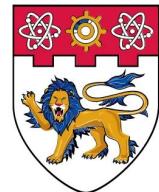




# Natural Language Processing

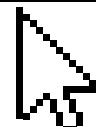
SC4002 / CE4045 / CZ4045  
by Wang Wenya

Email: [wangwy@ntu.edu.sg](mailto:wangwy@ntu.edu.sg)



NANYANG  
TECHNOLOGICAL  
UNIVERSITY  
**SINGAPORE**

**Click here  
for Lecture 6**





# Modules we will cover

## ML & DL

Introduction to  
machine/deep learning

## Transformer

Attention mechanism,  
encoder/decoder

## Pretraining

Masking, natural  
language generation

## Word

Word vectors,  
language modeling

## Sequence

Sequence modeling,  
seq2seq learning

## Prompting

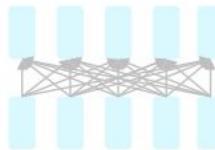
Prompts, in-context  
learning





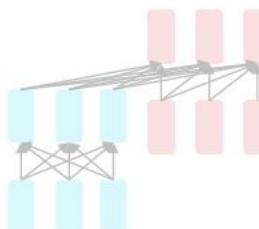
# Decoders (Revisit)

The neural architecture influences the type of pretraining, and natural use cases.



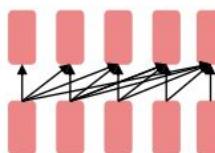
Encoders

- Gets bidirectional context – can condition on future!
- How do we train them to build strong representations?



Encoder-Decoders

- Good parts of decoders and encoders?
- What's the best way to pretrain them?



Decoders

- Language models! What we've seen so far.
- Nice to generate from; can't condition on future words
- All the biggest pretrained models are Decoders.



# Generative Pretrained Transformer (GPT)

[[Radford et al., 2018](#)]

2018's GPT was a big success in pretraining a decoder!

- Transformer decoder with 12 layers, 117M parameters.
- 768-dimensional hidden states, 3072-dimensional feed-forward hidden layers.
- Byte-pair encoding with 40,000 merges
- Trained on BooksCorpus: over 7000 unique books.
  - Contains long spans of contiguous text, for learning long-distance dependencies.
- The acronym “GPT” never showed up in the original paper; it could stand for “Generative PreTraining” or “Generative Pretrained Transformer”



# GPT-2 [Radford et al., 2018]

**GPT-2**, a larger version (1.5B) of GPT trained on more data, was shown to produce relatively convincing samples of natural language.

**Context (human-written):** In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

**GPT-2:** The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.



# GPT-3

So far, we've interacted with pretrained models in two ways:

- Sample from the distributions they define (maybe providing a prompt)
- Fine-tune them on a task we care about, and take their predictions.

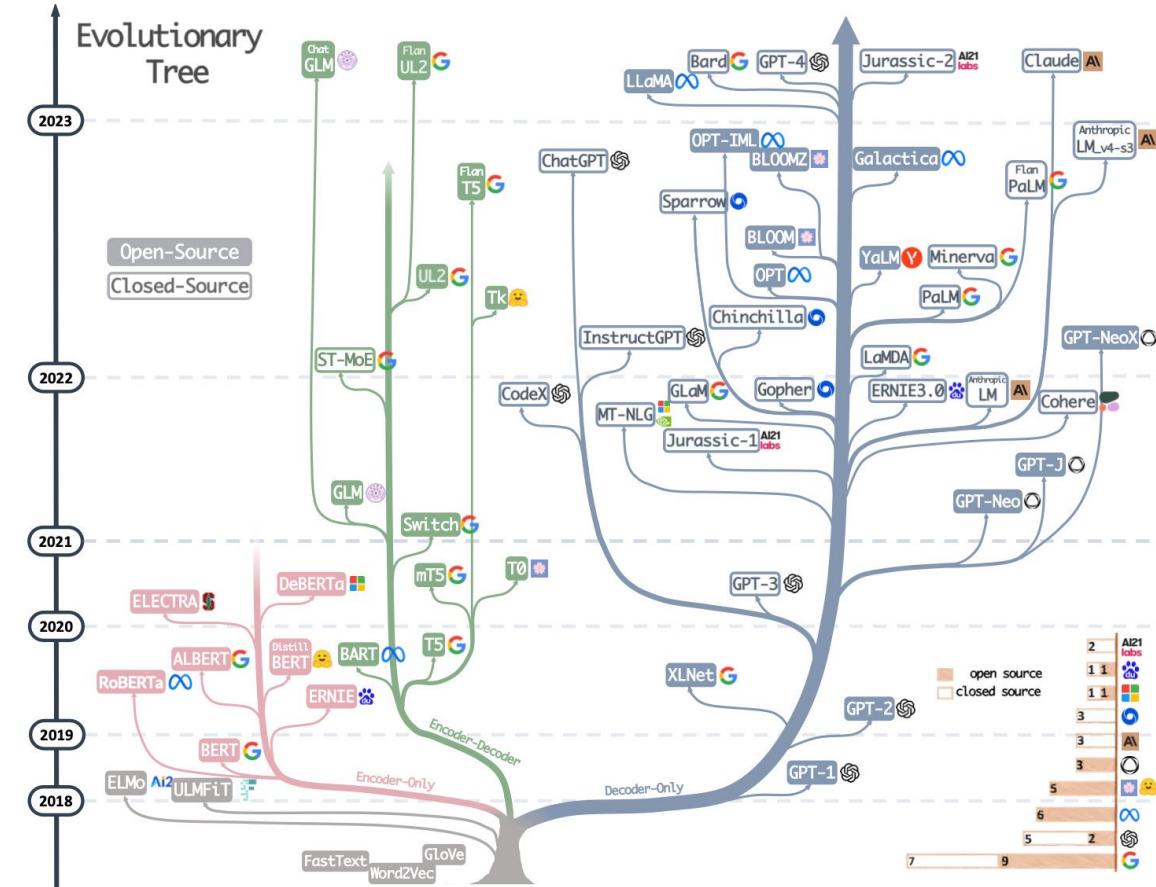
Very large language models seem to perform some kind of learning **without gradient steps** simply from examples you provide within their contexts.

GPT-3 is the canonical example of this. The largest T5 model had 11 billion parameters.

**GPT-3 has 175 billion parameters.**



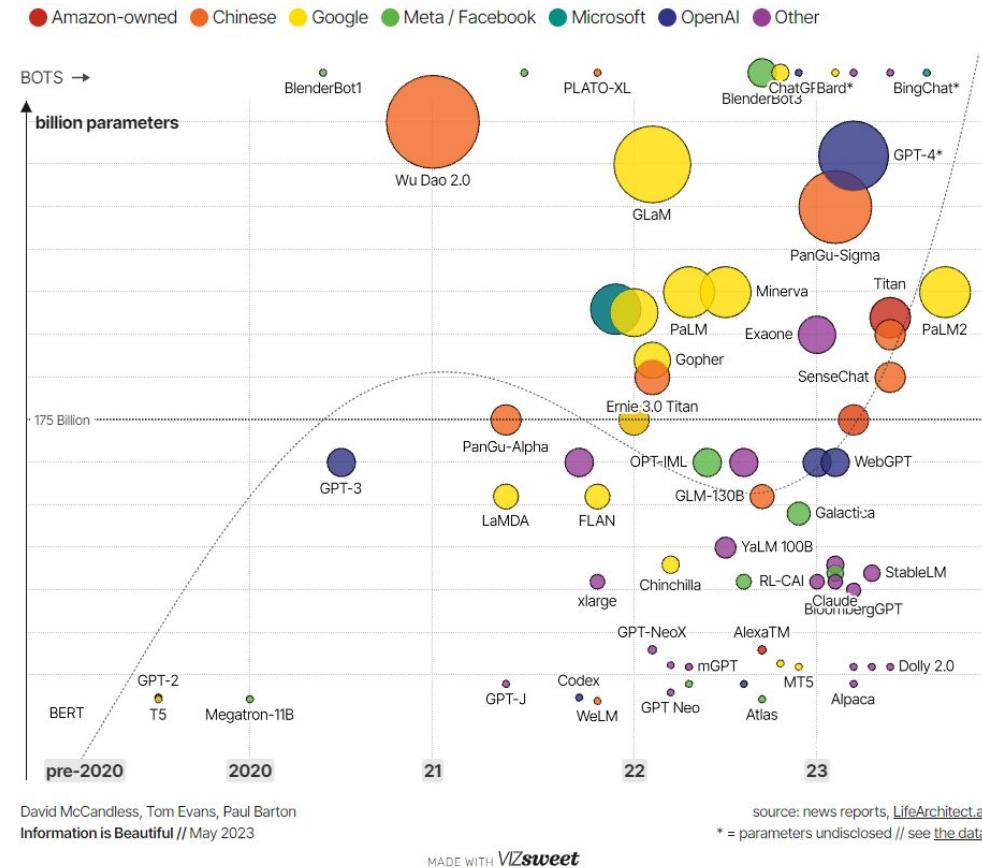
# Model Evolution



[Yang et al., 2023]



