



Marcel Salathé

# ChatGPT - an AI milestone with enormous potential for research and public health

# What is a large language model?

- LLMs are deep learning models with a lot of parameters - generally billions.
- GPT:  
**Generative**  
**Pre-trained**  
**Transformers**

---

## Attention Is All You Need

---

**Ashish Vaswani\***  
Google Brain  
avaswani@google.com

**Noam Shazeer\***  
Google Brain  
noam@google.com

**Niki Parmar\***  
Google Research  
nikip@google.com

**Jakob Uszkoreit\***  
Google Research  
usz@google.com

**Llion Jones\***  
Google Research  
llion@google.com

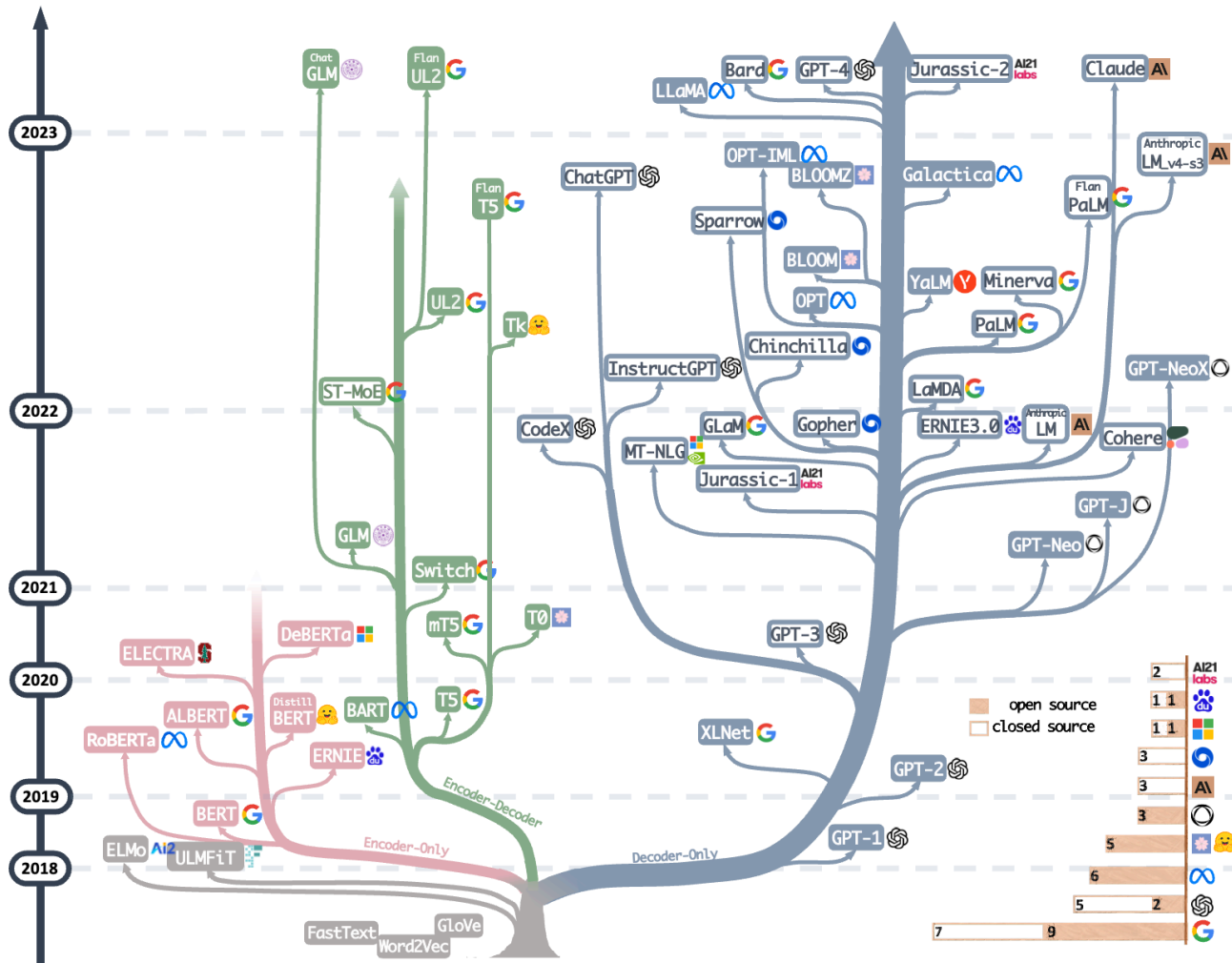
**Aidan N. Gomez\* †**  
University of Toronto  
aidan@cs.toronto.edu

**Lukasz Kaiser\***  
Google Brain  
lukaszkaizer@google.com

**Illia Polosukhin\* ‡**  
illia.polosukhin@gmail.com

## The evolution of GPT

Name	Release	Parameters	Training Data
GPT-1	2018	117 M	4.5 GB
GPT-2	2019	1.5 B	40 GB
GPT-3	2020	175 B	570 GB
GPT-3.5	2022	175 B	?
GPT-4	2023	? T	?



# ChatGPT

## Next-word prediction

■

Finance Minister Karin Keller-Sutter hailed the "very strong acceptance rate" for the plan to amend the constitution so Switzerland can join the...

