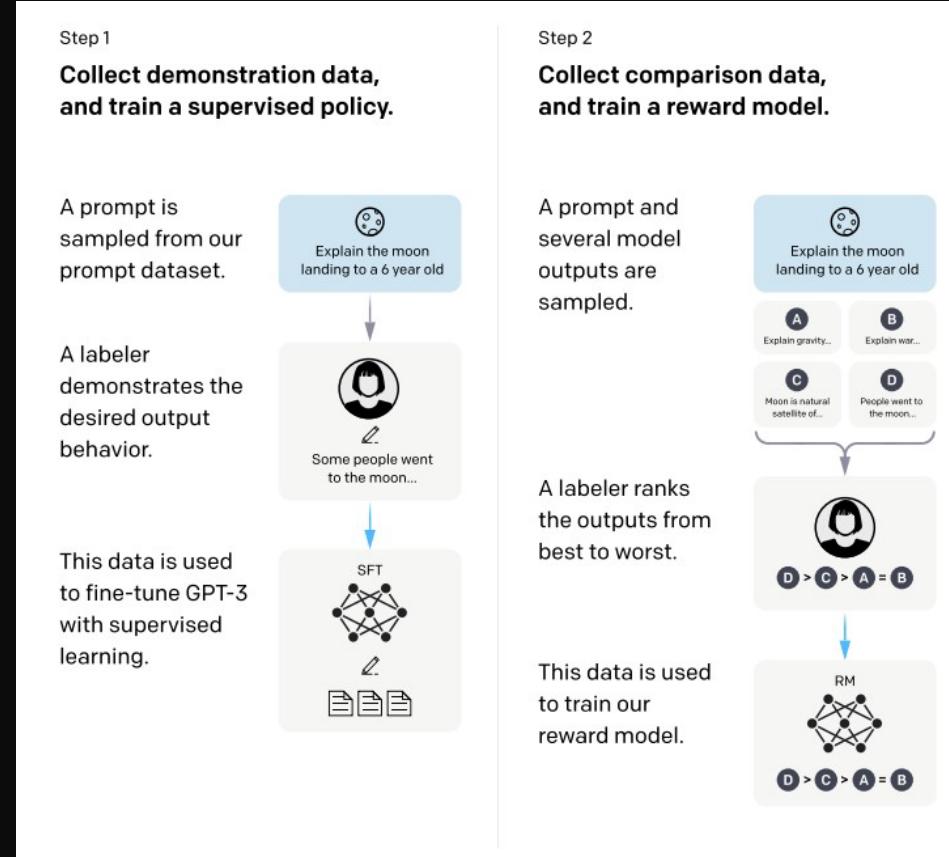
Two more interesting Concepts

- Training ChatGPT
- Chain of Thought Prompting (and extensions)

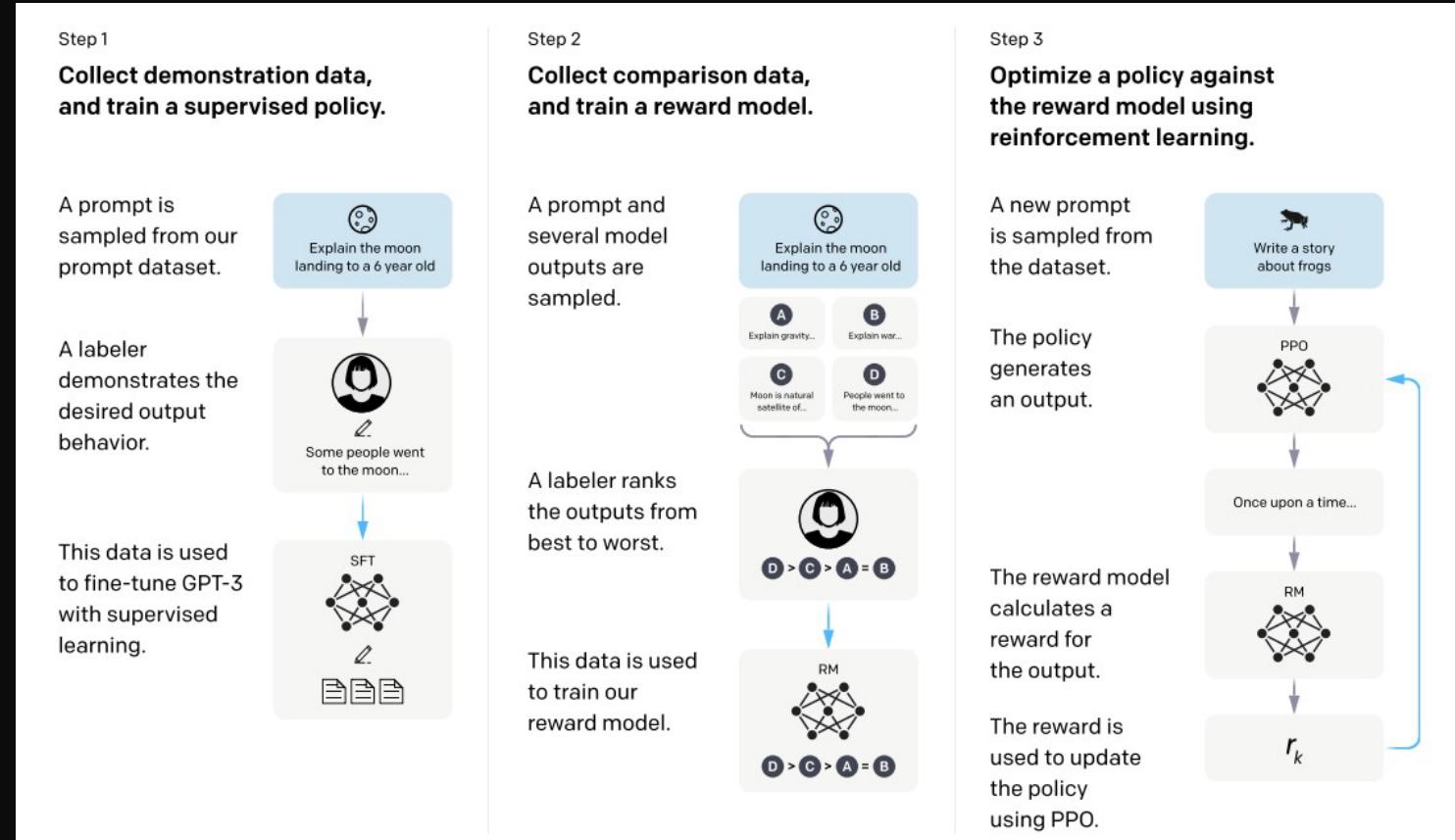
Training ChatGPT



Training ChatGPT



Training ChatGPT



Chain of Thought

We can help ChatGPT answer better!

