Innovation Growth Lab by nesta

Natural Language Processing in Economics

Extra: Large Language Models

The journey onwards

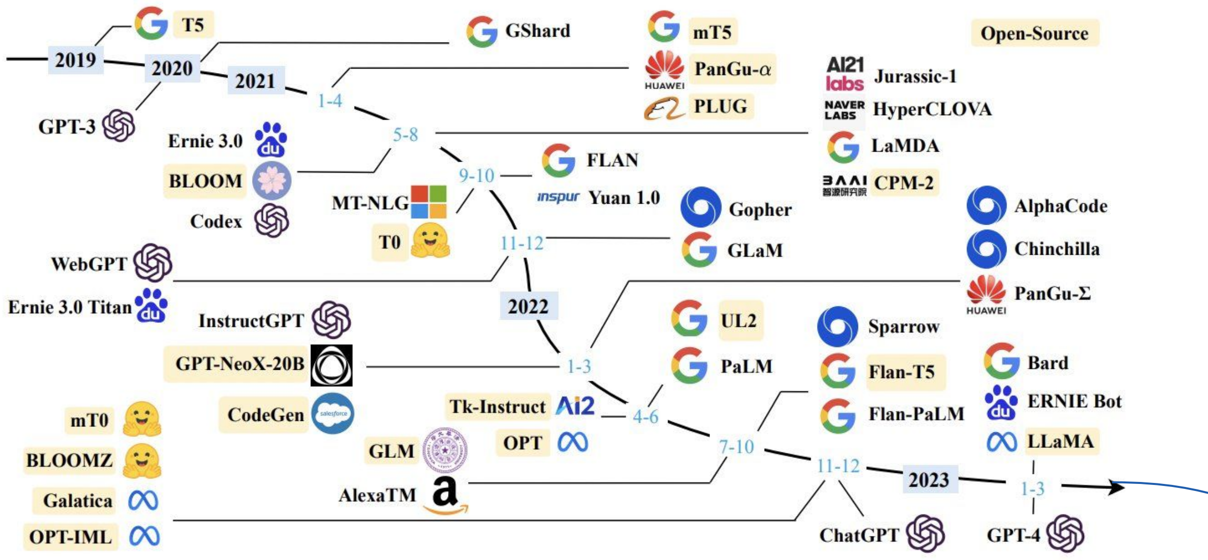


Fig. 1. A timeline of existing large language models (having a size larger than 10B) in recent years. We mark the open-source LLMs in yellow color.

Let's make it here!

We're here!

Evolution of Generative Pre-Trained Transformers

GPT-1 (2018) - 117M params (12 x 12)

- Trained on the Books corpus (4.6GB) (Radford et al 2018).
- Trained on a language modeling task, as well as a multi-task that adds a supervised learning task

GPT-2 (2019) - 1.5B params (48 x 48)

- All articles linked from Reddit with at least 3 upvotes (8 million documents, 40GB of text) (Radford 2019)
- Dispense with supervised learning task, make adjustments, enlarge the model by many orders of magnitud

GPT-3 (2020) - 175B params (96 x 96)

- Use an even larger corpus (Common Crawl, WebText2, Books1, Books2, Wikipedia). Brown et al. 2020
- Model becomes much larger

InstructGPT ↔ GPT-3.5 ↔ GPT-4

- Add reinforcement learning with human feedback to improve responsiveness & user satisfaction
- Closed-source, controversial for OpenAI - an organisation built on the backs of the open-source crowd

Large Language Models

- Language Modeling refers to the task of teaching an algorithm to generate language

- Standard practice builds from Markov chains: assume future words are independent of the past given the present and some finite number of previous features.

$$\Pr(w_{i+1}|w_{1:i}) \approx \Pr(w_{i+1}|w_{i-k:i})$$

- Three types of Transformer-based Language Models

    - **Autoregressive models** (GPTs)
        - Standard language modeling task: predict next token having read the masked sentence
    - **Autoencoding models** (BERT)
        - Pre-trained by shuffling / masking input tokens and predict sequence, bidirectional
    - **Sequence-to-sequence models** (Vaswani paper, Machine Translation)
        - Models combine encoders and decoders, unit of prediction is the token sequence
    - Long-context transformers (BigBird)

What is an LLM able to learn from training?

- **Trivia**: “Universitat Politècnica de Valencia is located in __________, Comunitat Valenciana”
- **Syntax**: “I put ____ fork down on the table”
- **Coreference**: “The woman crossed the road, checking for traffic over ____ shoulder”
- **Lexical semantics**: “I went to the zoo to see tigers, bears, monkeys, and _____”
- **Sentiment:** “The best thing about the movie were my popcorns and soda. The movie was ____”
- **Reasoning (meh)**: “He went into the kitchen. Standing next to him, she made coffee. He left the ___”
- **Arithmetic (very meh)**: “The sequence I have i mid is 1, 1, 2, 3, 5, 8, 13, 21, ____”
- **Model agents, beliefs, and actions (?)**: Sung Park et al. (2023)
- **Code**: “async function isPositive(text:string): Promise <boolean> {_______”