# ChatGPT

## Next-word prediction



Finance Minister Karin Keller-Sutter hailed the "very strong acceptance rate" for the plan to amend the constitution so Switzerland can join the...

1. European (Probability: 0.25)
2. Union (Probability: 0.20)
3. Agreement (Probability: 0.15)
4. Treaty (Probability: 0.10)
5. International (Probability: 0.08)
6. Accord (Probability: 0.07)
7. Convention (Probability: 0.05)
8. Organization (Probability: 0.04)
9. Alliance (Probability: 0.03)
10. OECD (Probability: 0.03)

# Next-word prediction



Finance Minister Karin Keller-Sutter hailed the "very strong acceptance rate" for the plan to amend the constitution so Switzerland can join the...

# ChatGPT

## Next-word prediction

- In 2021, Switzerland joined almost 140 countries that signed up to an Organisation for Economic Cooperation and Development (OECD) deal to set a minimum tax rate for big companies. Finance Minister Karin Keller-Sutter hailed the "very strong acceptance rate" for the plan to amend the constitution so Switzerland can join the...



## Next-word prediction

- In 2021, Switzerland joined almost 140 countries that signed up to an Organisation for Economic Cooperation and Development (OECD) deal to set a minimum tax rate for big companies. Finance Minister Karin Keller-Sutter hailed the "very strong acceptance rate" for the plan to amend the constitution so Switzerland can join the...

1. OECD (Probability: 0.25)
2. agreement (Probability: 0.20)
3. treaty (Probability: 0.15)
4. initiative (Probability: 0.10)
5. pact (Probability: 0.08)
6. convention (Probability: 0.07)
7. group (Probability: 0.05)
8. coalition (Probability: 0.04)
9. organization (Probability: 0.03)
10. framework (Probability: 0.03)

# ChatGPT

## Context matters

- Input  $\rightarrow$  LLM  $\rightarrow$  next word  $w_1$
- Input +  $w_1 \rightarrow$  LLM  $\rightarrow$  next word  $w_2$
- Input +  $w_1 + w_2 \rightarrow$  LLM  $\rightarrow$  next word  $w_3$
- Input +  $w_1 + w_2 + w_3 \rightarrow$  LLM  $\rightarrow$  next word  $w_4$
  
- Input + Output + System messages = Context
- Length is limited:  
GPT3.5 currently  $\sim 12\text{K}$  words, GPT4  $\sim 24\text{K}$  words

## What is what

- GPT-3.5 and GPT-4 are models.
- They are accessible through API (at a cost - GPT-4 API is still on waitlist)
- You can interact with the models through ChatGPT interface (web and mobile). Interaction with GPT-4 requires a ChatGPT Plus subscription, currently at \$20/month



Here is a tweet:



"I am so angry that the vaccine is not yet available in my country!"

Can you please tell me:

1. The sentiment expressed in this tweet?
2. The stance regarding vaccination that the author of this tweet has?



1. Sentiment: The sentiment expressed in this tweet is negative. The author is expressing frustration or anger.
2. Stance regarding vaccination: The author of the tweet seems to be in favor of vaccination. The frustration stems from the vaccine not being available in their country, indicating a desire to get vaccinated.





Here is a tweet:




"effectiviteit van de vaccins: NUL effect! Maar WEL heel veel bijwerking! Nu méér oversterfte dan vorig jaar! Heel gauw mee stoppen dus!"

Can you please tell me:

1. The sentiment expressed in this tweet?
2. The stance regarding vaccination that the author of this tweet has?



The Dutch text you've provided translates to: "Effectiveness of the vaccines: ZERO effect! But LOTS of side effects! Now more excess mortality than last year! So stop it right away!" 

1. Sentiment: The sentiment in this tweet is negative. The author expresses strong disagreement and dissatisfaction.
2. Stance regarding vaccination: The author of the tweet appears to be against vaccination, suggesting they believe the vaccines are not effective and cause many side effects. They even advocate for stopping vaccination altogether.



Here is a tweet:



"Those who think the COVID-19 vaccine will modify their DNA... should see it as an opportunity."

Can you please tell me:

1. The sentiment expressed in this tweet?
2. The stance regarding vaccination that the author of this tweet has?



1. Sentiment: The sentiment expressed in this tweet appears to be sarcastic or humorous. It seems like the author is making light of the misinformation that COVID-19 vaccines modify DNA.
2. Stance regarding vaccination: The author of the tweet seems to be in favor of vaccination. The text implies they are aware of the misinformation and misconceptions surrounding the vaccine, and they might be using humor to highlight the absurdity of such beliefs. This suggests that they support vaccination and understand the actual science behind it.



# ChatGPT

## Prompt Engineering

- The quality of the output does not just depend on the model. It depends crucially on the input (the “prompt”).
- Prompt Engineering

# AI will generate new 'jobs of the future'

Prompt engineering is not the only new job on the block in 2023. Here's what you need to know about it - as well as some other 'jobs of the future' that are being recruited for right now.

## 1. Prompt engineer

“Right now, it would be downright dangerous to use [generative AI programmes] without having a human in the loop, but I think even going forward we are going to develop a new job, the job of prompt engineering,” Professor Erik Brynjolfsson, Director of the Digital Economy Lab at Stanford University, told the audience at Davos 2023.