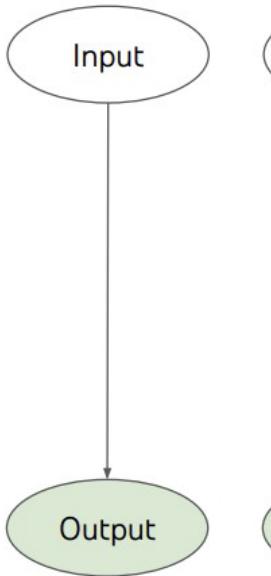
Intuition:

- Even humans don't immediately give the final answer
- Instead, they **reason** (with **thoughts**)

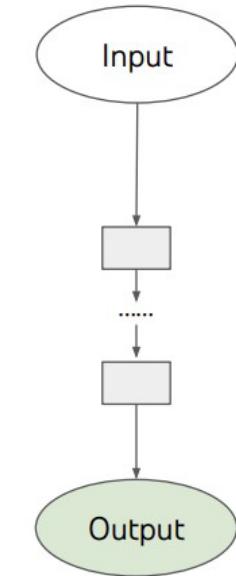
Chain of Thought

Intuition:

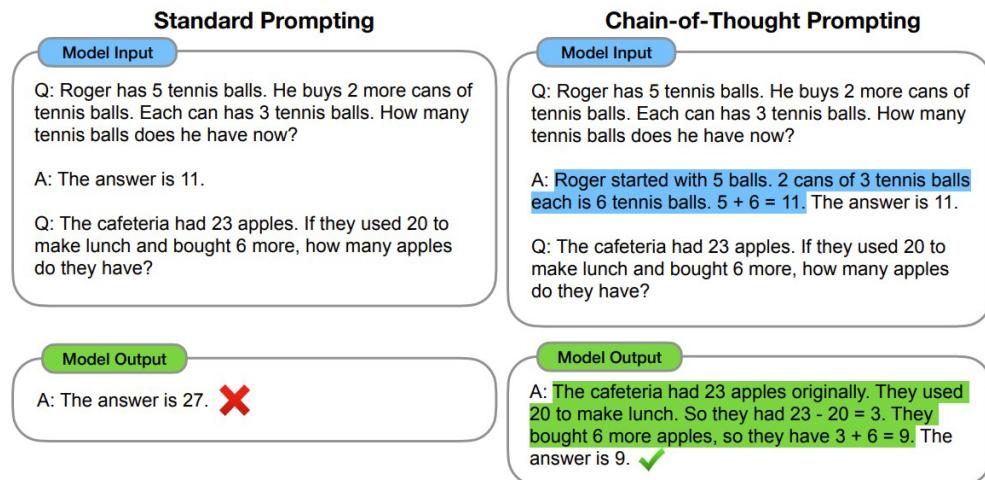
- Even humans don't immediately give the final answer
- Instead, they **reason (with thoughts)**



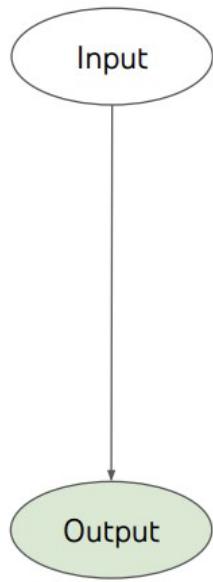
(a) Input-Output
Prompting (IO)



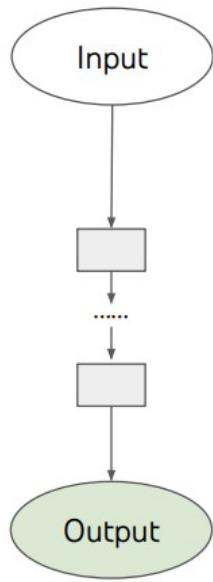
(c) Chain of Thought
Prompting (CoT)



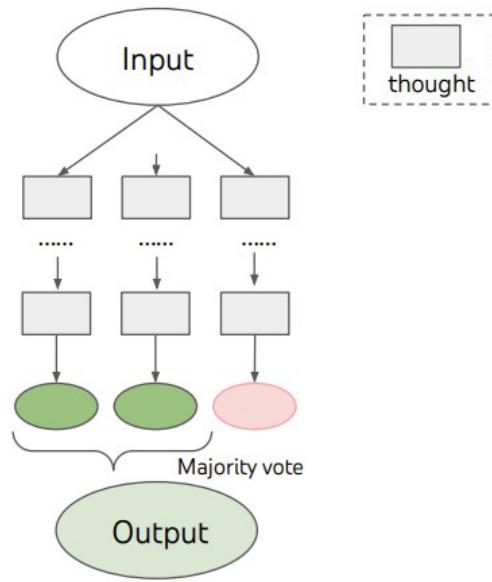
Chain of Thought



(a) Input-Output
Prompting (IO)



(c) Chain of Thought
Prompting (CoT)

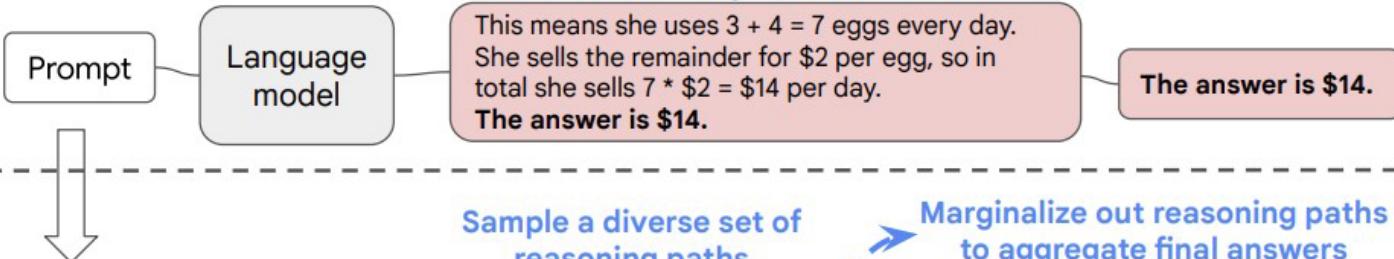


(c) Self Consistency
with CoT (CoT-SC)



Chain of Thought

Chain-of-thought prompting



Q: If there are 3 cars in the parking lot and 2 more cars arrive, how many cars are in the parking lot?
A: There are 3 cars in the parking lot already. 2 more arrive. Now there are $3 + 2 = 5$ cars. The answer is 5.
...
Q: Janet's ducks lay 16 eggs per day. She eats three for breakfast every morning and bakes muffins for her friends every day with four. She sells the remainder for \$2 per egg. How much does she make every day?
A:

Language model

Sample a diverse set of reasoning paths

Marginalize out reasoning paths to aggregate final answers

She has $16 - 3 - 4 = 9$ eggs left. So she makes $\$2 * 9 = \18 per day.

The answer is \$18.

This means she sells the remainder for $\$2 * (16 - 4 - 3) = \26 per day.

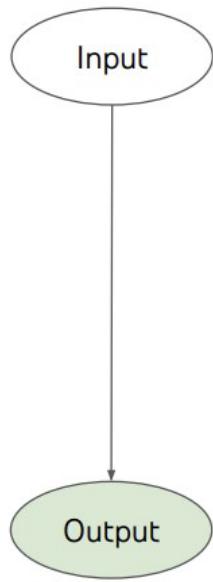
The answer is \$26.

She eats 3 for breakfast, so she has $16 - 3 = 13$ left. Then she bakes muffins, so she has $13 - 4 = 9$ eggs left. So she has $9 \text{ eggs} * \$2 = \18 .

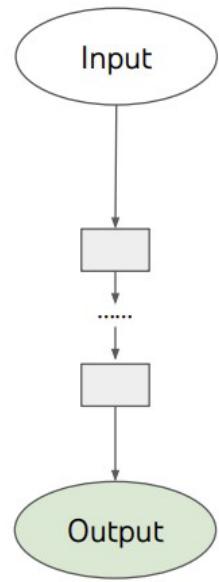
The answer is \$18.

The answer is \$18.