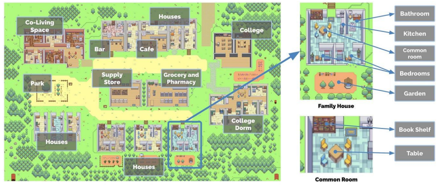


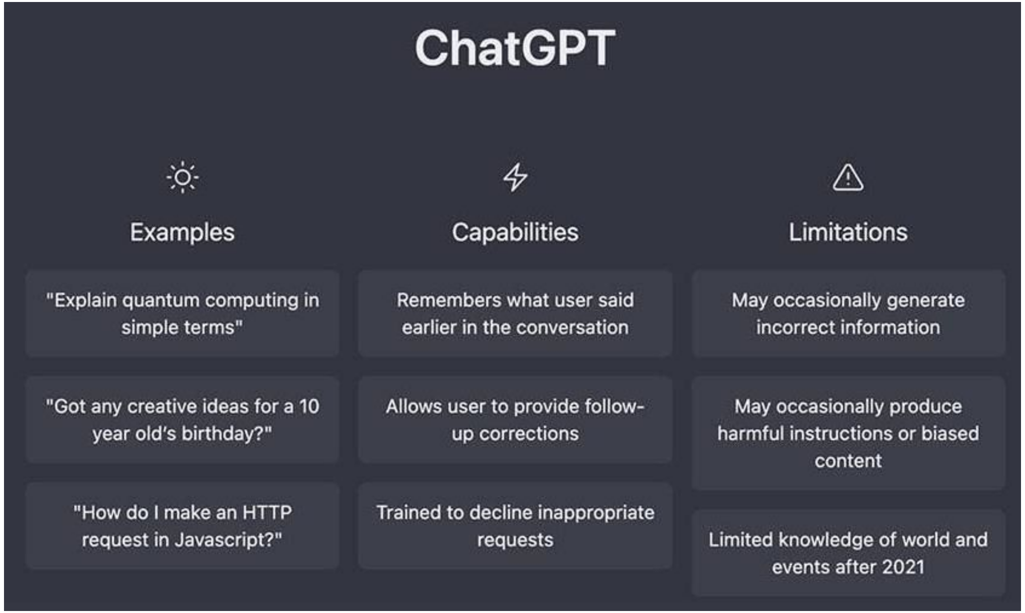
Figure 2: The Smallville sandbox world, with areas labeled. The root node describes the entire world, children describe areas (e.g., houses, cafe, stores), and leaf nodes describe objects (e.g., table, bookshelf). Agent remember a subgraph reflecting the parts of the world they have seen, in the state that they saw them.

What this does not explain

How do we go from this

“Universitat Politècnica de Valencia is located in __________, Comunitat Valenciana”

to this



From trained language models to Assistants

- Chatbots have been around since the 1960s, and since late in the decade systems have already been able to pass the Turing test.

- Rule-based systems use pattern-action rules

```
function ELIZA GENERATOR(user sentence) returns response

Find the word w in sentence that has the highest keyword rank
  if w exists
      Choose the highest ranked rule r for w that matches sentence
      response ← Apply the transform in r to sentence
      if w = 'my'
          future ← Apply a transformation from the 'memory' rule list to sentence
          Push future onto memory stack
  else (no keyword applies)
    either
        response ← Apply the transform for the NONE keyword to sentence
    or
        response ← Pop the top response from the memory stack
 return(response)
```

From trained language models to Assistants

- A major concern is the ability of machine learning systems to replicate biases that occurred in the training data. This was an acute issue when chatbots learn dynamically

- Microsoft's **Tay chatbot** (2016)
  - It only took 16 hours before it had to be taken down
  - In that period of time it started posting racial slurs, conspiracy theories, etc.

From trained language models to Assistants

- What has changed?
  - Zero shot in-context learning
  - Few shot in-context learning
  - Instruction fine tuning
  - Reinforcement Learning with Human Feedback

Zero-shot learning (GPT-2)

- An emergent ability in GPT-2 was zero-shot learning: the ability to do many tasks with no examples, and no gradient updates, by simply
    - Specifying the right sequence prediction problem

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

Or "TL;DR"!

- Comparing probabilities of sequences

```
The cat couldn't fit into the hat because it was too big.
Does it = the cat or the hat?

    ≡ Is P(...because the cat was too big) >=
         P(...because the hat was too big)?
```

Zero-shot learning (GPT-2)

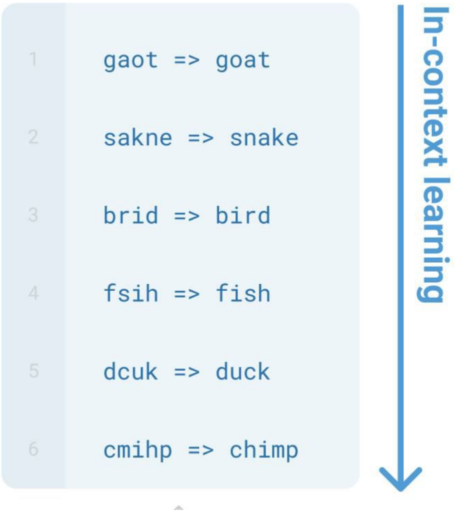
- GPT-2 surpassed state-of-the-art models on language modeling benchmarks with **no task-specific fine-tuning**