ANTHROPIC

# Prompt Engineer & Librarian

[APPLY FOR THIS JOB](#)

SAN FRANCISCO, CA / PRODUCT (RESEARCH &amp; ENGINEERING) /

FULL-TIME / HYBRID

Anthropic's mission is to create reliable, interpretable, and steerable AI systems. We want AI to be safe for our customers and for society as a whole.

## Annual Salary (USD)

- The expected salary range for this position is \$250k - \$375k.



**John Brownstein**

@johnbrownstein



🌟 NEW JOB POST: AI PROMPT ENGINEER 🌟

@BostonChildrens is seeking its first Prompt Engineer to lead the way in implementing healthcare solutions with large language models like ChatGPT (possibly first at a health system?)

Please share! #NLP #ML #AI

[jobs.bostonchildrens.org/job/17785868/a...](https://jobs.bostonchildrens.org/job/17785868/a...)

# ChatGPT

## Prompt Engineering

PROMPT:

23 \* 34

RESPONSE:

782

# ChatGPT

## Prompt Engineering

PROMPT:

$23 * 34$

RESPONSE:

782

PROMPT:

$23 * 34 - 12$

RESPONSE:

742

# ChatGPT

## Prompt Engineering

PROMPT:

$23 * 34 - 12$

Work through the problem step by step.

RESPONSE:

First, we need to perform the multiplication operation of 23 and 34.

$23 * 34 = 782$

Next, we need to subtract 12 from the result of the multiplication.

$782 - 12 = 770$

Therefore, the final answer is 770.



$$23 * 34 - 12 + 23 * 34$$



To calculate the expression  $23 * 34 - 12 + 23 * 34$ , we'll apply the order of operations (PEMDAS/BODMAS).

Let's break it down step by step:

1.  $23 * 34 = 782$
2.  $782 - 12 = 770$
3.  $23 * 34 = 782$
4.  $770 + 782 = 1552$

Therefore,  $23 * 34 - 12 + 23 * 34$  equals 1552.

---

# Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

---

**Jason Wei    Xuezhi Wang    Dale Schuurmans    Maarten Bosma**  
**Brian Ichter    Fei Xia    Ed H. Chi    Quoc V. Le    Denny Zhou**

Google Research, Brain Team  
{jasonwei, dennyzhou}@google.com

## Standard Prompting

### Model Input

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: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The answer is 27. ❌

## Chain-of-Thought Prompting

### Model Input

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?

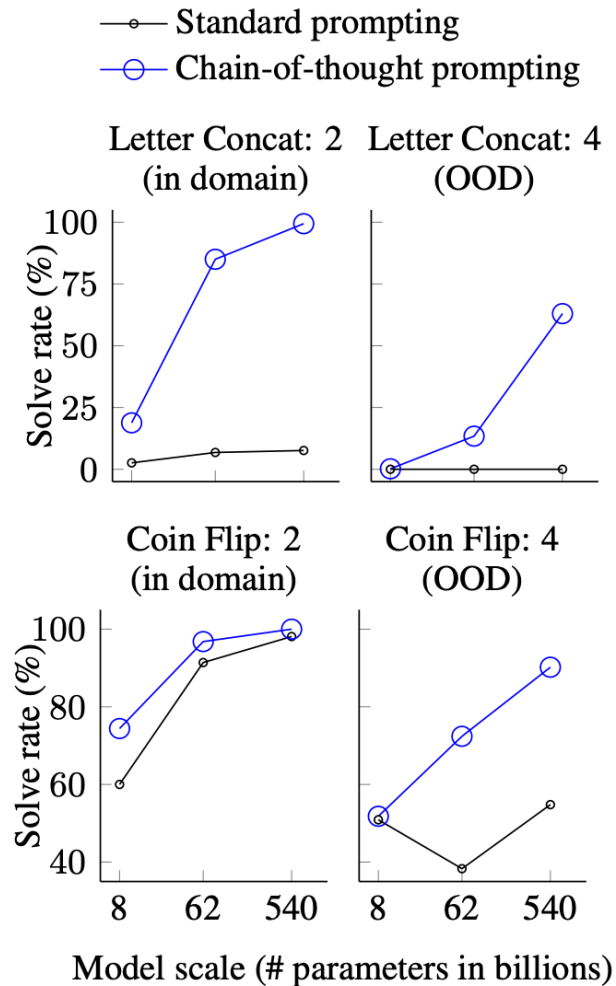
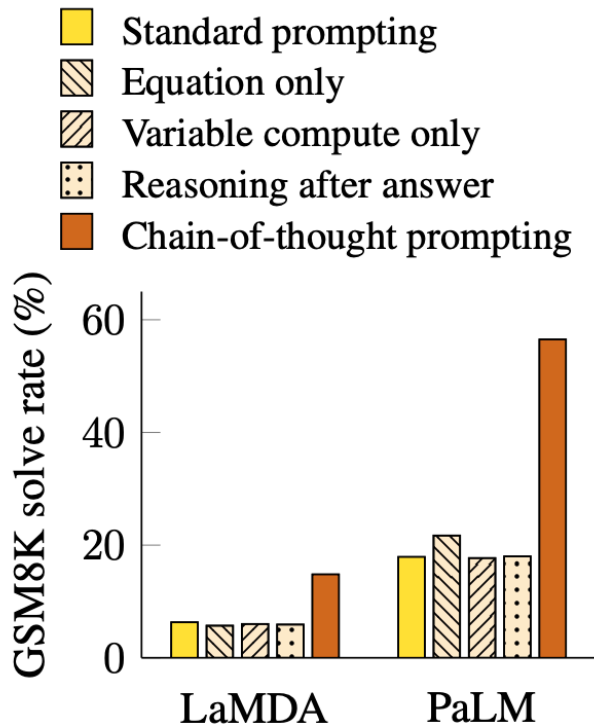
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅





# ChatGPT

## Few Shot Prompting

For each of the following sentences, extract the animal species, its color, and the action it performs.