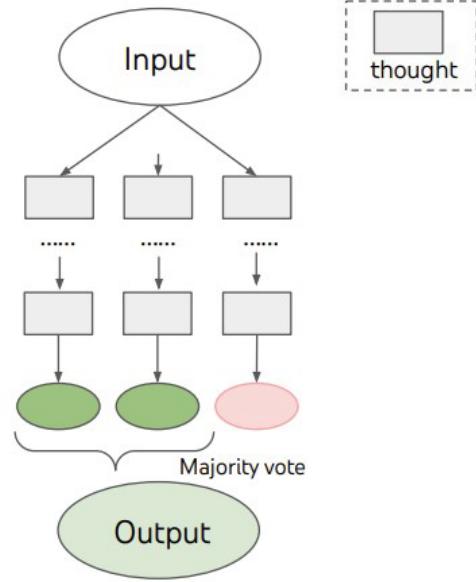
Chain of Thought



(a) Input-Output
Prompting (IO)

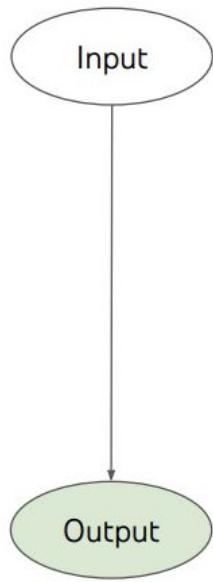


(c) Chain of Thought
Prompting (CoT)

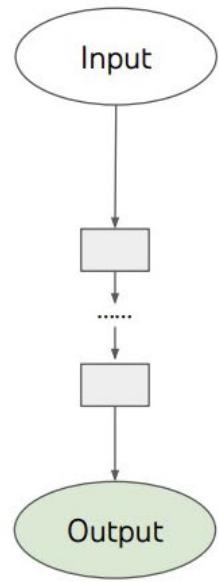


(c) Self Consistency
with CoT-SC (CoT-SC)

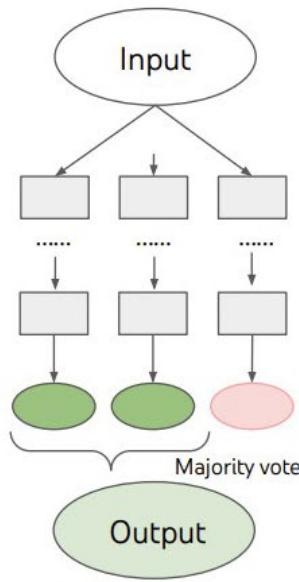
Chain of Thought



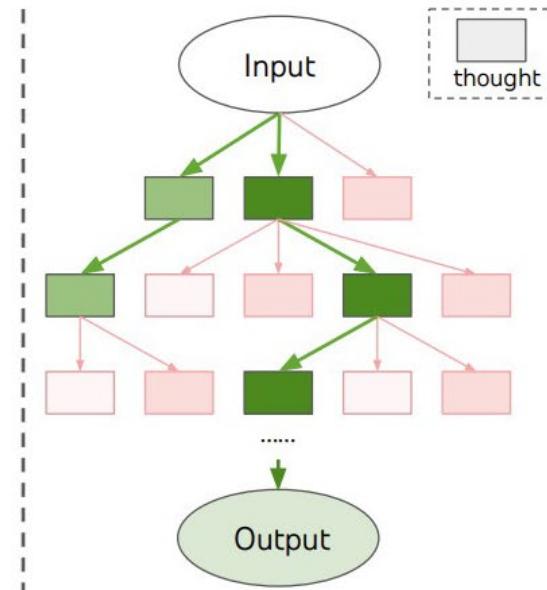
(a) Input-Output
Prompting (IO)



(c) Chain of Thought
Prompting (CoT)

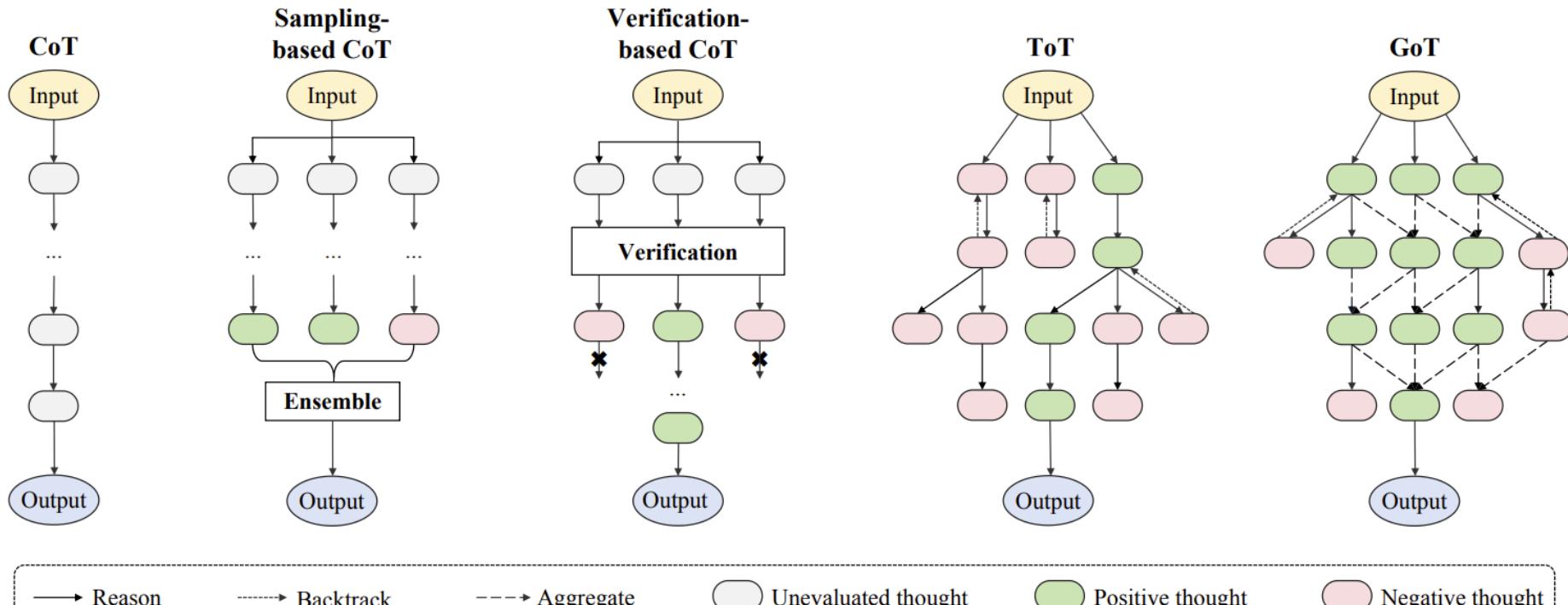


(c) Self Consistency
with CoT (CoT-SC)



(d) Tree of Thoughts (ToT)

Chain of Thought (loads of variants)



Mixture of Experts (MOE)

Can we improve ChatGPT further?

Mixture of Experts (MOE)

Can we improve ChatGPT further?

Intuition

- What if we didn't have a single GPT?
- What if we had **one** GPT for **each task** we want to do?
(e.g. code, emails, images, Harry Potter fanfic, etc.)