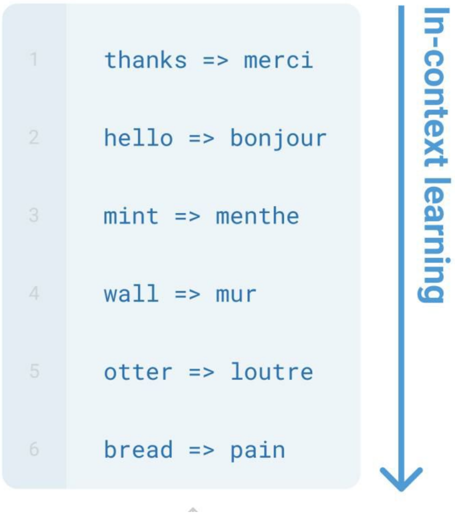
Context: "Why?" "I would have thought you'd find him rather dry," she said. "I don't know about that," said Gabriel. "He was a great craftsman," said Heather. "That he was," said Flannery.
Target sentence: "And Polish, to boot," said _____.
Target word: Gabriel

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

Few-shot learning (GPT-3)

- Specify a task by simply **prepending examples of the task before**
  - In-context learning = no gradient updates when learning a task

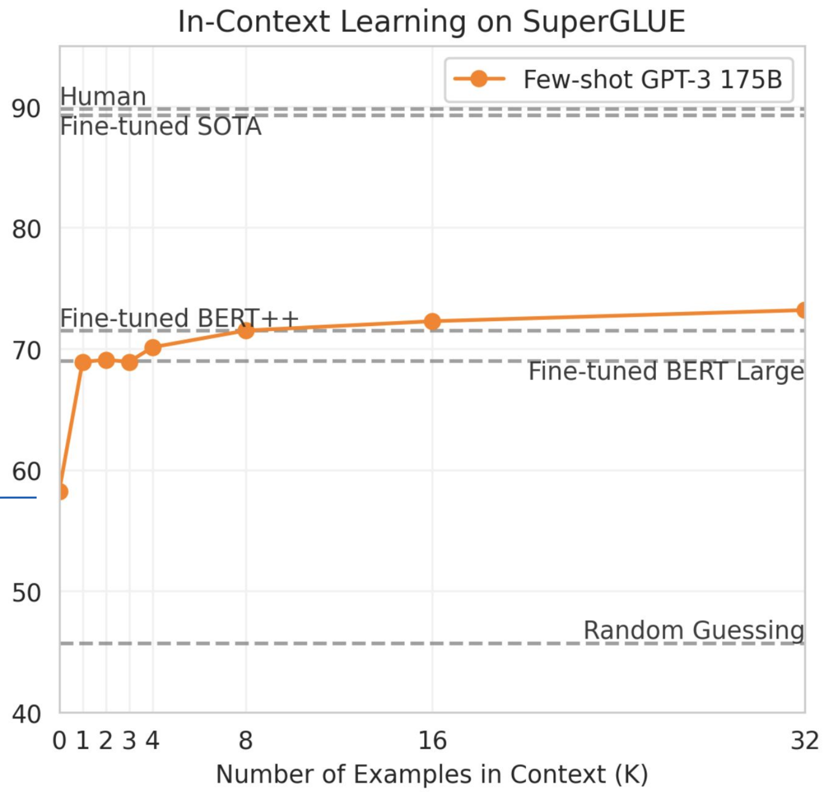




Few-shot learning (GPT-3)

Zero shot

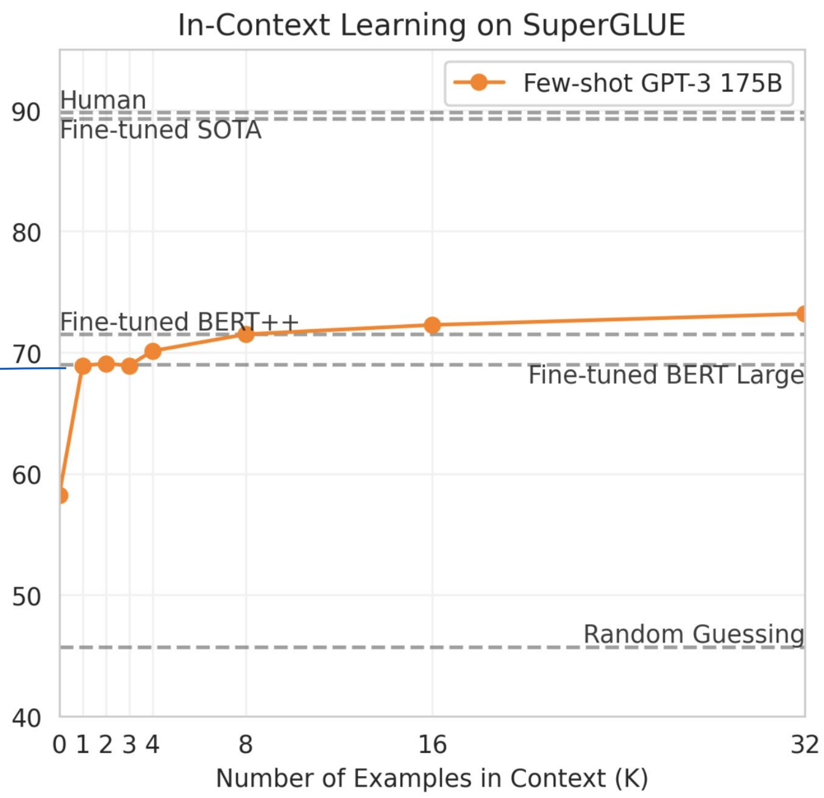
```
Translate English to French:
cheese =>
```