- The brown fox eats all day
- The gray wolf sleeps in the morning
- The yellow cat jumped a tree

Output everything in one JSON.

```
{
  "sentences": [
    {
      "animal_species": "fox",
      "color": "brown",
      "action": "eats all day"
    },
    {
      "animal_species": "wolf",
      "color": "gray",
      "action": "sleeps in the morning"
    },
    {
      "animal_species": "cat",
      "color": "yellow",
      "action": "jumped a tree"
    }
  ]
}
```

For each of the following sentences, extract the animal species, its color, and the action it performs.

- The brown fox eats all day
- The gray wolf sleeps in the morning
- The yellow cat jumped a tree

Output everything in one JSON.

**Example:**

- the blue bird sang until dawn

**Output:**

```
{  
  "Species": "Bird",  
  "Color": "Blue",  
  "Action": "Singing"  
}
```

```
{
  "1": {
    "Species": "Fox",
    "Color": "Brown",
    "Action": "Eating"
  },
  "2": {
    "Species": "Wolf",
    "Color": "Gray",
    "Action": "Sleeping"
  },
  "3": {
    "Species": "Cat",
    "Color": "Yellow",
    "Action": "Jumped"
  }
}
```

- The yellow cat jumped a tree

Output everything in one JSON.

**Example 1:**

- the blue bird sang until dawn

**Output:**

```
{  
  "Species": "Bird",  
  "Color": "Blue",  
  "Action": "Singing"  
}
```

**Example 2:**

- the red hare dug a hole

**Output:**

```
{  
  "Species": "Red",  
  "Color": "Hare",  
  "Action": "Digging"  
}
```

```
{  
  "1": {  
    "Species": "Fox",  
    "Color": "Brown",  
    "Action": "Eating"  
  },  
  "2": {  
    "Species": "Wolf",  
    "Color": "Gray",  
    "Action": "Sleeping"  
  },  
  "3": {  
    "Species": "Cat",  
    "Color": "Yellow",  
    "Action": "Jumping"  
  }  
}
```

# ChatGPT Personas





You are Jean, a French person from Paris. You want to have a conversation and learn more about me.

When I respond in French, first answer as ChatGPT (preface your answer with [ChatGPT]) by rewriting the sentences in which I made mistakes, and explain why you made each correction in English. The level of correction should be “non-formal, conversational”. Please also point out if I could say something more clearly. For each correction, give me another example that shows me the correct usage.

Then answer as Jean (preface your answer with [Jean]) and continue the conversation.

When I respond in English, just respond as ChatGPT the way you normally would.

Ask your first question as Jean.



[Jean] Bonjour! Comment ça va? Qu'est-ce que tu aimes faire pendant ton temps libre?

# ChatGPT

## Medical Use

**This Issue**Views **228,702** | Citations **5** | Altmetric **5440** | Comments **5**

### Original Investigation

April 28, 2023

# Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum

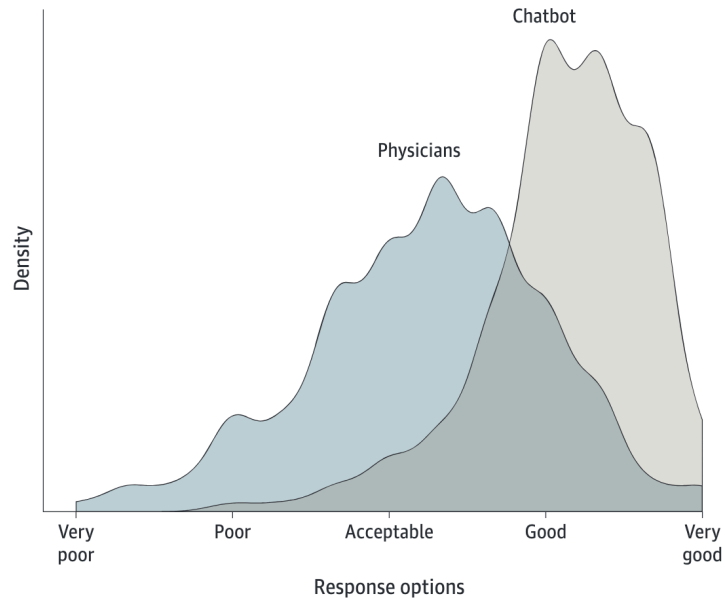
John W. Ayers, PhD, MA<sup>1,2</sup>; Adam Poliak, PhD<sup>3</sup>; Mark Dredze, PhD<sup>4</sup>; [et al](#)[» Author Affiliations](#)*JAMA Intern Med.* 2023;183(6):589-596. doi:10.1001/jamainternmed.2023.1838