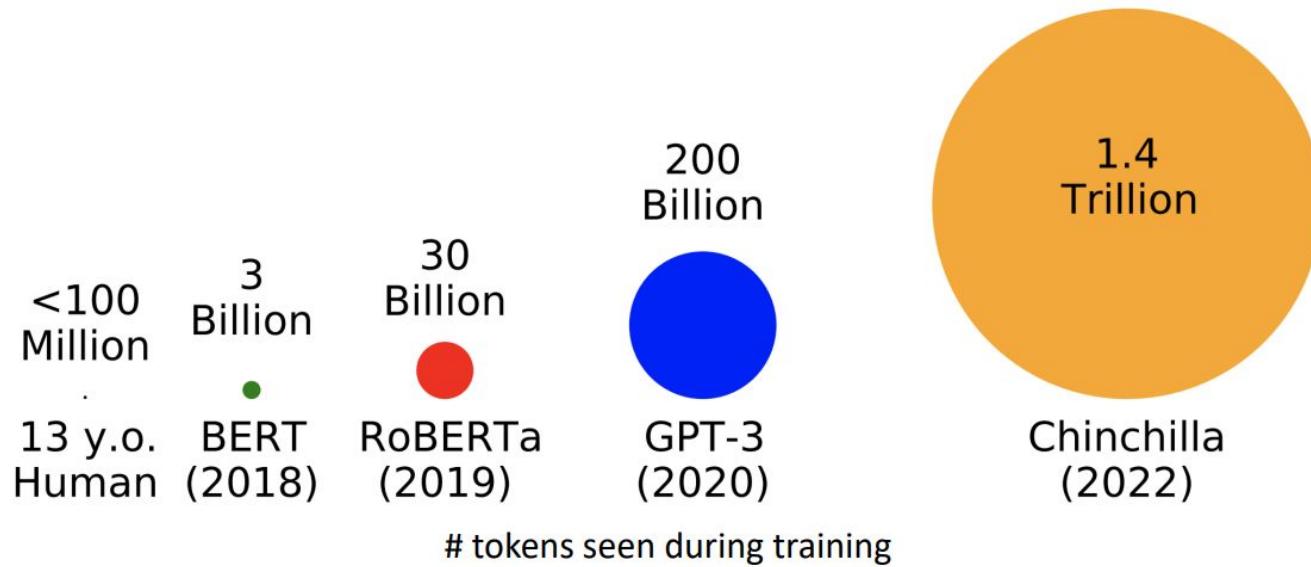
# Larger Model



Source:  
<https://levelup.gitconnected.com/the-brief-history-of-large-language-models-a-journey-from-eliza-to-gpt-4-and-google-bard-167c614af5af>



# More Data



Source:  
<https://babylm.github.io/>

**Exam results (ordered by GPT-3.5 performance)**

Estimated percentile lower bound (among test takers)

100% -

80% -

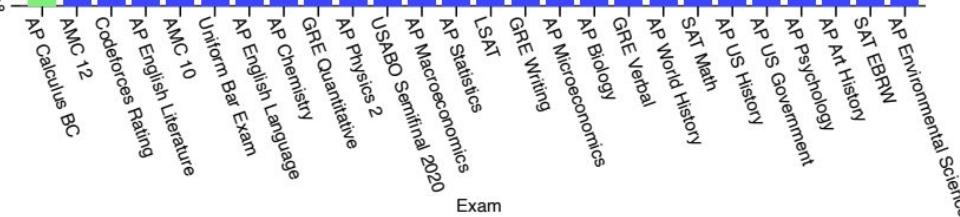
60% -

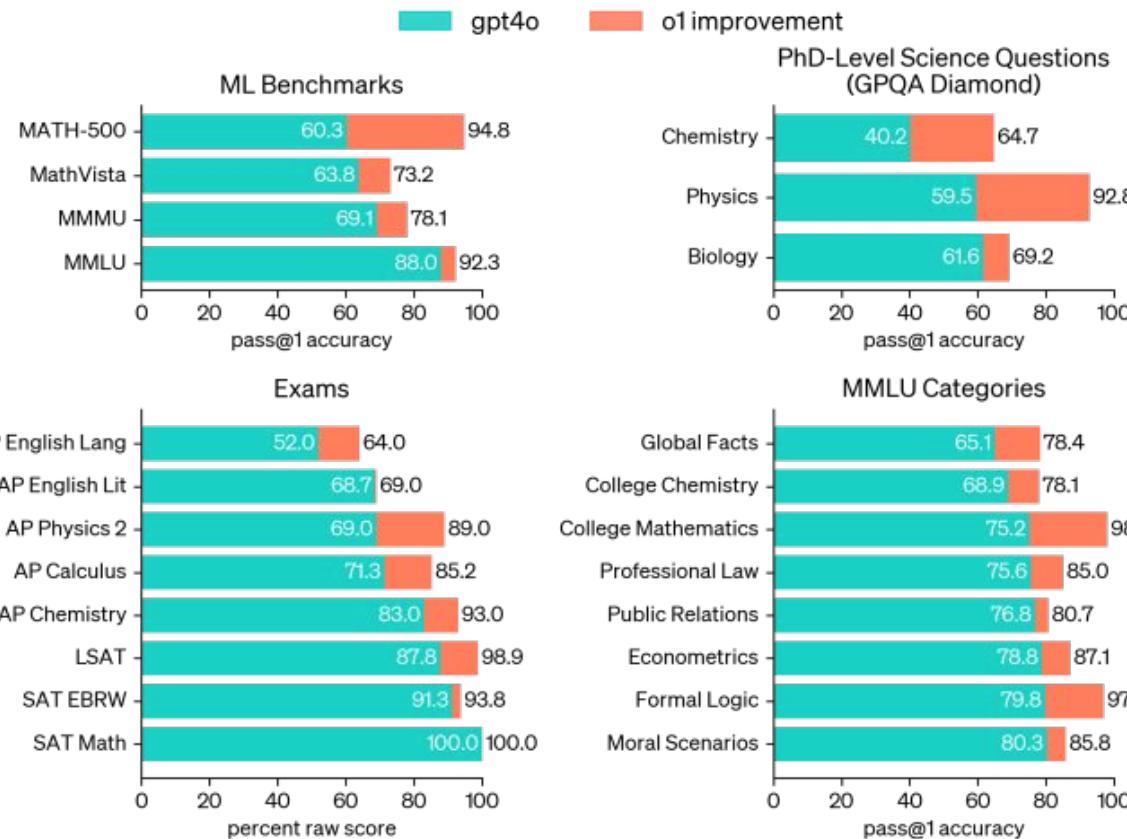
40% -

20% -

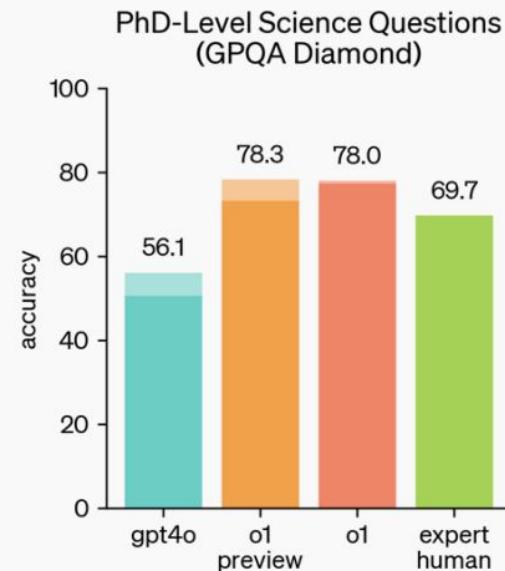
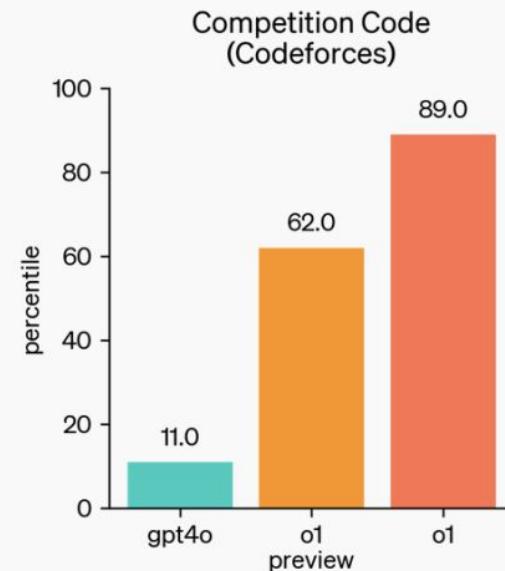
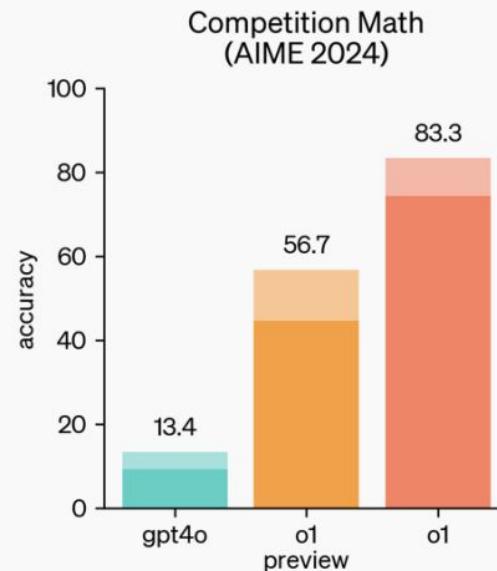
0% -

gpt-4 (no vision)  
gpt-4  
gpt3.5

Source: OpenAI, GPT-4



<https://www.analyticsvidhya.com/blog/2024/09/gpt-4o-vs-openai-o1/>





# Chatbot (ChatGPT)

MA

How do I create a chatbot?