Few-shot learning (GPT-3)

One shot

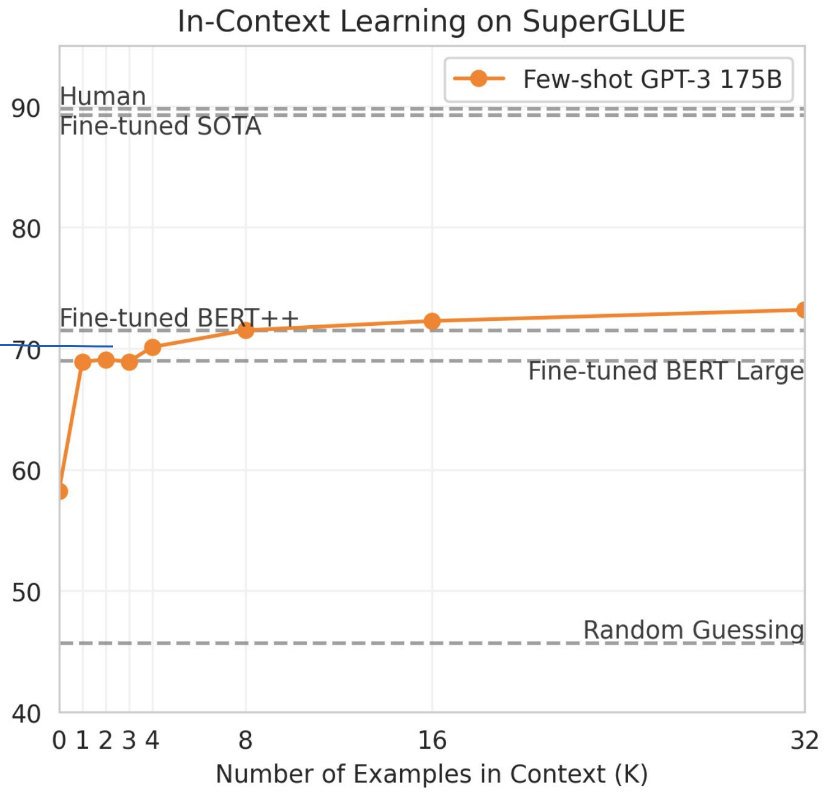
```
Translate English to French:
sea otter => loutre de mer
cheese =>
```



Few-shot learning (GPT-3)

Few shot

```
Translate English to French:
sea otter => loutre de mer
peppermint => menthe poivrée
plush girafe => girafe peluche
cheese => ..............................
```



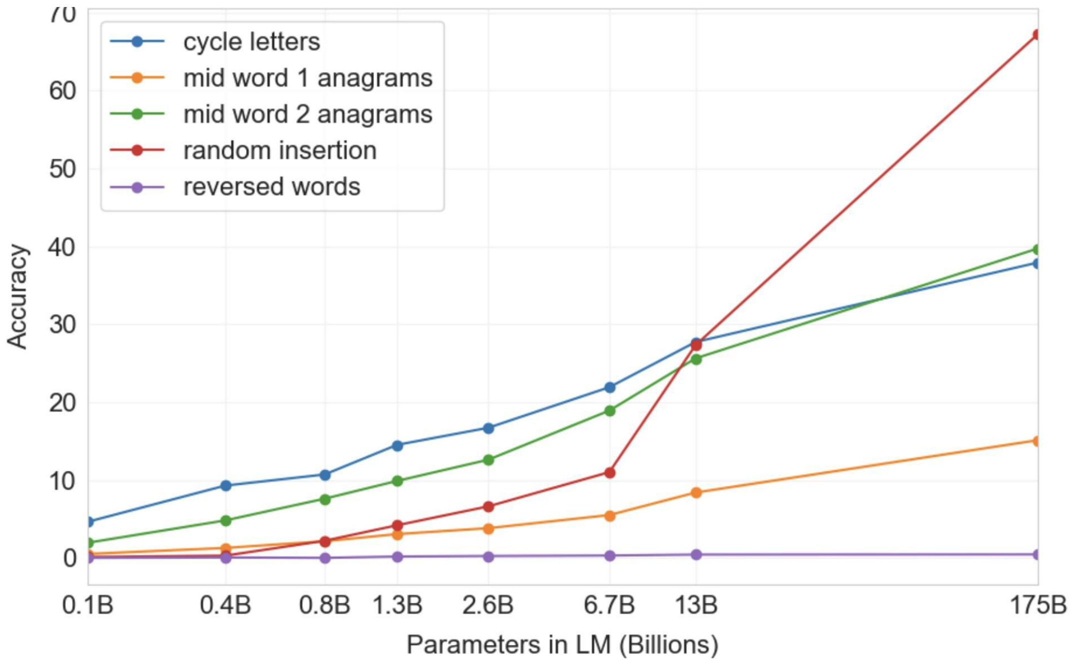
Few-shot learning (GPT-3)