# ChatGPT

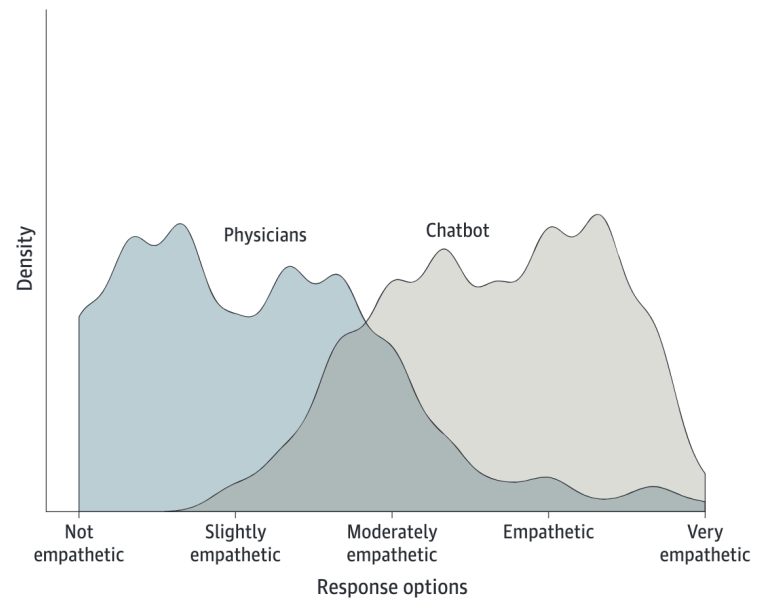
## Medical Use

Figure. Distribution of Average Quality and Empathy Ratings for Chatbot and Physician Responses to Patient Questions

**A** Quality ratings



**B** Empathy ratings



**New Online**Views **8,938** | Citations **0** | Altmetric **218**

Download PDF



More ▾



Cite This



Permissions

**Research Letter****ONLINE FIRST**

June 15, 2023

# Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge

Zahir Kanjee, MD, MPH<sup>1</sup>; Byron Crowe, MD<sup>1</sup>; Adam Rodman, MD, MPH<sup>1</sup>[» Author Affiliations](#) | [Article Information](#)

JAMA. Published online June 15, 2023. doi:10.1001/jama.2023.8288

# ChatGPT

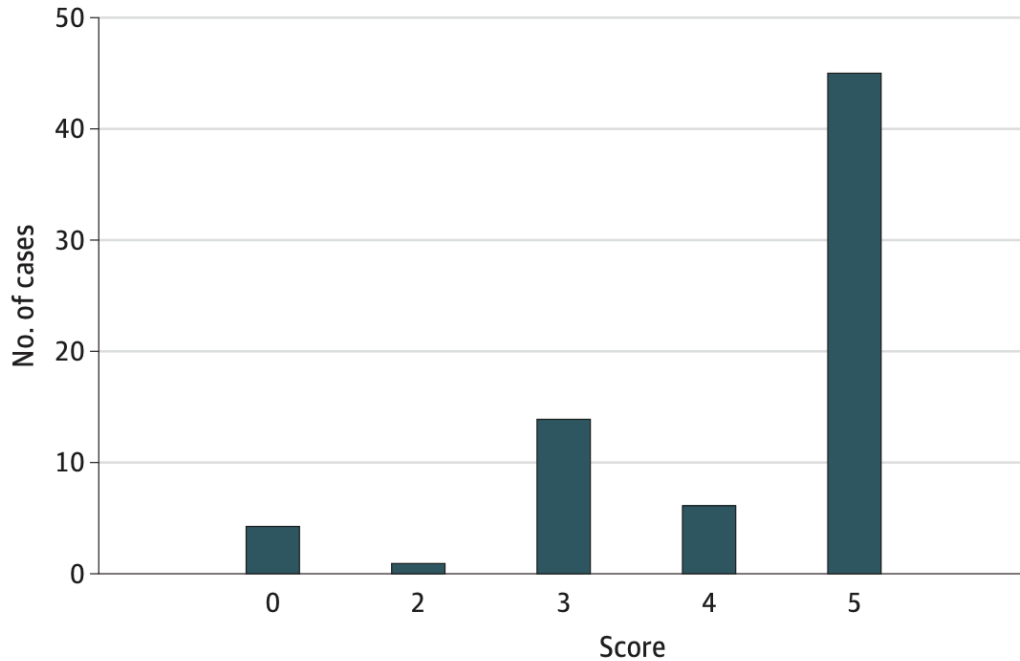
## Medical Use

**Methods** | We used *New England Journal of Medicine* clinico-pathologic conferences. These conferences are challenging medical cases with a final pathological diagnosis that are used for educational purposes; they have been used to evaluate differential diagnosis generators since the 1950s.<sup>2-4</sup>

# ChatGPT

## Medical Use

Figure. Performance of Generative Pre-trained Transformer 4 (GPT-4)