Mixture of Experts (MOE)

Intuition

- What if we didn't have a single GPT?
- What if we had **one** GPT for **each task** we want to do?
(e.g. code, emails, images, Harry Potter fanfic, etc.)
- That's what ChatGPT-4 is (**rumored to be***) doing

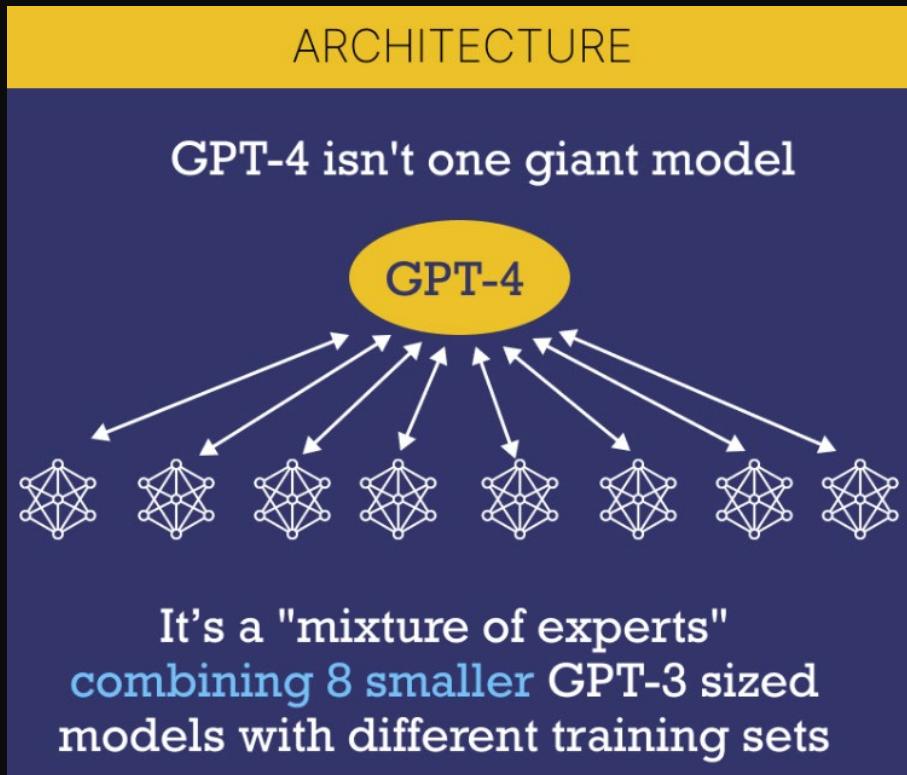
Mixture of Experts (MOE)

Intuition

- What if we didn't have a single GPT?
- What if we had **one** GPT for **each task** we want to do?
(e.g. code, emails, images, Harry Potter fanfic, etc.)
- That's what ChatGPT-4 is (**rumored to be***) doing

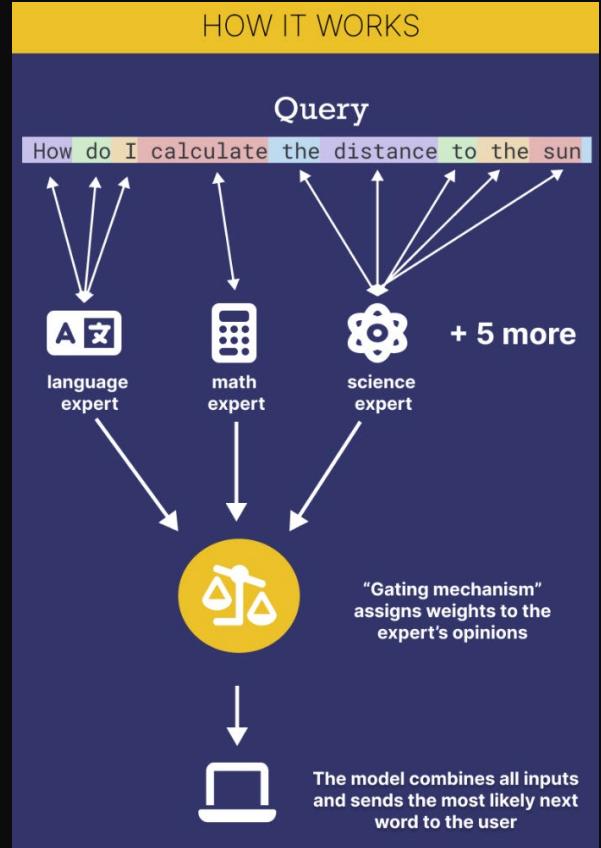
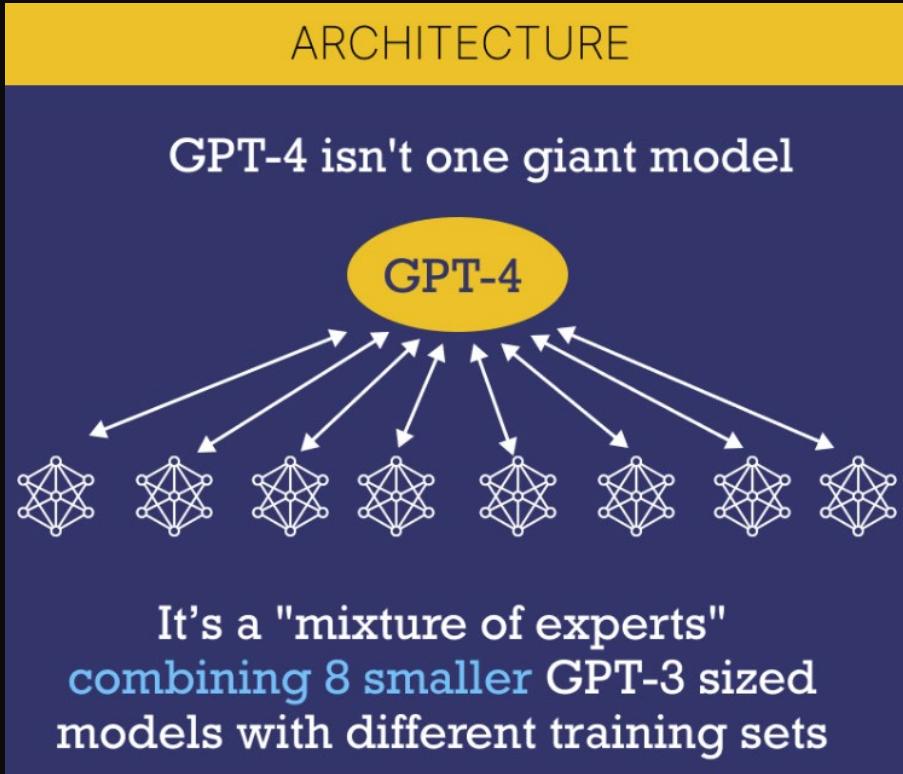
*XOXO,
-Gossip girl

Mixture of Experts (MOE)



*Hypothetical architecture

Mixture of Experts (MOE)



Plan

Concepts

- A little bit of science
- Is ChatGPT useful (and when)?
- How to utilize it

Discussion (Non-technical)

- Ethics
- Responsible use
- Outlook of humanity

Is ChatGPT useful (and when)?

Is ChatGPT useful (and when)?

Who thinks it *is* useful?

Is ChatGPT useful (and when)?

Who thinks it *isn't* useful?

Is ChatGPT useful (and when)?

Some conventional wisdom (possibly from twitter)

1. Unless you understand enough about the field you're asking about, you shouldn't trust GPT's answer

Is ChatGPT useful (and when)?

Some conventional wisdom (possibly from twitter)

1. Unless you understand enough about the field you're asking about, you shouldn't trust GPT's answer
2. Schrodingers GPT:
ChatGPT is both **less useful** than people claim

Is ChatGPT useful (and when)?