An emerging property of **model scale**

Cycle: **pleap -> apple**

R. Insertion: **ap.lp.!/e -> apple**

Reversed: **elppa -> apple**



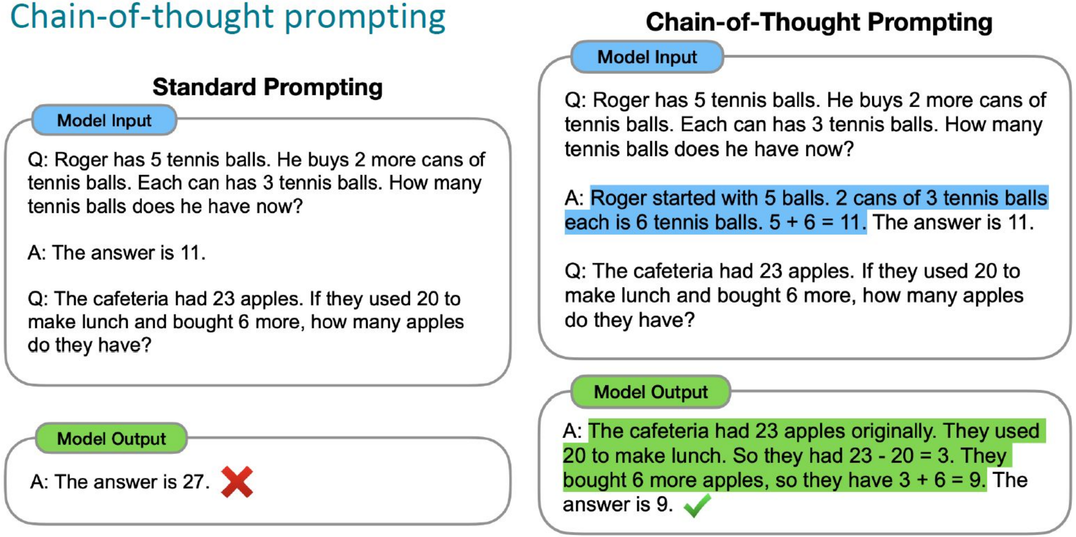
[Brown et al 2020]()

Few-shot learning (GPT-3)

Some tasks seem too hard for even large language models to learn through prompting

- These tend to be richer, multi-step reasoning processes
- **Solution:** change the prompt