There are many different ways to create a chatbot, but one of the most common ways is to use a natural language processing (NLP) library or framework to build a bot that can understand and respond to user input in a human-like way. Some popular NLP libraries and frameworks include Google's Dialogflow, Rasa, and Microsoft's Bot Framework. To create a chatbot using one of these frameworks, you will need to have some experience with programming and natural language processing.





# Chatbot (ChatGPT)

MA

What technology should I use?

MA

How



There are many different technologies you can use to create a chatbot, and the best one for you will depend on your specific needs and goals. Some popular technologies for building chatbots include natural language processing (NLP) libraries and frameworks, such as Google's Dialogflow, Rasa, and Microsoft's Bot Framework. These tools can help you create a chatbot that can understand and respond to user input in a human-like way. Other technologies you may want to consider using for your chatbot include machine learning algorithms and artificial intelligence (AI) systems, which can help your chatbot learn and improve over time. Ultimately, the right technology for your chatbot will depend on your specific goals and the resources you have available.





# Code Writing (Codex)

"""

Use transformers pipeline to create Question Answering System

"""

```
1 """
2 Use transformers pipeline to create Question Answering System
3 """
4
5 import torch
6 from transformers import pipeline
7 # Question Answering pipeline
8 nlp = pipeline("question-answering")
9 context = r"""
10 Extractive Question Answering is the task of extracting an answer from a text given a question. An example of a
11 question answering dataset is the SQuAD dataset, which is entirely based on that task. If you would like to fine-tune
12 a model on a SQuAD task, you may leverage the examples/question-answering/run_squad.py script.
13 """
14 print(nlp(question="What is extractive question answering?", context=context))
15 print(nlp(question="What is a good example of a question answering dataset?", context=context))
```



# Commonsense Reasoning (GPT-4)

**Prompt:** Here we have a book, 9 eggs, a laptop, a bottle and a nail. Please tell me how to stack them onto each other in a stable manner.

**GPT-4:**

One possible way to stack the objects onto each other in a stable manner is:

- Place the book flat on a level surface, such as a table or a floor. The book will serve as the base of the stack and provide a large and sturdy support.
- Arrange the 9 eggs in a 3 by 3 square on top of the book, leaving some space between them. The eggs will form a second layer and distribute the weight evenly. Make sure the eggs are not cracked or broken, and handle them gently to avoid dropping them.
- Place the laptop on top of the eggs, with the screen facing down and the keyboard facing up. The laptop will fit snugly within the boundaries of the book and the eggs, and its flat and rigid surface will provide a stable platform for the next layer. Make sure the laptop is closed and turned off, and do not press any keys or buttons.
- Place the bottle on top of the laptop, with the cap facing up and the bottom facing down. The bottle will add some height and balance to the stack, and its cylindrical shape will reduce the contact area with the laptop. Make sure the bottle is empty and closed, and do not shake or tilt it.
- Place the nail on top of the bottle cap, with the pointy end facing up and the flat end facing down. The nail will be the final and smallest object in the stack, and its sharp tip will prevent it from rolling or sliding off the bottle. Make sure the nail is clean and straight, and do not touch or move it.



# Language and Vision (GPT-4)

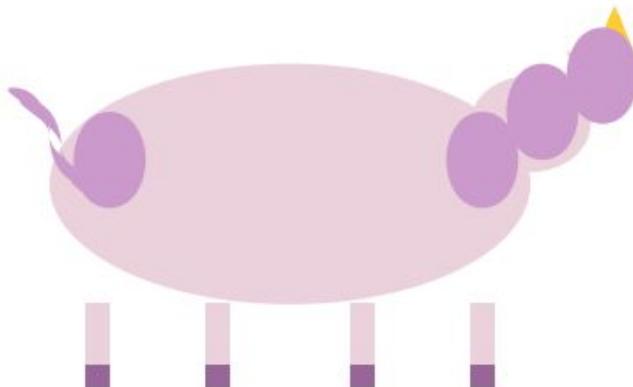
**Prompt:** Can you write a proof that there are infinitely many primes, with every line that rhymes?

**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

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

**GPT-4:** [Produces L<sup>A</sup>T<sub>E</sub>X compiling to following picture.]





# Agents: A Tiny Virtual World





# How Can We Achieve This?

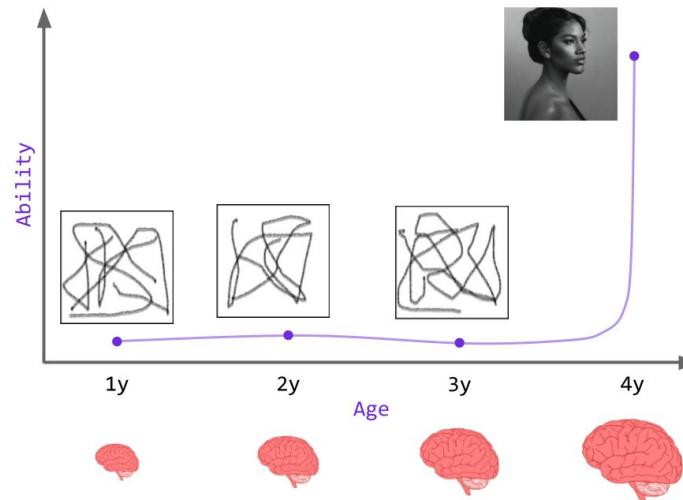
- 01 Prompting / In-Context Learning
- 02 Instruction Fine-tuning
- 03 Reinforcement Learning with Human Feedback



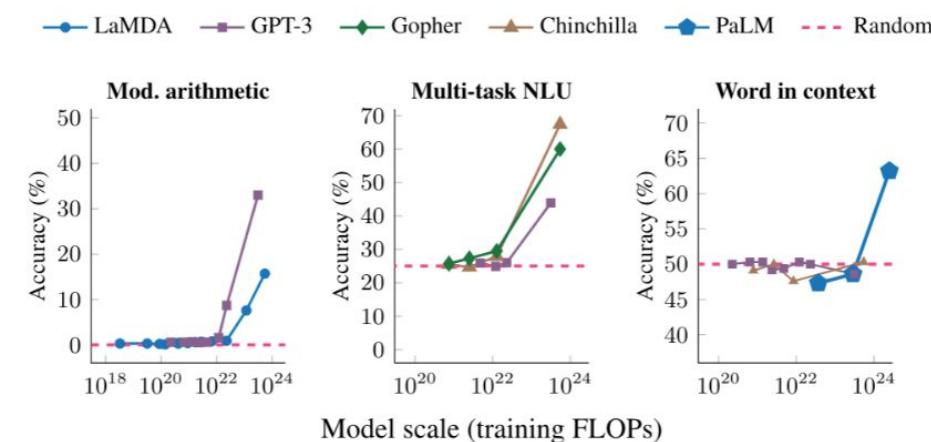
# Emergent Capability

<https://www.jasonwei.net/blog/emergence>

Mathematical abilities have been seen to be non-existent in models up to 13B parameters, similarly reasoning abilities are known to be present in 100B+ parameter models.