Some conventional wisdom (possibly from twitter)

1. Unless you understand enough about the field you're asking about, you shouldn't trust GPT's answer
2. Schrodingers GPT:
ChatGPT is both **less useful** than people claim and **more useful** than people* claim

Cold Hard Facts*

- ChatGPT is **not creative**

*Actually, some are just my (cold hard) opinions

Cold Hard Facts*

- ChatGPT is **not creative**
- If ChatGPT generates something, this exists somewhere on the internets

*Actually, some are just my (cold hard) opinions

Cold Hard Facts*

- ChatGPT is **not creative**
- If ChatGPT generates something, this exists somewhere on the internets
- ChatGPT is only as useful as your prompts.

Cold Hard Facts*

- ChatGPT is **not creative**
- If ChatGPT generates something, this exists somewhere on the internets
- ChatGPT is only as useful as your prompts.
- ChatGPT is simply not very good (yet) at some things

*Actually, some are just my (cold hard) opinions

Cold Hard Facts*

- ChatGPT is **not creative**
- If ChatGPT generates something, this exists somewhere on the internets
- ChatGPT is only as useful as your prompts.
- ChatGPT is simply not very good (yet) at some things
- ChatGPT is not magic.

*Actually, some are just my (cold hard) opinions

Cold Hard Facts*

- ChatGPT is **not creative**
- If ChatGPT generates something, this exists somewhere on the internets
- ChatGPT is only as useful as your prompts.
- ChatGPT is simply not very good (yet) at some things
- ChatGPT is not magic.
- **You have to work to make ChatGPT work**

*Actually, some are just my (cold hard) opinions

Cold Hard Facts*

- ChatGPT is **not magic**.
- **You** have to work to make **ChatGPT** work

*Actually, some are just my (cold hard) opinions

Limitations

Limitations

- Writing

Limitations

- Writing



Limitations

- Writing
 - Its really bad at creative/academic writing

Me spending more time trying to make ChatGPT write an email than it would have taken me to do it myself 😤



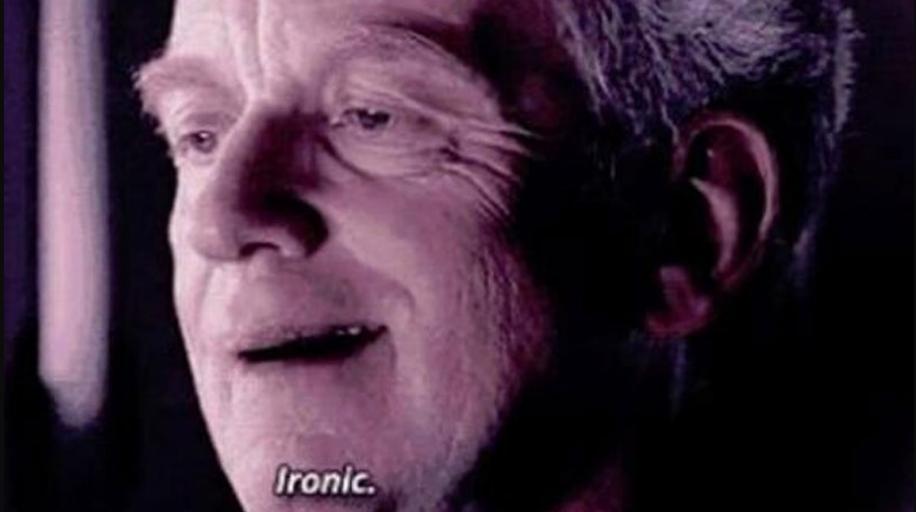
Limitations

- Writing
 - Its really bad at creative/academic writing
- Thinking



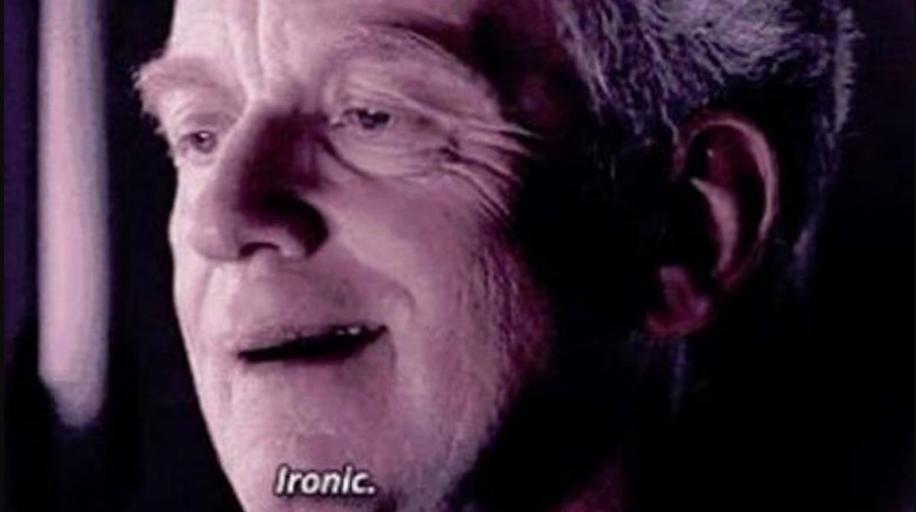
Limitations

- Writing
 - Its really bad at creative/academic writing
- Thinking
- Math



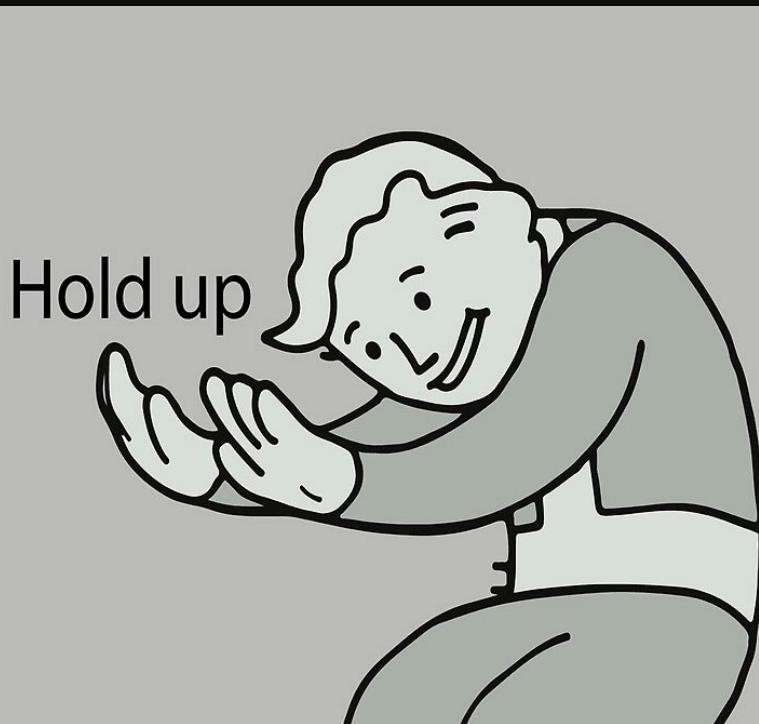
Limitations

- Writing
 - Its really bad at creative/academic writing
- Thinking
- Math
- Software Engineering



Limitations

- Writing
 - Its really bad at creative/academic writing
- Thinking
- Math
- Software Engineering



Hold up, that's what a PhD is all about! Then why did we lose our time today??