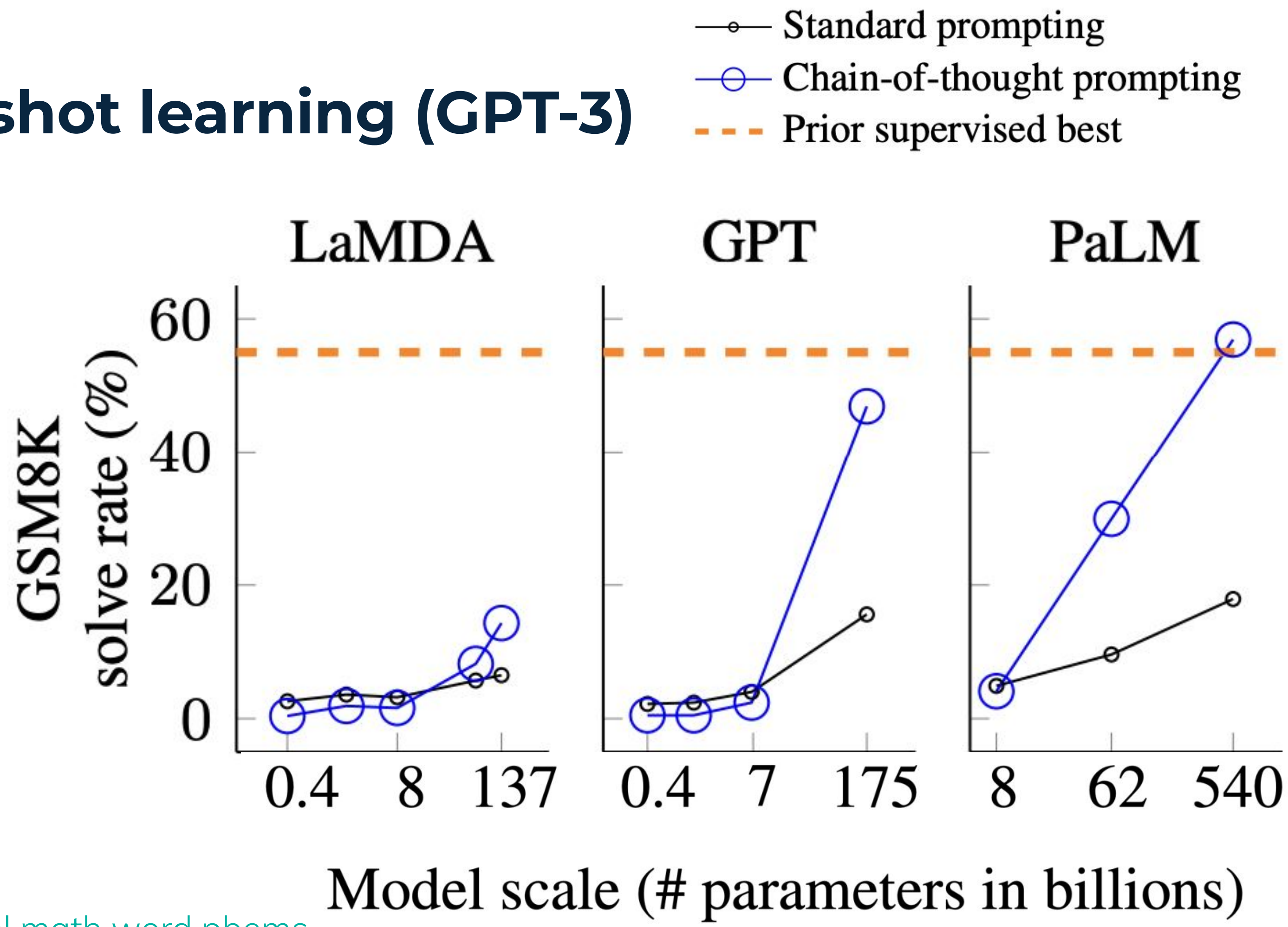
Chain-of-thought prompting



Or just say "Let's think step by step"!

Wei et al 2022

Few-shot learning (GPT-3)



[Middle school math word pbems]

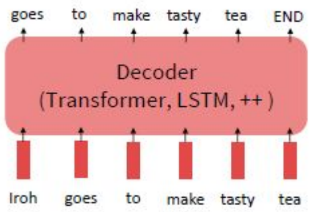
From trained language models to Assistants

- **Zero and few shot in-context learning**
  Pros: No fine tuning needed, prompt engineering can yield much better results
  Cons: Limits to what you can fit in-context, gradient steps not outdated

- Instruction fine tuning

- Reinforcement Learning with Human Feedback

Language models do not assist users



PROMPT	*Explain the moon landing to a 6 year old in a few sentences.*
COMPLETION	GPT-3