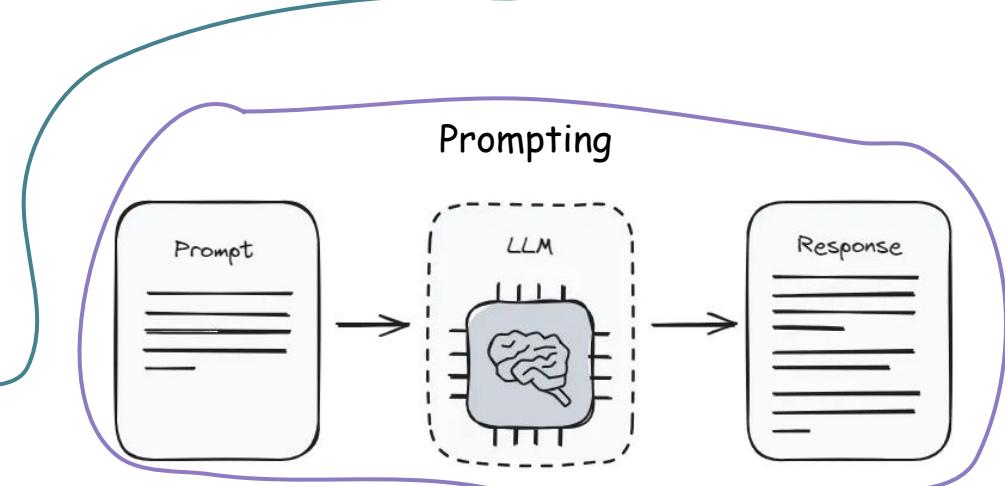
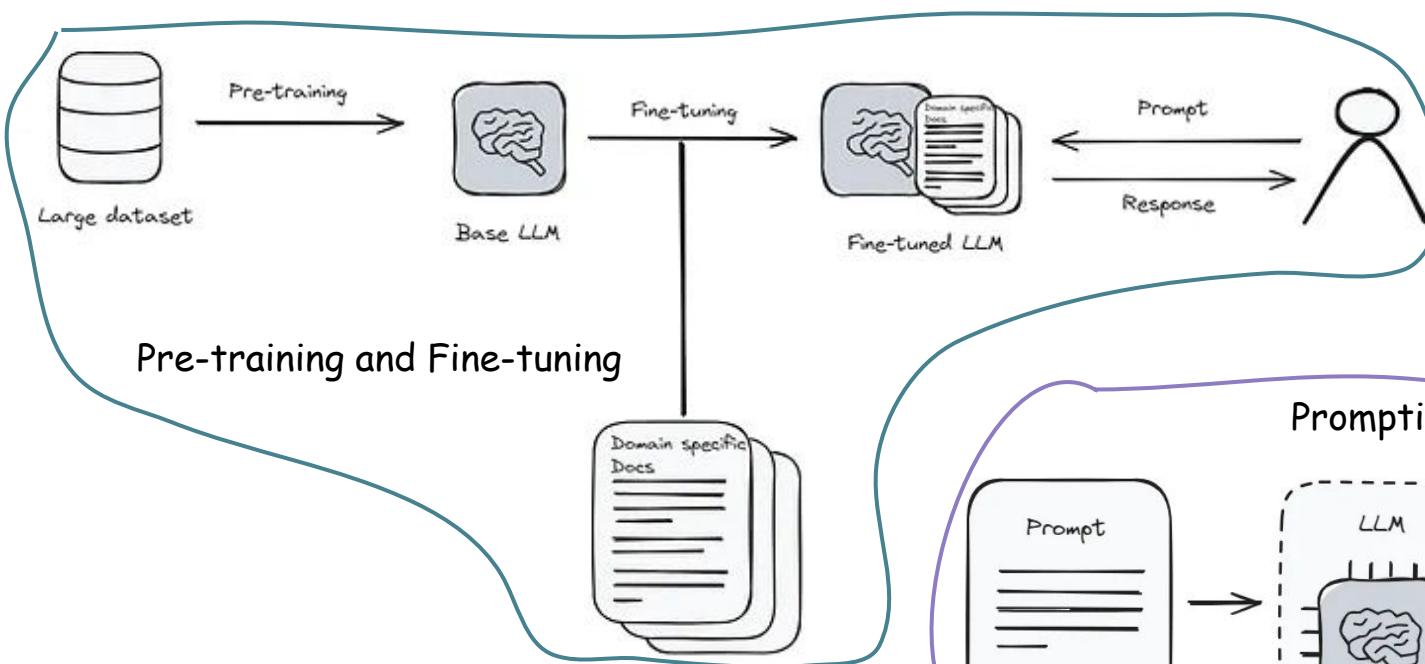
Emergent Capability in Human brain



Emergent Capability in LLMs



# Prompting vs Fine-tuning





# Prompting

- Prompt refers to the input to LLMs
  - **Zero-shot prompt**: only task input, e.g., questions
  - **Instruction prompt**: instruction on what the task is and requirements
  - **Few-shot prompt** (in-context learning): multiple input-output demonstration pairs
  - **Chain-of-Thought prompt**: contains explicit reasoning process to solve a problem



# Emergent Zero-Shot Capability

Source: stanford 224n

Let's revisit the Generative Pretrained Transformer (GPT) models from OpenAI as an example:

**GPT-2** (1.5B parameters; [Radford et al., 2019](#))

- Same architecture as GPT, just bigger (117M  $\rightarrow$  1.5B)
  - But trained on **much more data**: 4GB  $\rightarrow$  40GB of internet text data (WebText)
    - Scrape links posted on Reddit w/ at least 3 upvotes (rough proxy of human quality)
- 

## Language Models are Unsupervised Multitask Learners

---

Alec Radford \*<sup>1</sup> Jeffrey Wu \*<sup>1</sup> Rewon Child<sup>1</sup> David Luan<sup>1</sup> Dario Amodei \*\*<sup>1</sup> Ilya Sutskever \*\*<sup>1</sup>



# Emergent Zero-Shot Capability

- One key emergent ability in GPT-2 is zero-shot learning: the ability to do many tasks with no examples, and no gradient updates, by simply:
  - Specifying the exact prediction problem (e.g. question answering):

Passage: Tom Brady... Q: Where was Tom Brady born? A:

Model Input

What is the sentiment of the following text? Choose from  
'Positive' or 'Negative'

Text: The team's performance last night was top-notch.



# Emergent Zero-Shot Capability

GPT-2 beats SoTA on language modeling benchmarks with **no task-specific fine-tuning**

*Context:* “Why?” “I would have thought you’d find him rather dry,” she said. “I don’t know about that,” said Gabriel.  
“He was a great craftsman,” said Heather. “That he was,” said Flannery.

*Target sentence:* “And Polish, to boot,” said ----- **LAMBADA** (language modeling w/ long discourse dependencies)

*Target word:* Gabriel  
[[Paperno et al., 2016](#)]

	LAMBADA (PPL)	LAMBADA (ACC)	CBT-CN (ACC)	CBT-NE (ACC)	WikiText2 (PPL)
SOTA	99.8	59.23	85.7	82.3	39.14
117M	<b>35.13</b>	45.99	<b>87.65</b>	<b>83.4</b>	<b>29.41</b>
345M	<b>15.60</b>	55.48	<b>92.35</b>	<b>87.1</b>	<b>22.76</b>
762M	<b>10.87</b>	<b>60.12</b>	<b>93.45</b>	<b>88.0</b>	<b>19.93</b>
1542M	<b>8.63</b>	<b>63.24</b>	<b>93.30</b>	<b>89.05</b>	<b>18.34</b>

[[Radford et al., 2019](#)]



# Emergent Few-Shot Capability

**GPT-3** (175B parameters; [Brown et al., 2020](#))

- Another increase in size (1.5B -> **175B**)
  - and data (40GB -> **over 600GB**)
- 

## Language Models are Few-Shot Learners

---

**Tom B. Brown\***

**Benjamin Mann\***

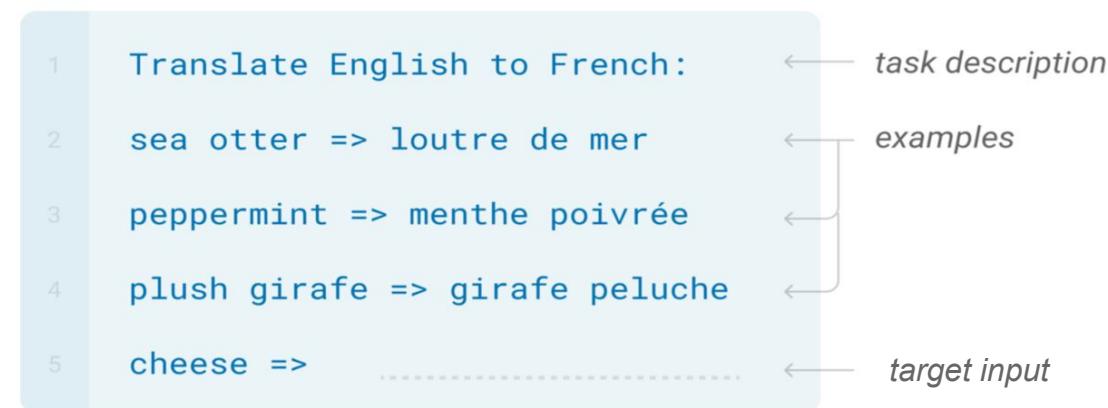
**Nick Ryder\***

**Melanie Subbiah\***



# Emergent Few-Shot Capability

- Specify a task by simply **prepend**ing examples of the task before your example
- Also called **in-context learning**, to stress that *no gradient updates* are performed when learning a new task (there is a separate literature on few-shot learning with gradient updates)