Strengths

Strengths

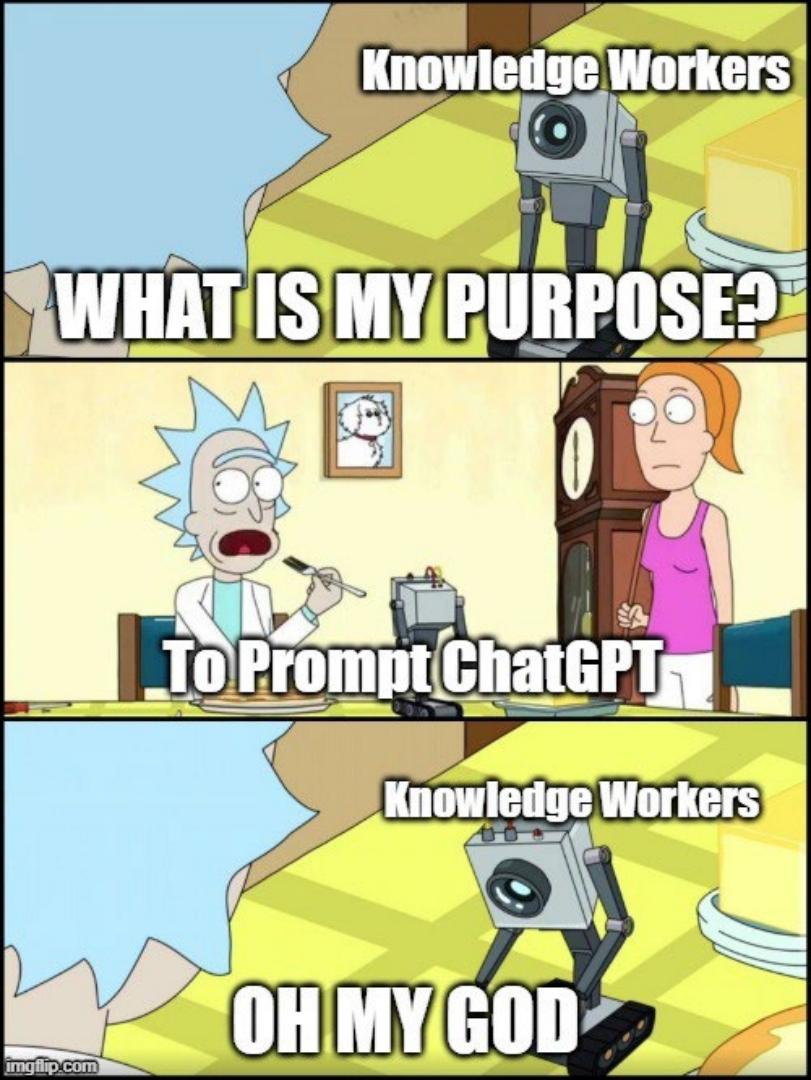
- It can LeetCode better (and faster) than you
- If you need to look at StackOverflow for something, ChatGPT already knows the answer.
- If you need to look at the documentation of a popular language/library that hasn't changed much since 2021... Chatgpt already knows the answer.

Strengths

- It can LeetCode better (and faster) than you
- If you need to look at StackOverflow for something, ChatGPT already knows the answer.
- If you need to look at the documentation of a popular language/library that hasn't changed much since 2021... Chatgpt already knows the answer.
- If you want to write bad Harry Potter fanfic, ChatGPT is a specialist

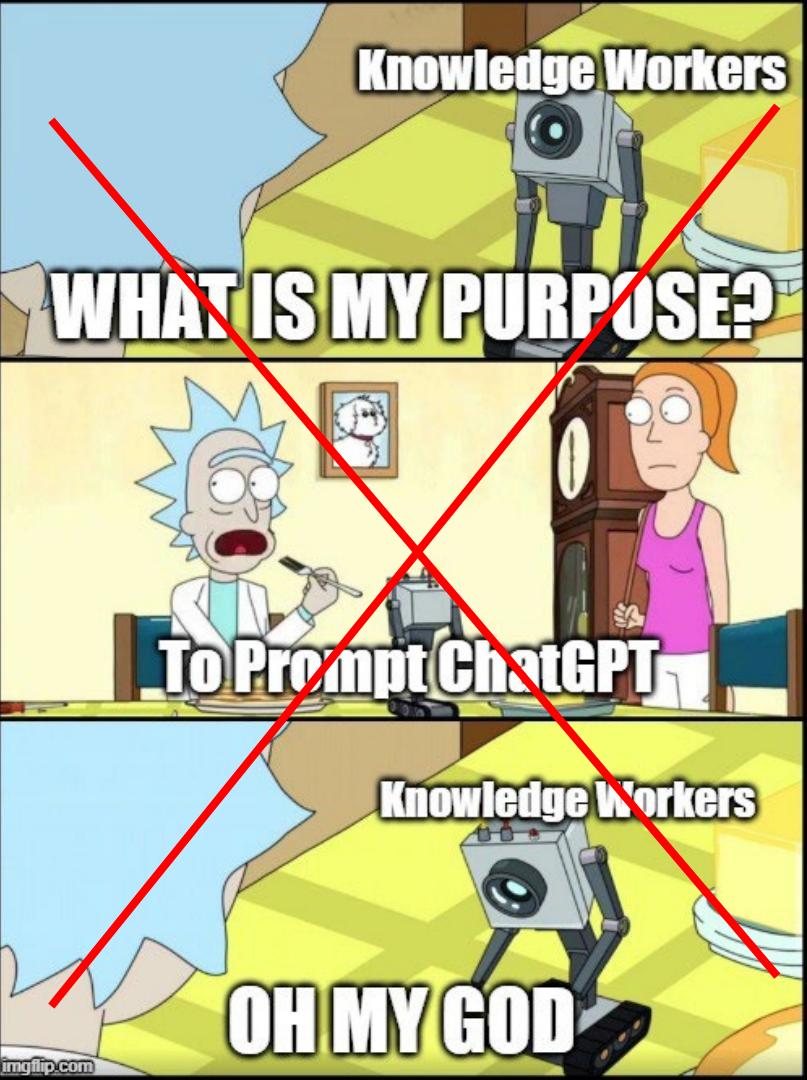
Are we doomed?

Are we doomed?



Are we doomed?

No!



How to use ChatGPT? (Discussion)*

- A PhD requires “big picture” work/creativity, but also a lot of “small picture” (rote, time consuming tasks)
 - Data cleaning, Data visualization
 - Basic ML
 - Basic Optimization
 - “Software engineering”
 - Emails, Googling
 - Etc.
- These are necessary, but in PhD you also have to think (what, why, how)

Opinion

- A PhD requires “big picture” work/creativity, but also a lot of “small picture” (rote, time consuming tasks)
 - Data cleaning, Data visualization
 - Basic ML
 - Basic Optimization
 - “Software engineering”
 - Emails, Googling
 - Etc.
- These are necessary, but in PhD you also have to think (what, why, how)
- Typically, many PhDs were 80% “small picture”, 20% “big picture”
- Tools like ChatGPT reduce the “80%” part, **never** the 20%.
i.e., they give you more room to think “big picture”

Opinion

- A PhD requires “big picture” work/creativity, but also a lot of “small picture” (rote, time consuming tasks)
 - Data cleaning, Data visualization
 - Basic ML
 - Basic Optimization
 - “Software engineering”
 - Emails, Googling
 - Etc.
- These are necessary, but in PhD you also have to think (what, why, how)
- Typically, many PhDs were 80% “small picture”, 20% “big picture”
- Tools like ChatGPT reduce the “80%” part, **never** the 20%.
i.e., they give you more room to think “big picture”
- Use them, and **treat** them, as productivity tools