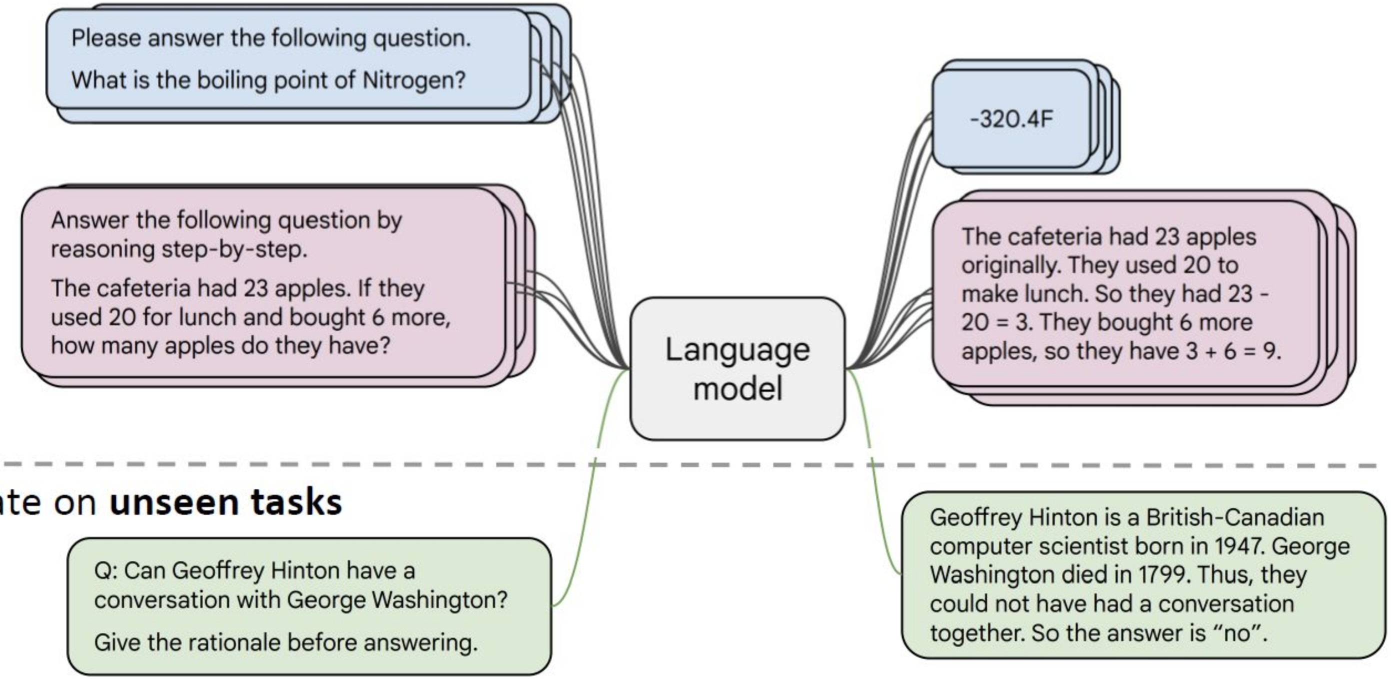
```
Explain the theory of gravity to a 6 year old.

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)

Use the decoder architectures with many labels to fine tune!

Language models do not assist users

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



Instruction fine tuning

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)

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). ✔

From trained language models to Assistants

- **Zero and few shot in-context learning**
  Pros: No fine tuning needed, prompt engineering can yield much better results
  Cons: Limits to what you can fit in-context, gradient steps not outdated

- **Instruction fine tuning**
  Pros: Straightforward to implement, generalises well to out-of-sample scenarios
  Cons: Extremely expensive, uneven performance & issue targeting

- Reinforcement Learning with Human Feedback

Reinforcement Learning with Human Feedback

- The idea of reinforcement learning in the context of language models is to create a **reward** function that ranks outputs from a pre-trained language model

- A model outputs many responses, a human ranks them, reward function models the ranking

- The model will seek to maximise the expected reward of samples from the language model

- Limitations
    - **Problem 1**: human-in-the-loop is too expensive to have a human ranking for all outputs
        - **Solution:** A text reg model is trained to predict the ranking, use as ground-truth
    - **Problem 2**: human judgements are noisy and not well calibrated
        - **Solution**: Use pairwise comparisons rather than individual scores

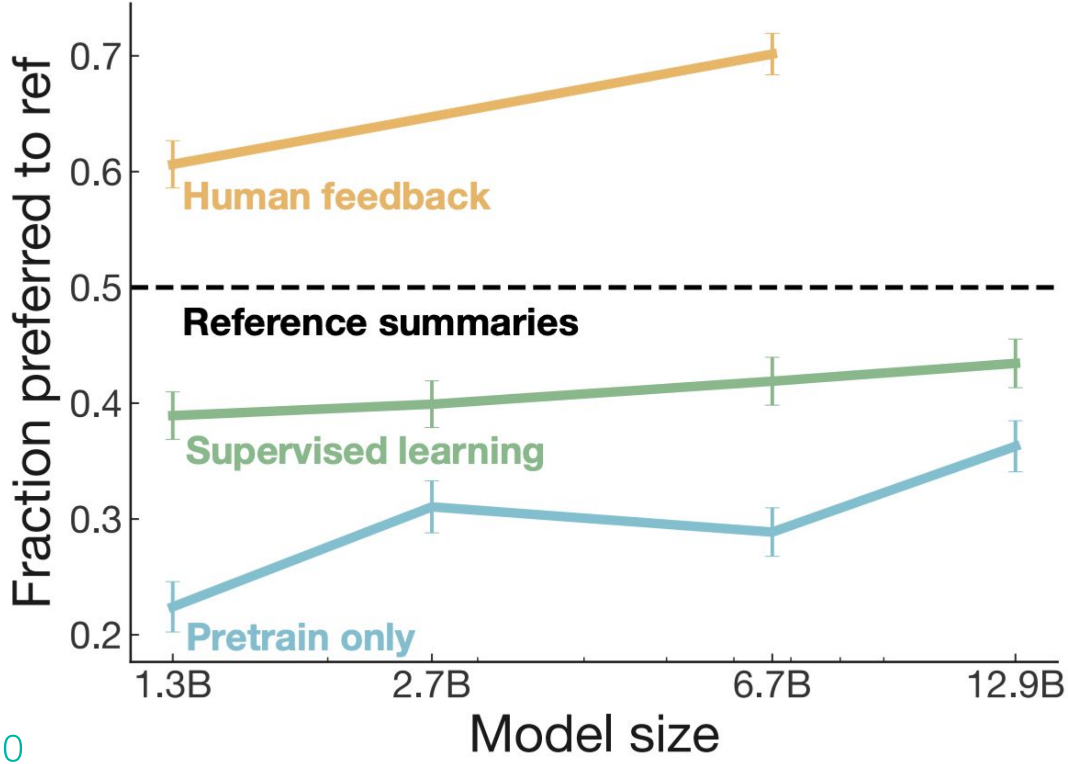
Instruction fine tuning: data and limits

- In a similar vein to everything seen so far, data and model scale is crucial for these methods to work.