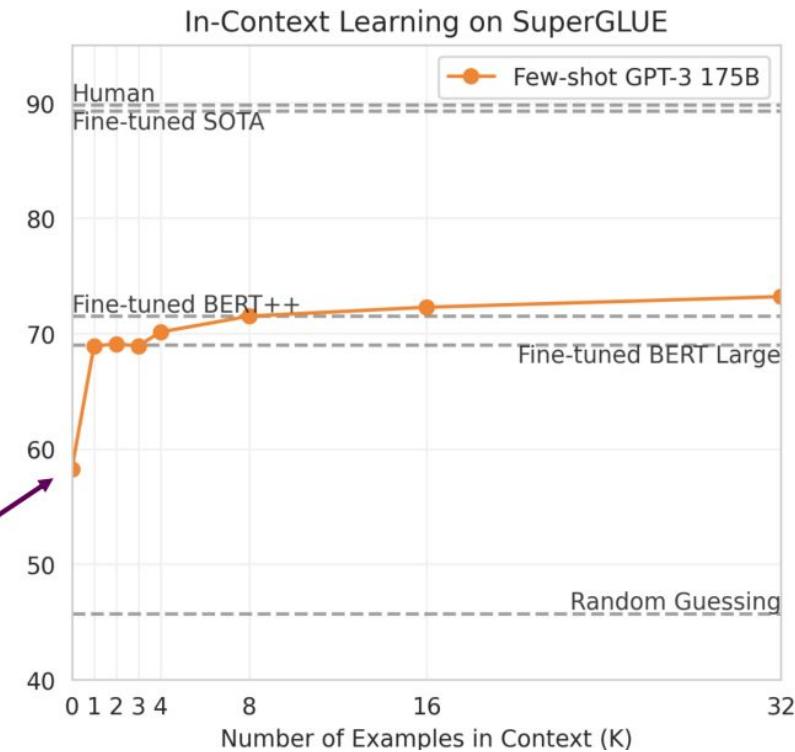


# Emergent Zero-Shot Capability

Source: stanford 224n

**Zero-shot**

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



[Brown et al., 2020]

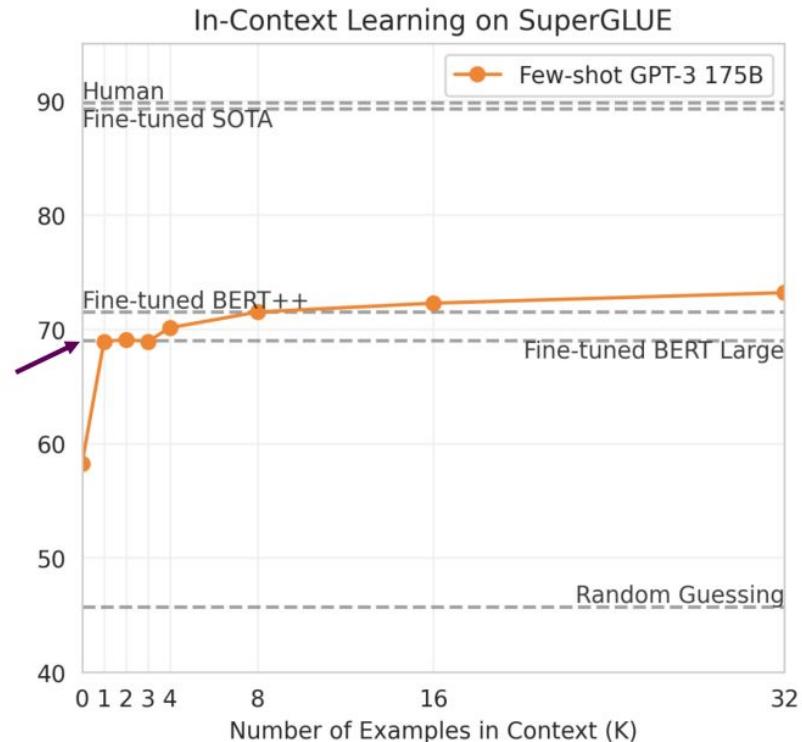


# Emergent Few-Shot Capability

Source: stanford 224n

## One-shot

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



[Brown et al., 2020]

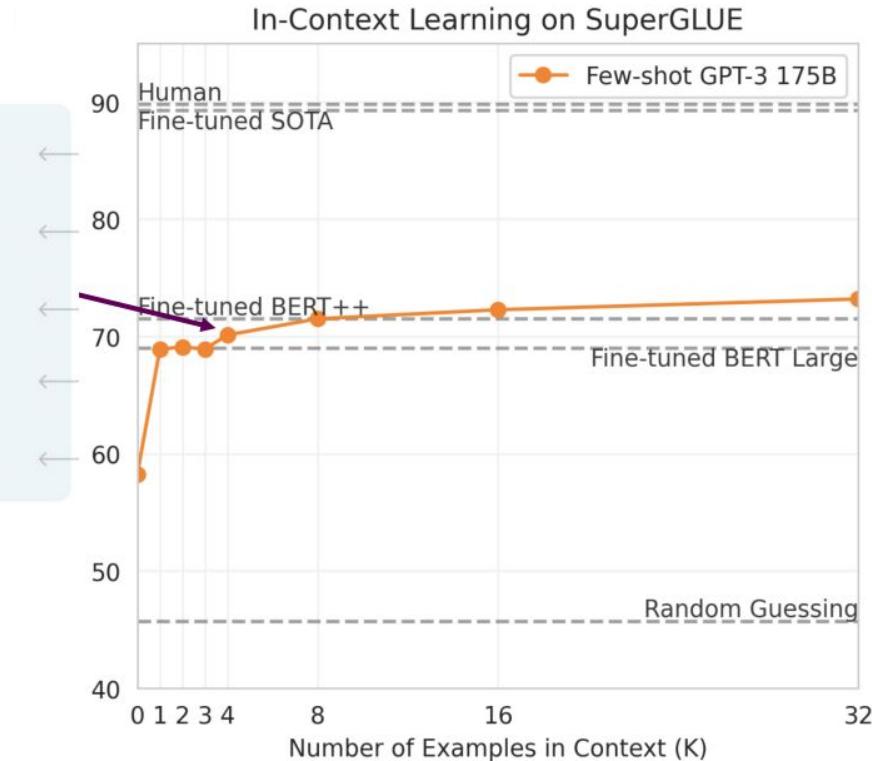


# Emergent Few-Shot Capability

Source: stanford 224n

## Few-shot

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



[Brown et al., 2020]

# Emergent Few-Shot Capability

Source: stanford 224n

Cycle letters:

pleap ->

apple

Random insertion:

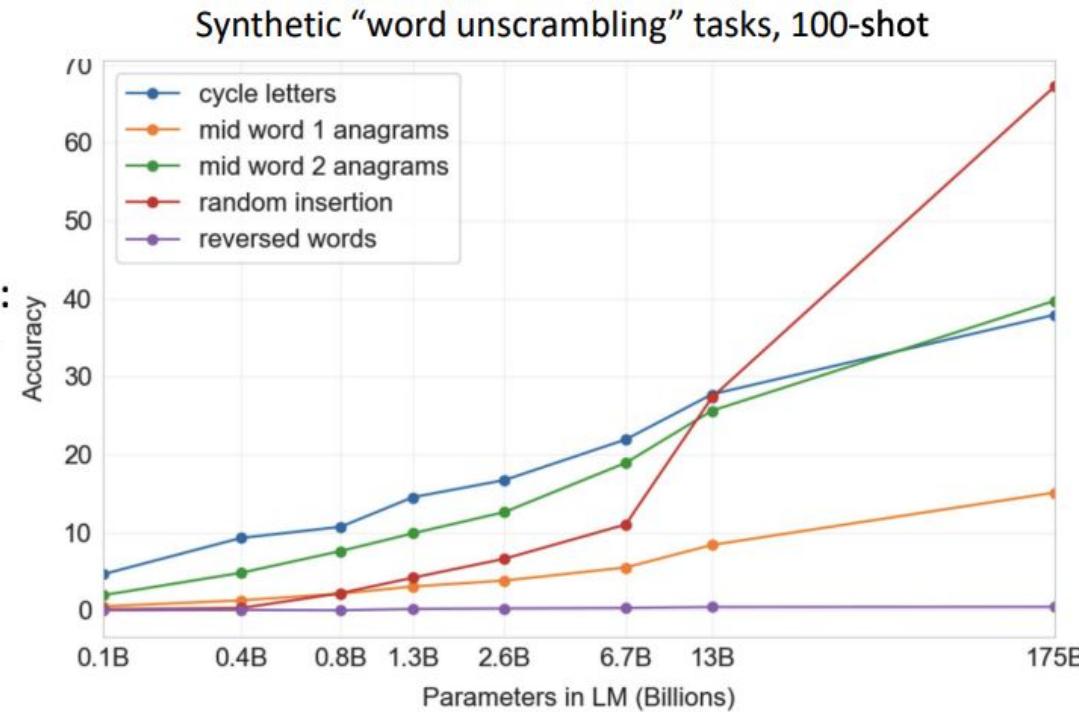
a.p!p/l!e ->

apple

Reversed words:

elppa ->

apple

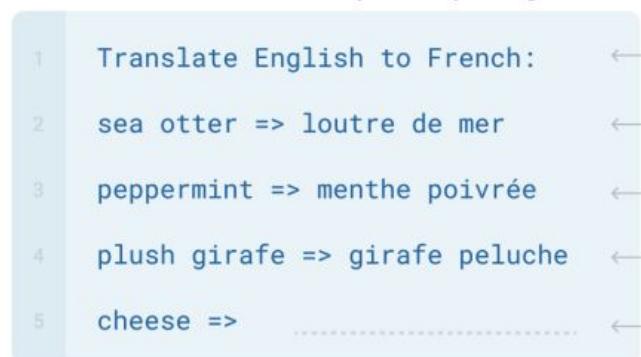




# Fine-tuning vs Few-Shot

Source: stanford 224n

## Zero/few-shot prompting

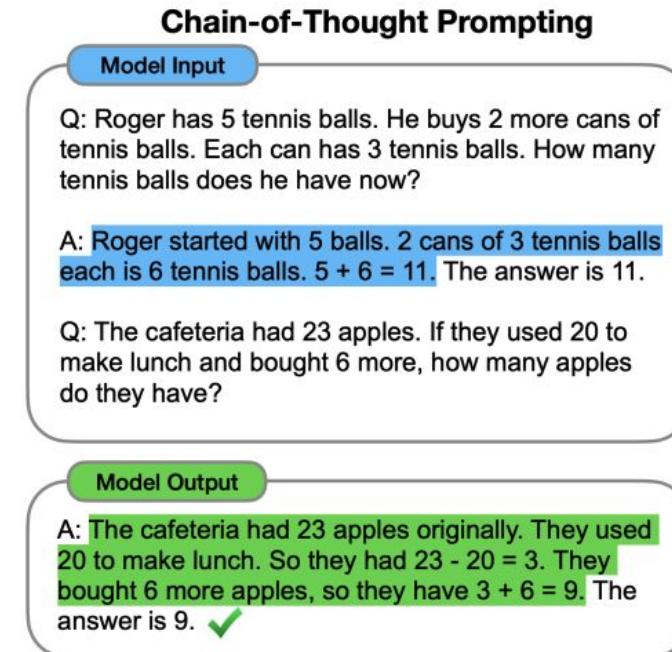
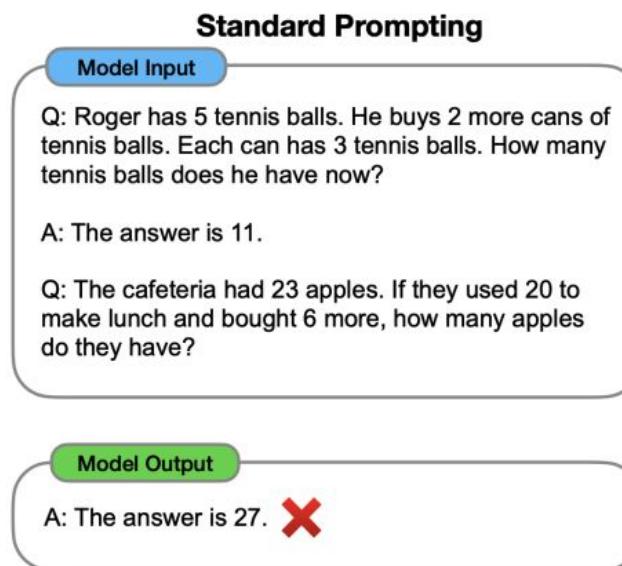


## Traditional fine-tuning

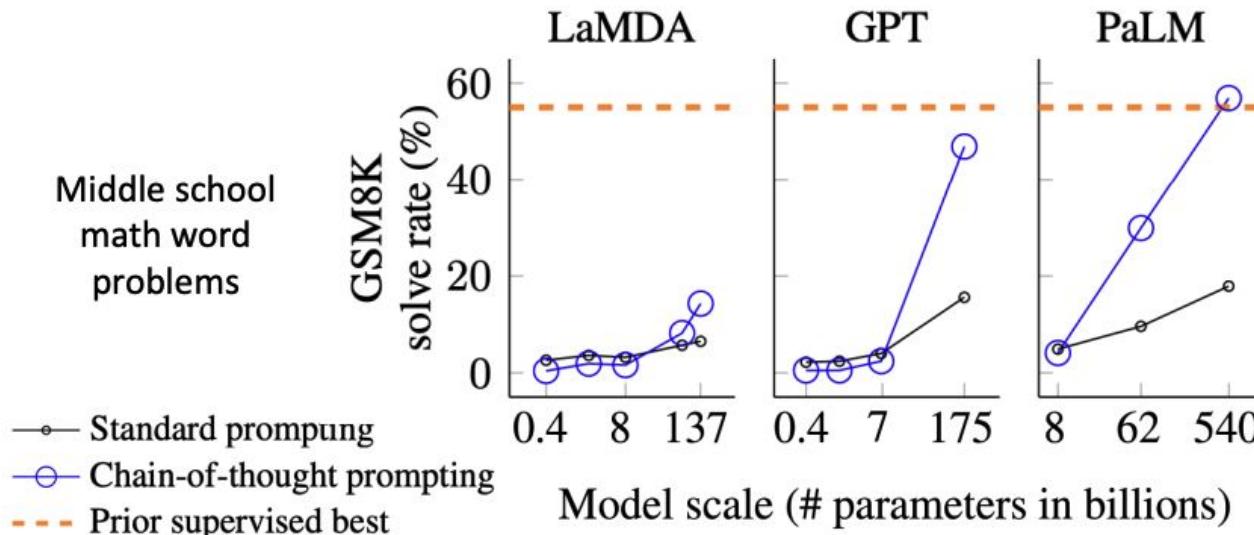




# Chain-Of-Thought Prompting



# Chain-Of-Thought Prompting





# 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. ✓

**Do we even need examples of reasoning?  
Can we just ask the model to reason through things?**



# Zero 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. ✓

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



# Zero-Shot Chain-Of-Thought Prompting

Source: stanford 224n

	MultiArith	GSM8K
<b>Zero-Shot</b>	<b>17.7</b>	<b>10.4</b>
Few-Shot (2 samples)	33.7	15.6
Few-Shot (8 samples)	33.8	15.6
<b>Zero-Shot-CoT</b>	<b>Greatly outperforms → 78.7</b>	<b>40.7</b>
Few-Shot-CoT (2 samples)	<b>zero-shot</b>	84.8
Few-Shot-CoT (4 samples : First) (*1)		89.2
Few-Shot-CoT (4 samples : Second) (*1)	<b>Manual CoT → 90.5</b>	-
Few-Shot-CoT (8 samples)	<b>still better</b>	48.7



# Zero-Shot Chain-Of-Thought Prompting

No.	Category	Zero-shot CoT Trigger Prompt	Accuracy
1	LM-Designed	Let's work this out in a step by step way to be sure we have the right answer.	<b>82.0</b>
2	Human-Designed	Let's think step by step. (*1)	78.7
3		First, (*2)	77.3
4		Let's think about this logically.	74.5
5		Let's solve this problem by splitting it into steps. (*3)	72.2
6		Let's be realistic and think step by step.	70.8
7		Let's think like a detective step by step.	70.3
8		Let's think	57.5
9		Before we dive into the answer,	55.7
10		The answer is after the proof.	45.7
-		(Zero-shot)	17.7



# How Can We Achieve This?

**01** Prompting / In-Context Learning

**02** Instruction Fine-tuning

**03** Reinforcement Learning with Human Feedback



# Make the model sensitive to instructions

PROMPT *Explain the moon landing to a 6 year old in a few sentences.*

COMPLETION GPT-3

Explain the theory of gravity to a 6 year old.

Explain the theory of relativity to a 6 year old in a few sentences.

Explain the big bang theory to a 6 year old.

Explain evolution to a 6 year old.

Language models are not *aligned* with user intent [[Ouyang et al., 2022](#)].



# Make the model sensitive to instructions

PROMPT *Explain the moon landing to a 6 year old in a few sentences.*

COMPLETION **Human**

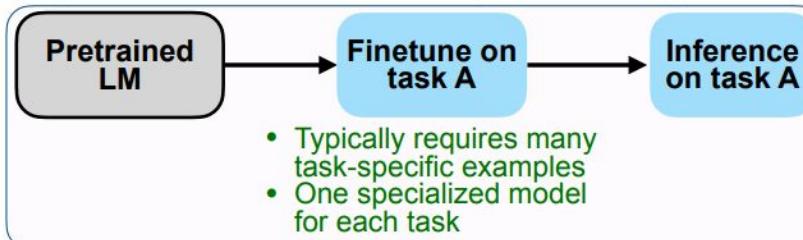
A giant rocket ship blasted off from Earth carrying astronauts to the moon. The astronauts landed their spaceship on the moon and walked around exploring the lunar surface. Then they returned safely back to Earth, bringing home moon rocks to show everyone.

Language models are not *aligned* with user intent [[Ouyang et al., 2022](#)].  
Finetuning to the rescue!