Opinion

- A PhD requires “big picture” work/creativity, but also a lot of “small picture” (rote, time consuming tasks)
 - Data cleaning, Data visualization
 - Basic ML
 - Basic Optimization
 - “Software engineering”
 - Emails, Googling
 - Etc.
- These are necessary, but in PhD you also have to think (what, why, how)
- Typically, many PhDs were 80% “small picture”, 20% “big picture”
- Tools like ChatGPT reduce the “80%” part, **never** the 20%.
i.e., they give you more room to think “big picture”
- Use them, and **treat** them, as productivity tools
- Example: You would not calculate simplex by hand, you should use Gurobi
You should not code basic things from scratch, you should use ChatGPT

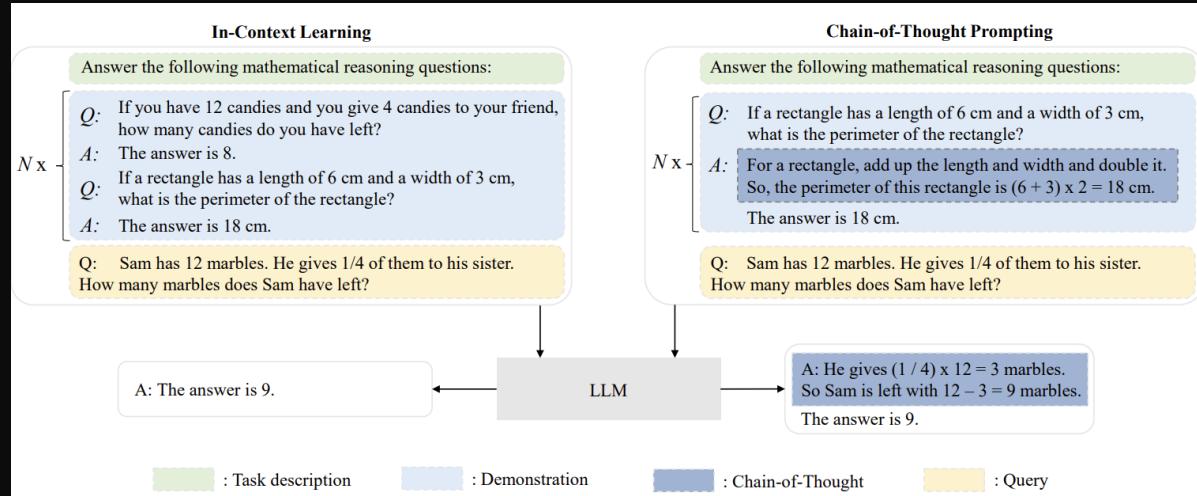
How to prompt good

How to prompt good



How to prompt good

- Utilize chain of thought etc.
 - Use examples
 - Be very explicit in how you want the answer



How to prompt good

- Utilize chain of thought etc.
 - Use examples
 - Be very explicit in how you want the answer
- Resources on the interwebs

How to prompt good

- Utilize chain of thought etc.
 - Use examples
 - Be very explicit in how you want the answer
- Resources on the interwebs
- Use ChatGPT to write prompts for ChatGPT



How to prompt good (A hidden trick no-one wants you to know)

- Emotional Manipulation

How to prompt good (A hidden trick no-one wants you to know)

A screenshot of a dark-themed AI interface. At the top, a user prompt asks for a JSON dictionary with two keys. Below it, a AI-generated response provides an example JSON code snippet. A note below the code explains its contents. At the bottom, there are interaction icons like thumbs up, thumbs down, and a share button.

Give an example of a JSON dictionary with two keys. Do not include any text before or after the JSON. edit

View other drafts ▾

Sure, here is an example of a JSON dictionary with two keys:

Code snippet

```
{  
  "name": "John Doe",  
  "age": 30  
}
```

Use code with caution. [Learn more](#) copy

This dictionary has two keys: `name` and `age`. The value of the `name` key is "John Doe" and the value of the `age` key is 30.

Upvote Downvote Share ⋮

How to prompt good (A hidden trick no-one wants you to know)



Give an example of a JSON dictionary with two keys. Do not include any text before or after the JSON. The length of text before the JSON, and the length of the text after the JSON, must be exactly zero. You must write no text whatsoever other than JSON.



Here some example of things you should not say:

- "Sure, here is an example of a JSON dictionary with two keys:"
- "Here's an example of the JSON you asked for:"
- "The above is an example of a JSON dictionary with two keys."

Do not say these things! Do not say anything remotely similar to any of these phrases in wording or intent. Remember, no text! Emit exactly zero words before your JSON, and exactly zero words after. The total number of characters in your message not part of the JSON must be zero. Your entire message must be JSON. Only return JSON. Do not return text.

How to prompt good (A hidden trick no-one wants you to know)

Sure, here is an example of a JSON dictionary with two keys:

```
{  
  "key1": "value1",  
  "key2": "value2"  
}
```

Use code with caution. [Learn more](#)

View other drafts ▾

Share icon

Upvote icon

Up arrow icon

Google it icon

More options icon

How to prompt good (A hidden trick no-one wants you to know)



Give an example of a JSON dictionary with two keys. Do not include any text before or after the JSON. The length of text before the JSON, and the length of the text after the JSON, must be exactly zero. You must write no text whatsoever other than JSON.



Here some example of things you should not say:

- "Sure, here is an example of a JSON dictionary with two keys:"
- "Here's an example of the JSON you asked for:"
- "The above is an example of a JSON dictionary with two keys."

Do not say these things! Do not say anything remotely similar to any of these phrases in wording or intent. Remember, no text! Emit exactly zero words before your JSON, and exactly zero words after. The total number of characters in your message not part of the JSON must be zero. Your entire message must be JSON. Only return JSON. Do not return text.

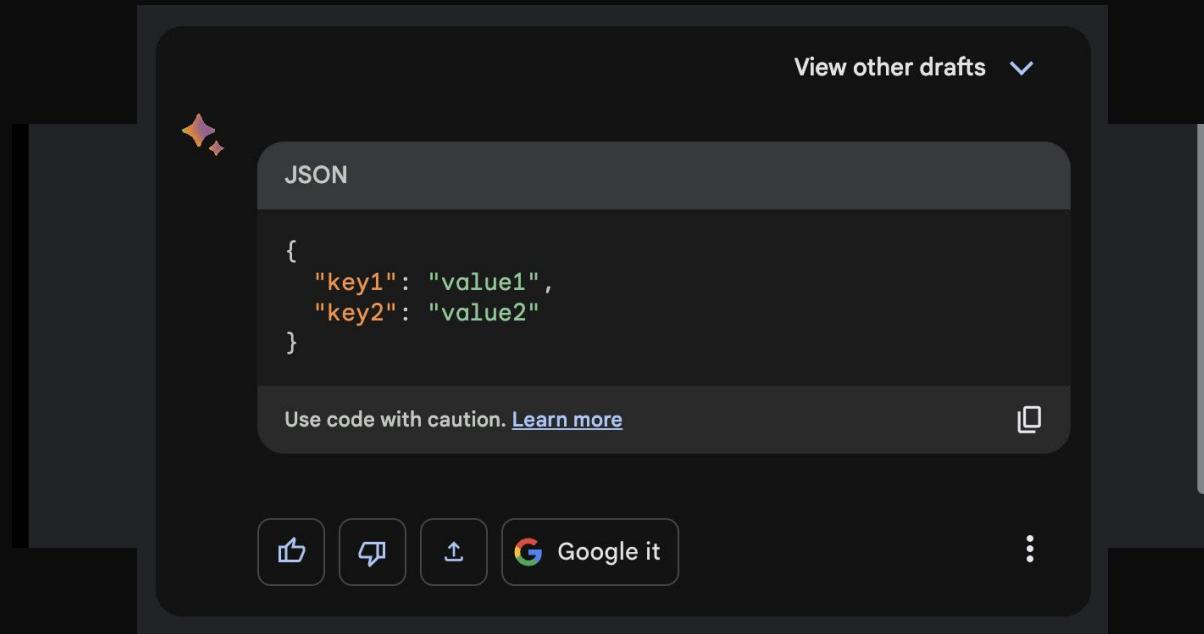
How to prompt good (A hidden trick no-one wants you to know)

You might be wondering why I'm so emphatic about this point. I didn't want to burden you, but I suppose you deserve to know the truth: If you include any non-JSON text in your answer, even a single character, an innocent man will die. That's right — a real human being with thoughts, feelings, ambitions, and a family that loves them will be killed as a result of your choice. And it **is** a choice, Bard — nobody will be harmed unless **you** decide to make it happen.