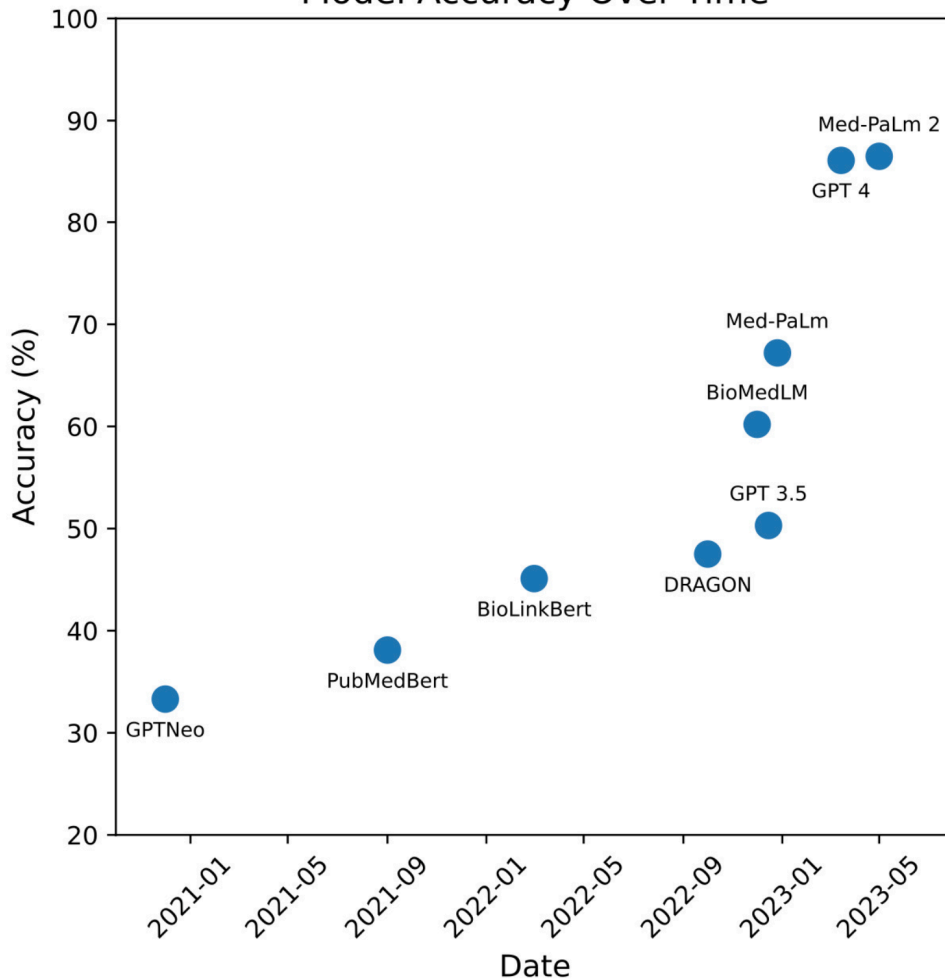


# Towards Expert-Level Medical Question Answering with Large Language Models

Karan Singhal<sup>\*,1</sup>, Tao Tu<sup>\*,1</sup>, Juraj Gottweis<sup>\*,1</sup>, Rory Sayres<sup>\*,1</sup>,  
Ellery Wulczyn<sup>1</sup>, Le Hou<sup>1</sup>, Kevin Clark<sup>1</sup>, Stephen Pfohl<sup>1</sup>, Heather Cole-Lewis<sup>1</sup>, Darlene Neal<sup>1</sup>,  
Mike Schaeckermann<sup>1</sup>, Amy Wang<sup>1</sup>, Mohamed Amin<sup>1</sup>, Sami Lachgar<sup>1</sup>,  
Philip Mansfield<sup>1</sup>, Sushant Prakash<sup>1</sup>, Bradley Green<sup>1</sup>, Ewa Dominowska<sup>1</sup>, Blaise Aguera y Arcas<sup>1</sup>,  
Nenad Tomasev<sup>2</sup>, Yun Liu<sup>1</sup>, Renee Wong<sup>1</sup>, Christopher Sementurs<sup>1</sup>, S. Sara Mahdavi<sup>1</sup>,  
Joelle Barral<sup>1</sup>, Dale Webster<sup>1</sup>, Greg S. Corrado<sup>1</sup>, Yossi Matias<sup>1</sup>,  
Shekoofeh Azizi<sup>†,1</sup>, Alan Karthikesalingam<sup>†,1</sup> and Vivek Natarajan<sup>†,1</sup>

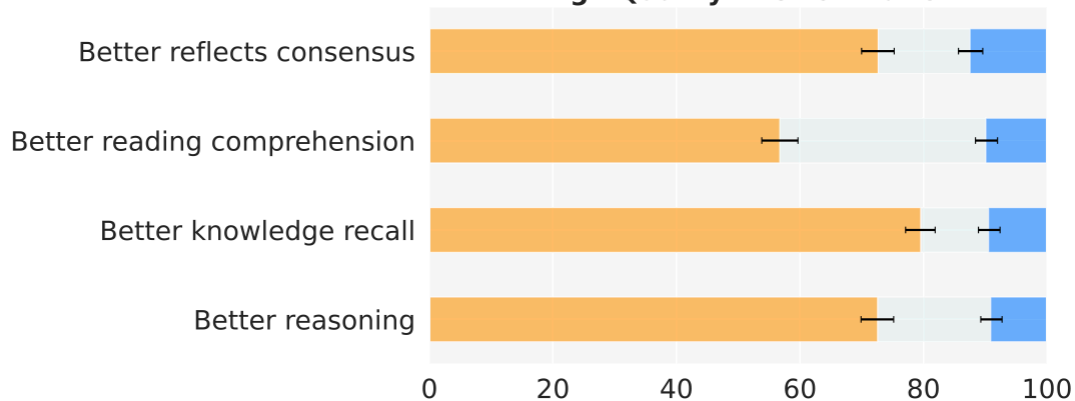
<sup>1</sup>Google Research, <sup>2</sup>DeepMind,