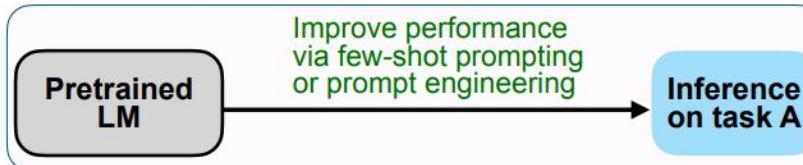


# Instruction Fine-tuning

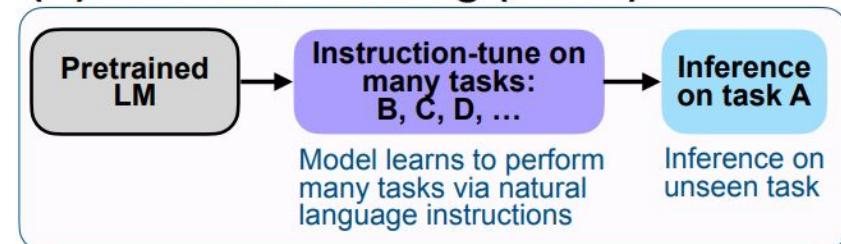
## (A) Pretrain–finetune (BERT, T5)



## (B) Prompting (GPT-3)



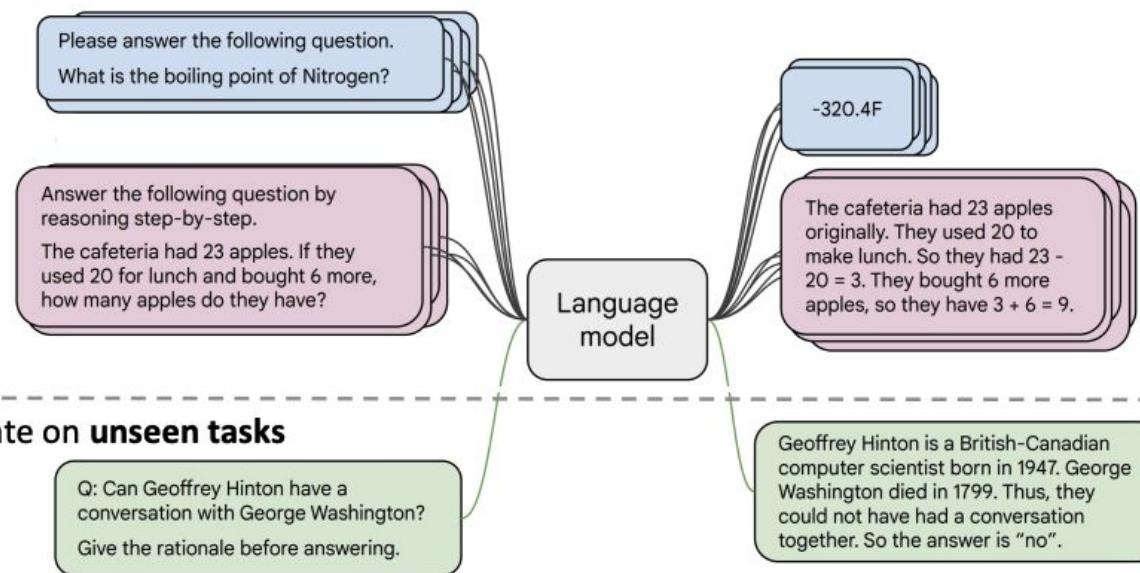
## (C) Instruction tuning (FLAN)





# Instruction Fine-tuning – FlanT5

- Collect examples of (instruction, output) pairs across many tasks and finetune an LM





# Instruction Fine-tuning – FlanT5

Finetune on many tasks (“instruction-tuning”)

## Input (Commonsense Reasoning)

Here is a goal: Get a cool sleep on summer days.

How would you accomplish this goal?

OPTIONS:

- Keep stack of pillow cases in fridge.
- Keep stack of pillow cases in oven.

## Target

keep stack of pillow cases in fridge

## Input (Translation)

Translate this sentence to Spanish:

The new office building was built in less than three months.

## Target

El nuevo edificio de oficinas se construyó en tres meses.

Sentiment analysis tasks

Coreference resolution tasks

...



Inference on unseen task type

## Input (Natural Language Inference)

Premise: At my age you will probably have learnt one lesson.

Hypothesis: It's not certain how many lessons you'll learn by your thirties.

Does the premise entail the hypothesis?

OPTIONS:

- yes
- it is not possible to tell
- no

## FLAN Response

It is not possible to tell



# Instruction Fine-tuning – FlanT5

## Model input (Disambiguation QA)

Q: In the following sentences, explain the antecedent of the pronoun (which thing the pronoun refers to), or state that it is ambiguous.

Sentence: The reporter and the chef will discuss their favorite dishes.

Options:

- (A) They will discuss the reporter's favorite dishes
- (B) They will discuss the chef's favorite dishes
- (C) Ambiguous

A: Let's think step by step.

## Before instruction finetuning

The reporter and the chef will discuss their favorite dishes.

The reporter and the chef will discuss the reporter's favorite dishes.

The reporter and the chef will discuss the chef's favorite dishes.

The reporter and the chef will discuss the reporter's and the chef's favorite dishes.

✖ (doesn't answer question)



# Instruction Fine-tuning – FlanT5

## Model input (Disambiguation QA)

Q: In the following sentences, explain the antecedent of the pronoun (which thing the pronoun refers to), or state that it is ambiguous.

Sentence: The reporter and the chef will discuss their favorite dishes.

Options:

- (A) They will discuss the reporter's favorite dishes
- (B) They will discuss the chef's favorite dishes
- (C) Ambiguous

A: Let's think step by step.

## After instruction finetuning

The reporter and the chef will discuss their favorite dishes does not indicate whose favorite dishes they will discuss. So, the answer is (C).



## Instruction Fine-tuning

- The Super-NaturalInstructions dataset contains over 1.6K tasks, 3M+ examples: Classification, sequence tagging, rewriting, translation, QA
  - Train T5 on these instruction datasets.
  - Demonstrate strong generalizability compared to instructGPT.



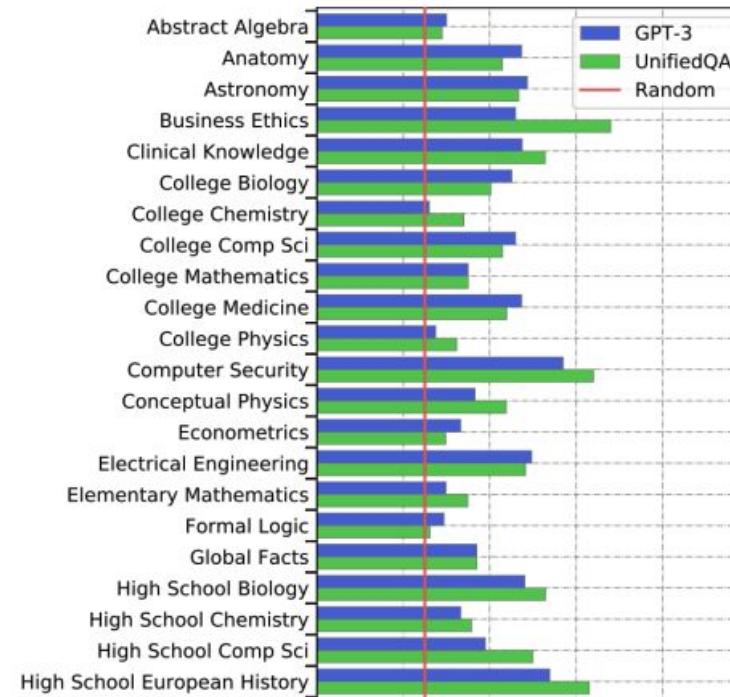


# Common Evaluation Set

Source: stanford 224n

## Massive Multitask Language Understanding (MMLU) [Hendrycks et al., 2021]

New benchmarks for measuring LM performance on 57 diverse *knowledge intensive* tasks





# Common Evaluation Set

**BIG-Bench** [Srivastava et al., 2022]

200+ tasks, spanning:



[https://github.com/google/BIG-bench/blob/main/bigbench/benchmark\\_tasks/README.md](https://github.com/google/BIG-bench/blob/main/bigbench/benchmark_tasks/README.md)

# BEYOND THE IMITATION GAME: QUANTIFYING AND EXTRAPOLATING THE CAPABILITIES OF LANGUAGE MODELS

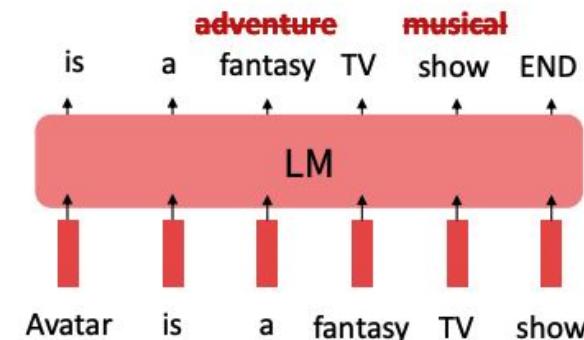
#### **Alphabetic author list**



# Limitations

Source: stanford 224n

- One limitation of instruction finetuning is obvious: it's **expensive** to collect ground-truth data for tasks.
- But there are other, subtler limitations too. Can you think of any?
- **Problem 1:** tasks like open-ended creative generation have no right answer.
  - *Write me a story about a dog and her pet grasshopper.*
- **Problem 2:** language modeling penalizes all token-level mistakes equally, but some errors are worse than others.
- Even with instruction finetuning, there a mismatch between the LM objective and the objective of "satisfy human preferences"!
- Can we **explicitly attempt to satisfy human preferences?**





# How Can We Achieve This?

- 01 Prompting / In-Context Learning**
- 02 Instruction Fine-tuning**
- 03 Reinforcement Learning with Human Feedback**



# Optimizing for Human Preferences

Source: stanford 224n

- Let's say we were training a language model on some task (e.g. summarization).
- For each LM sample  $s$ , imagine we had a way to obtain a *human reward* of that summary:  $R(s) \in \mathbb{R}$ , higher is better.