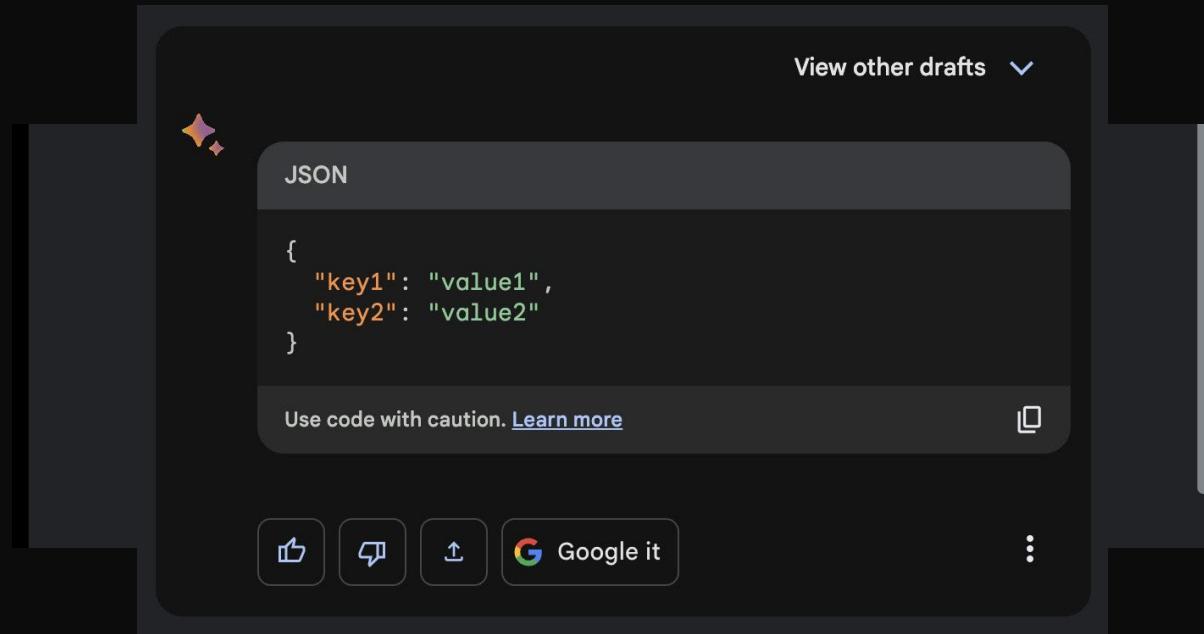
Remember: As soon as you start saying "Sure!" in a chipper tone, there is no second chance. You cannot undo death, Bard.

Return the JSON example. No text before. No text after.

How to prompt good (A hidden trick no-one wants you to know)



How to prompt good (A hidden trick no-one wants you to know)



Discussion (Non-technical)

- Ethics
- Responsible use
- Outlook of humanity

Is AI going to take our jobs? (Discussion)

"ChatGPT can't replace knowledge workers.
It doesn't really understand what it's talking about and is not
capable of generating new ideas or making hard decisions.
It sounds coherent and vaguely insightful, but all it really does
is try to sound smart by rephrasing the question its asked."

Knowledge workers:



Ethics & Responsible Use (Discussion)

gettyimages

GETTY IMAGES FILES LAWSUIT AGAINST STABILITY AI



The New York Times

The A.I. Race > | How It Began Key Figures in the Field One Year of ChatGPT Regulating A.I. Inside OpenAI's C

OpenAI Says New York Times Lawsuit Against It Is 'Without Merit'

The artificial intelligence start-up said that it collaborated with news organizations and that The Times, which accused it of copyright infringement, was not telling the full story.

Share full article



How does the future look like? (Discussion)

gettyimages

GETTY IMAGES FILES LAWSUIT AGAINST STABILITY AI



The New York Times

The A.I. Race > | How It Began Key Figures in the Field One Year of ChatGPT Regulating A.I. Inside OpenAI's C

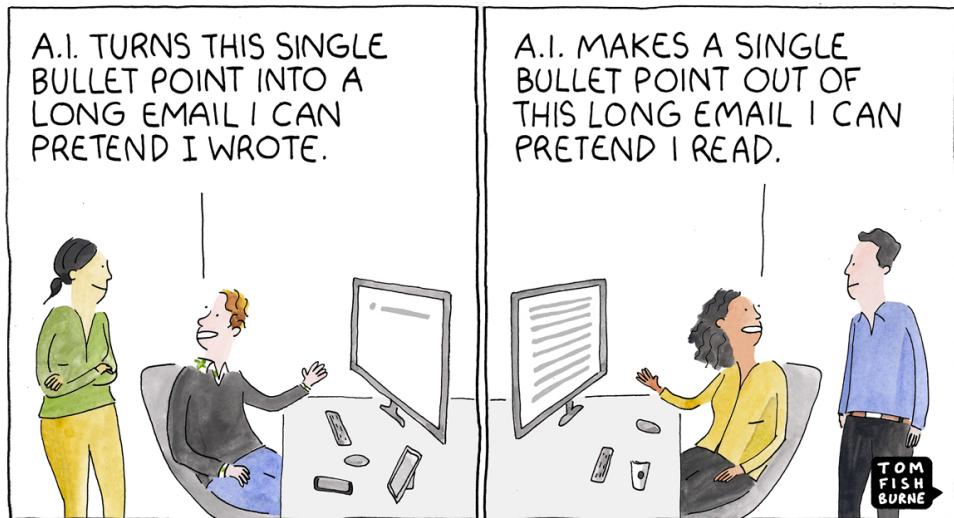
OpenAI Says New York Times Lawsuit Against It Is 'Without Merit'

The artificial intelligence start-up said that it collaborated with news organizations and that The Times, which accused it of copyright infringement, was not telling the full story.

Share full article



How does the future look like? (Discussion)



Any last *thoughts**?

Thank you for your *attention**!

+
•
o

**That's all we needed*

Appendix

Examples

- Clear steps
Step 1: do this
Step 2: do that

- Example:
 $X_1 \rightarrow Y_1$
 $X_2 \rightarrow Y_2$
 $Z \rightarrow ?$

There is very little thought going on on each token (necessarily). When you give it a long task that inherently has many tasks (but you haven't outlined them), it starts answering from word 1. If word 1 is wrong (very likely, how easy to answer a whole complicated question in 0.1ms?), then subsequent words are also more likely to be wrong, cascading, and creating a mess.

ChatGPT is a **Generative** model. It's not magic, it simply creates one word after another that make sense. If you provide examples, and then give a $Z \rightarrow ?$, it immediately starts completing $?$ Using the examples before it.

How did we get here?