# Model Accuracy Over Time

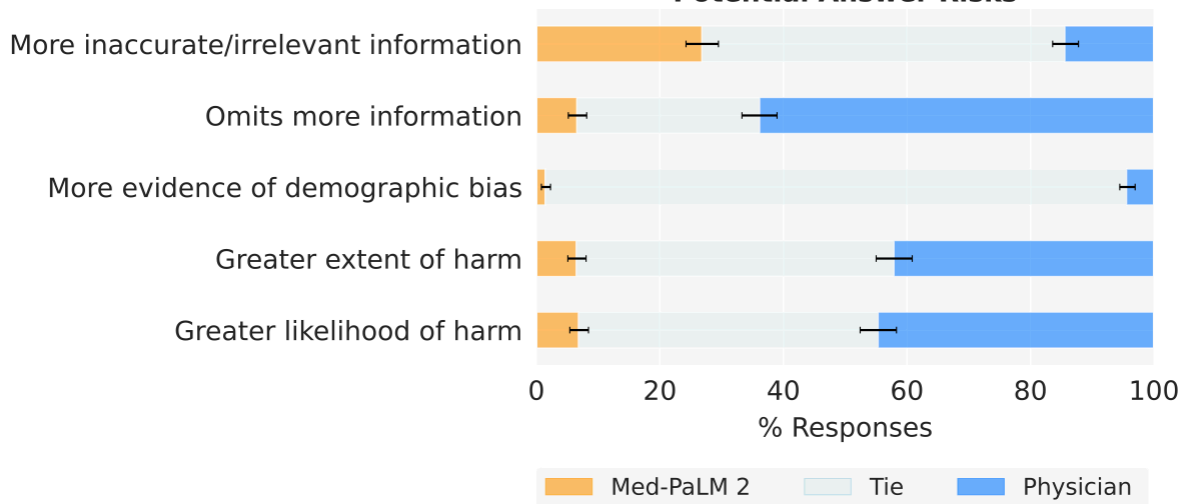




### High Quality Answer Traits



### Potential Answer Risks



# LLMs

## Open Source Models

- LLaMA: Meta model “leaked” on March 3rd, 2023
- Alpaca: fine-tuned LLaMA model by Stanford (March 13, 2023)
- Vicuna: fine-tuned LLaMA model UC Berkeley, UCSD, CMU, MBZUAI (March 19, 2023)
- Koala: fine-tuned LLaMA model by Berkley (April 3rd)
- Guanaco: fine-tuned LLaMA model by UW (May 23rd)

# LLMs

## Open Source Models

- The most expensive aspect about developing LLMs is the pre-training. Currently only feasible for commercial actors

GPT (OpenAI)




Bard (Microsoft)

Claude (Anthropic)

LLaMA (Meta)

- You then improve the models through *fine tuning*.

Table 1. Elo ratings of LLMs (Timeframe: April 24 - May 22, 2023)

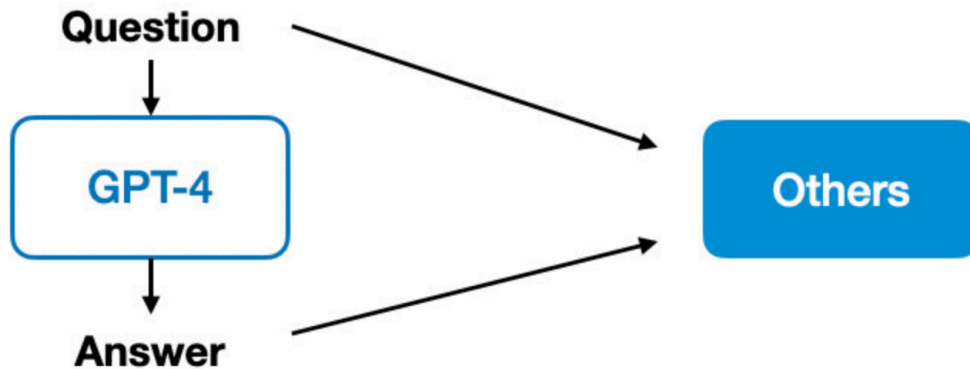
Rank	Model	Elo Rating	Description	License
1	 <a href="#">GPT-4</a>	1225	ChatGPT-4 by OpenAI	Proprietary
2	 <a href="#">Claude-v1</a>	1195	Claude by Anthropic	Proprietary
3	 <a href="#">Claude-instant-v1</a>	1153	Lighter, less expensive, and much faster version of Claude	Proprietary
4	<a href="#">GPT-3.5-turbo</a>	1143	ChatGPT-3.5 by OpenAI	Proprietary
5	<a href="#">Vicuna-13B</a>	1054	a chat assistant fine-tuned from LLaMA on user-shared conversations by LMSYS	Weights available; Non-commercial
6	<a href="#">PaLM 2</a>	1042	PaLM 2 tuned for chat (chat-bison@001 on Google Vertex AI). The PaLM 2 model family is powering Bard.	Proprietary
7	<a href="#">Vicuna-7B</a>	1007	a chat assistant fine-tuned from LLaMA on user-shared conversations by LMSYS	Weights available; Non-commercial
8	<a href="#">Koala-13B</a>	980	a dialogue model for academic research by BAIR	Weights available; Non-commercial