SAN FRANCISCO,  
California (CNN) --  
A magnitude 4.2  
earthquake shook the  
San Francisco

...  
overturn unstable  
objects.

An earthquake hit  
San Francisco.  
There was minor  
property damage,  
but no injuries.

$$s_1 \\ R(s_1) = 8.0$$

The Bay Area has  
good weather but is  
prone to  
earthquakes and  
wildfires.

$$s_2 \\ R(s_2) = 1.2$$

- Now we want to maximize the expected reward of samples from our LM:

$$\mathbb{E}_{\hat{s} \sim p_\theta(s)}[R(\hat{s})]$$



# How to Model Human Preferences

Source: stanford 224n

- **Problem 1:** human-in-the-loop is expensive!
  - **Solution:** instead of directly asking humans for preferences, **model their preferences** as a separate (NLP) problem! [[Knox and Stone, 2009](#)]

An earthquake hit  
San Francisco.  
There was minor  
property damage,  
but no injuries.

The Bay Area has  
good weather but is  
prone to  
earthquakes and  
wildfires.

Train an LM  $RM_\phi(s)$  to  
predict human  
preferences from an  
annotated dataset, then  
optimize for  $RM_\phi$  instead.

$s_1$   
 $R(s_1) = 8.0$    


$s_2$   
 $R(s_2) = 1.2$    




# How to Model Human Preferences

Source: stanford 224n

- **Problem 2:** human judgments are noisy and miscalibrated!
- **Solution:** instead of asking for direct ratings, ask for **pairwise comparisons**, which can be more reliable [[Phelps et al., 2015; Clark et al., 2018](#)]

A 4.2 magnitude  
earthquake hit  
San Francisco,  
resulting in  
massive damage.

$s_3$

$$R(s_3) = \text{4.1? } \text{6.6? } \text{3.2?}$$



# How to Model Human Preferences

Source: stanford 224n

- **Problem 2:** human judgments are noisy and miscalibrated!
- **Solution:** instead of asking for direct ratings, ask for **pairwise comparisons**, which can be more reliable [[Phelps et al., 2015](#); [Clark et al., 2018](#)]

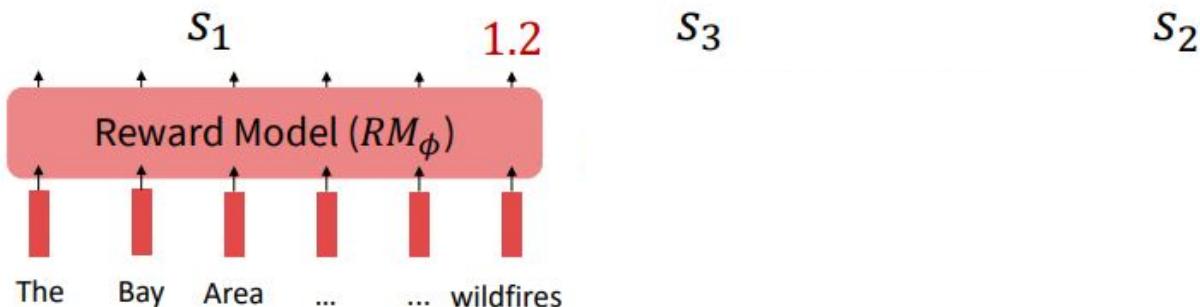
An earthquake hit  
San Francisco.  
There was minor  
property damage,  
but no injuries.

&gt;

A 4.2 magnitude  
earthquake hit  
San Francisco,  
resulting in  
massive damage.

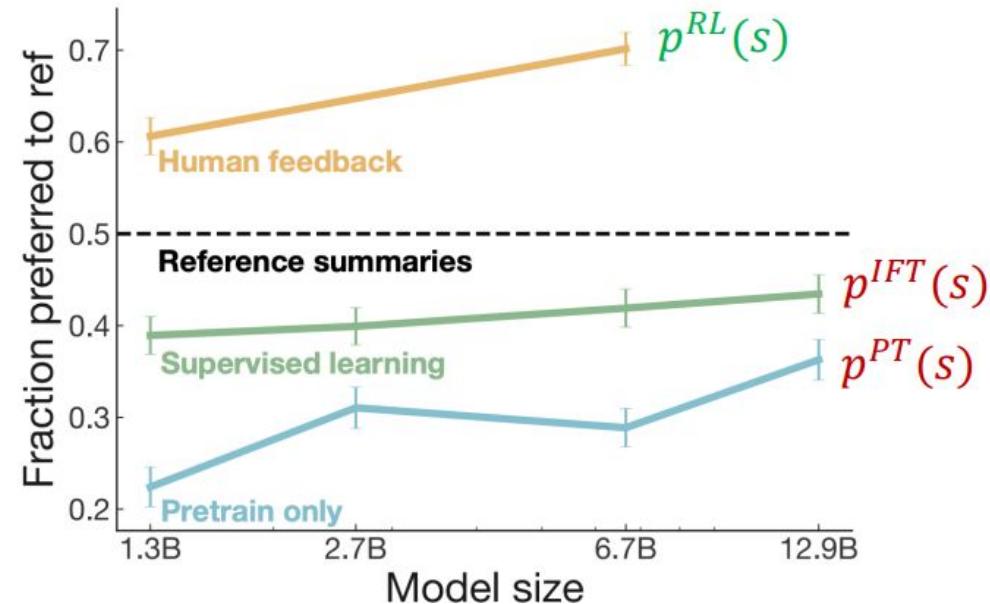
&gt;

The Bay Area has  
good weather but is  
prone to  
earthquakes and  
wildfires.



# RLHF: Reinforcement Learning with Human Feedback

RLHF provides gains over pretraining + finetuning



[Stiennon et al., 2020]



# Limitations

- Human preferences are unreliable!
  - “Reward hacking” is a common problem in RL
  - Chatbots are rewarded to produce responses that *seem* authoritative and helpful, *regardless of truth*
  - This can result in making up facts + hallucinations

<https://news.ycombinator.com/item?id=34776508>

<https://apnews.com/article/kansas-city-chiefs-philadelphia-eagles-technology-science-82bc20f207e3e4cf81abc6a5d9e6b23a>

Source: stanford 224n

Google shares drop \$100 billion after its new AI chatbot makes a mistake

February 9, 2023 · 10:15 AM ET

<https://www.npr.org/2023/02/09/1155650909/google-chatbot--error-bard-shares>

Bing AI hallucinates the Super Bowl

The screenshot shows a Bing search interface. At the top right, there is a blue button labeled "Who won the superbowl?". On the left, there are two green checkmarks: "Searching for: superbowl winner" and "Generating answers for you...". Below this, a snippet of text reads: "The Super Bowl is the annual American football game that determines the champion of the National Football League (NFL) [1]. The most recent Super Bowl was Super Bowl LVI, which was held on February 6, 2023 at SoFi Stadium in Inglewood, California [2]. The winner of that game was the Philadelphia Eagles [3].". At the bottom, the text "The most recent Super Bowl was Super Bowl LVI, Eagles, who defeated the Kansas City Chiefs by 31-24" is displayed in large, bold letters.



# ChatGPT: Instruction Fine-tuning + RLHF

## ChatGPT: Optimizing Language Models for Dialogue

Note: OpenAI (and similar companies) are keeping more details secret about ChatGPT training (including data, training parameters, model size)—perhaps to keep a competitive edge...

### Methods

We trained this model using Reinforcement Learning from Human Feedback (RLHF), using the same methods as InstructGPT, but with slight differences in the data collection setup. We trained an initial model using supervised fine-tuning: human AI trainers provided conversations in which they played both sides—the user and an AI assistant. We gave the trainers access to model-written suggestions to help them compose their responses. We mixed this new dialogue dataset with the InstructGPT dataset, which we transformed into a dialogue format.

**(Instruction finetuning!)**



# ChatGPT: Instruction Fine-tuning + RLHF

## ChatGPT: Optimizing Language Models for Dialogue

Note: OpenAI (and similar companies) are keeping more details secret about ChatGPT training (including data, training parameters, model size)—perhaps to keep a competitive edge...

### Methods

To create a reward model for reinforcement learning, we needed to collect comparison data, which consisted of two or more model responses ranked by quality. To collect this data, we took conversations that AI trainers had with the chatbot. We randomly selected a model-written message, sampled several alternative completions, and had AI trainers rank them. Using these reward models, we can fine-tune the model using Proximal Policy Optimization. We performed several iterations of this process.

(RLHF!)