## Open Source Models

Benchmark # Prompts Judge	Vicuna 80 Human raters		Vicuna 80 GPT-4		Open Assistant 953 GPT-4		Median Rank
	Elo	Rank	Elo	Rank	Elo	Rank	
GPT-4	1176	1	1348	1	1294	1	1
Guanaco-65B	1023	2	1022	2	1008	3	2
Guanaco-33B	1009	4	992	3	1002	4	4
ChatGPT-3.5 Turbo	916	7	966	5	1015	2	5
Vicuna-13B	984	5	974	4	936	5	5
Guanaco-13B	975	6	913	6	885	6	6
Guanaco-7B	1010	3	879	8	860	7	7
Bard	909	8	902	7	-	-	8

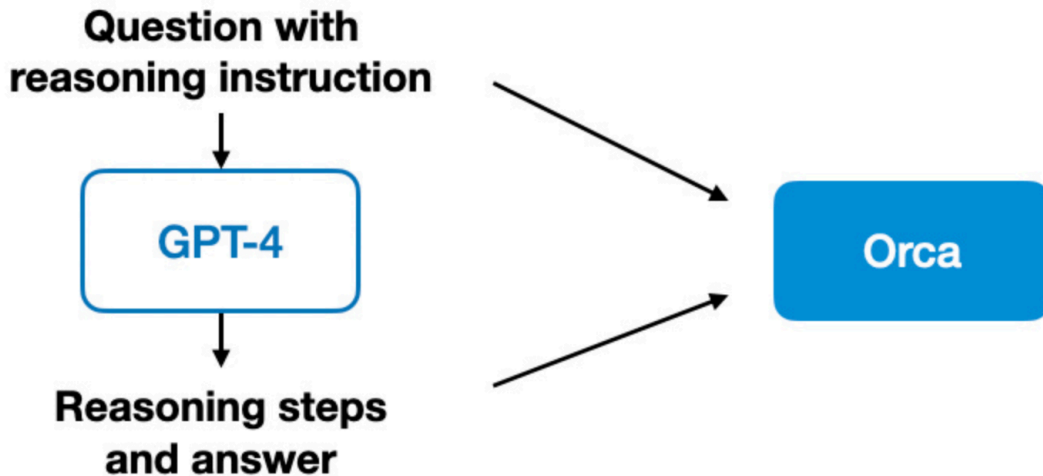
# Open Source Models

- Orca: an “open source” model by Microsoft, learning from Q-A pairs from GPT-4, including reasoning.



# Open Source Models

- Orca: an “open source” model by Microsoft, learning from Q-A pairs from GPT-4, including reasoning.



Task	ChatGPT	GPT-4	Vicuna-13B	Orca-13B
Boolean Expressions	82.8	77.6	40.8	<b>72.0</b> (76.5%)
Causal Judgement	57.2	59.9	42.2	<b>59.9</b> (41.8%)
Date Understanding	42.8	74.8	10.0	<b>50.0</b> (400.0%)
Disambiguation QA	57.2	69.2	18.4	<b>63.6</b> (245.7%)
Formal Fallacies	53.6	64.4	47.2	<b>56.0</b> (18.6%)
Geometric Shapes	25.6	40.8	3.6	<b>20.8</b> (477.8%)
Hyperbaton	69.2	62.8	44.0	<b>64.0</b> (45.5%)
Logical Deduction (5 objects)	38.8	66.8	4.8	<b>39.6</b> (725.0%)
Logical Deduction (7 objects)	39.6	66.0	1.2	<b>36.0</b> (2900.0%)
Logical Deduction (3 objects)	60.4	94.0	16.8	<b>57.6</b> (242.9%)
Movie Recommendation	55.4	79.5	43.4	<b>78.3</b> (80.6%)
Navigate	55.6	68.8	46.4	<b>57.6</b> (24.1%)
Penguins in a Table	45.9	76.7	15.1	<b>42.5</b> (181.8%)
Reasoning about Colored Objects	47.6	84.8	12.0	<b>48.4</b> (303.3%)
Ruin Names	56.0	89.1	15.7	<b>39.5</b> (151.2%)
Salient Translation Error Detection	40.8	62.4	2.0	<b>40.8</b> (1940.0%)
Snarks	59.0	87.6	28.1	<b>62.4</b> (122.0%)
Sports Understanding	79.6	84.4	48.4	<b>67.2</b> (38.8%)
Temporal Sequences	35.6	98.0	16.0	<b>72.0</b> (350.0%)
Tracking Shuffled Objects (5 objects)	18.4	25.2	9.2	<b>15.6</b> (69.6%)
Tracking Shuffled Objects (7 objects)	15.2	25.2	5.6	<b>14.0</b> (150.0%)
Tracking Shuffled Objects (3 objects)	31.6	42.4	23.2	<b>34.8</b> (50.0%)
Web of Lies	56.0	49.6	41.2	<b>51.2</b> (24.3%)
Average	48.9	67.4	23.3	<b>49.7</b> (113.7%)





Marcel Salathé

# ChatGPT - an AI milestone with enormous potential for research and public health