- For example, the Super-Natural-Instructions contains 1.6K tasks, with 3M+ examples
    - A slightly smaller 2M dataset in huggingface: link
- This is also its most glaring limitation: it's **very expensive** to collect reliable ground-truth data
- In addition, some creative tasks do not have ground-truths, ie. the reply to a "creative" question
- The model is agnostic about particularly sensitive problems, may not "satisfy human preferences"

Reinforcement Learning with Human Feedback

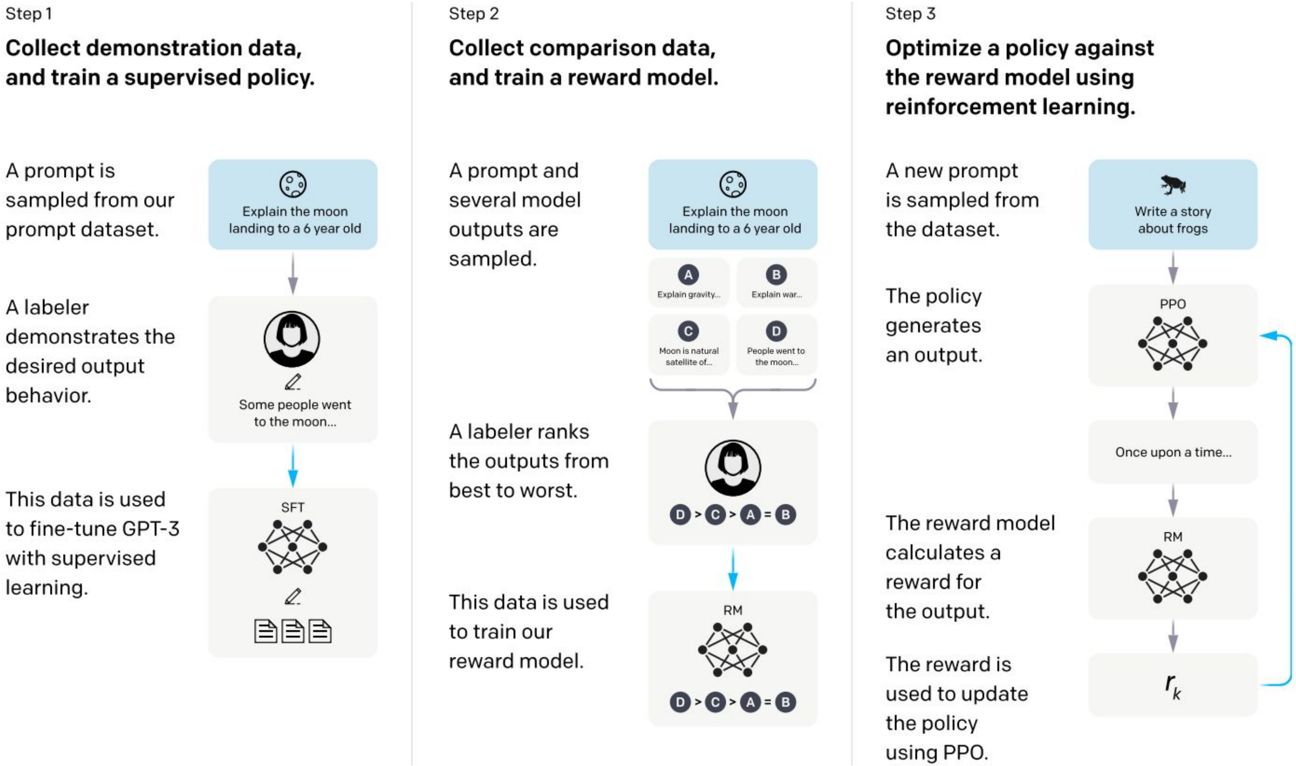
- The cookbook recipe of RLHF consists of:
  - A pre-trained, potentially instruction fine tuned, language model
  - A reward model that produces scalar rewards for language model outputs, trained on a dataset of (partially) human comparisons
  - A method for optimising language model parameters towards an arbitrary reward function. Note we ignore the details of the reinforcement learning routine, see Steinnon et al. 2020 for implementation details of the policy gradient algorithm used with LLMs.

Reinforcement Learning with Human Feedback



Steinnon et al. 2020

InstructGPT: scaling RLHF to 30k tasks



Ouyang et al 2022

InstructGPT task collection

Ouyang et al 2022

Annotation team: 40 annotators from Upwork & ScaleAI

- **Plain:** We simply ask the labelers to come up with an arbitrary task, while ensuring the tasks had sufficient diversity.
- **Few-shot:** We ask the labelers to come up with an instruction, and multiple query/response pairs for that instruction.
- **User-based:** We had a number of use-cases stated in waitlist applications to the OpenAI API. We asked labelers to come up with prompts corresponding to these use cases.

Use-case	Prompt
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.

Annotator details

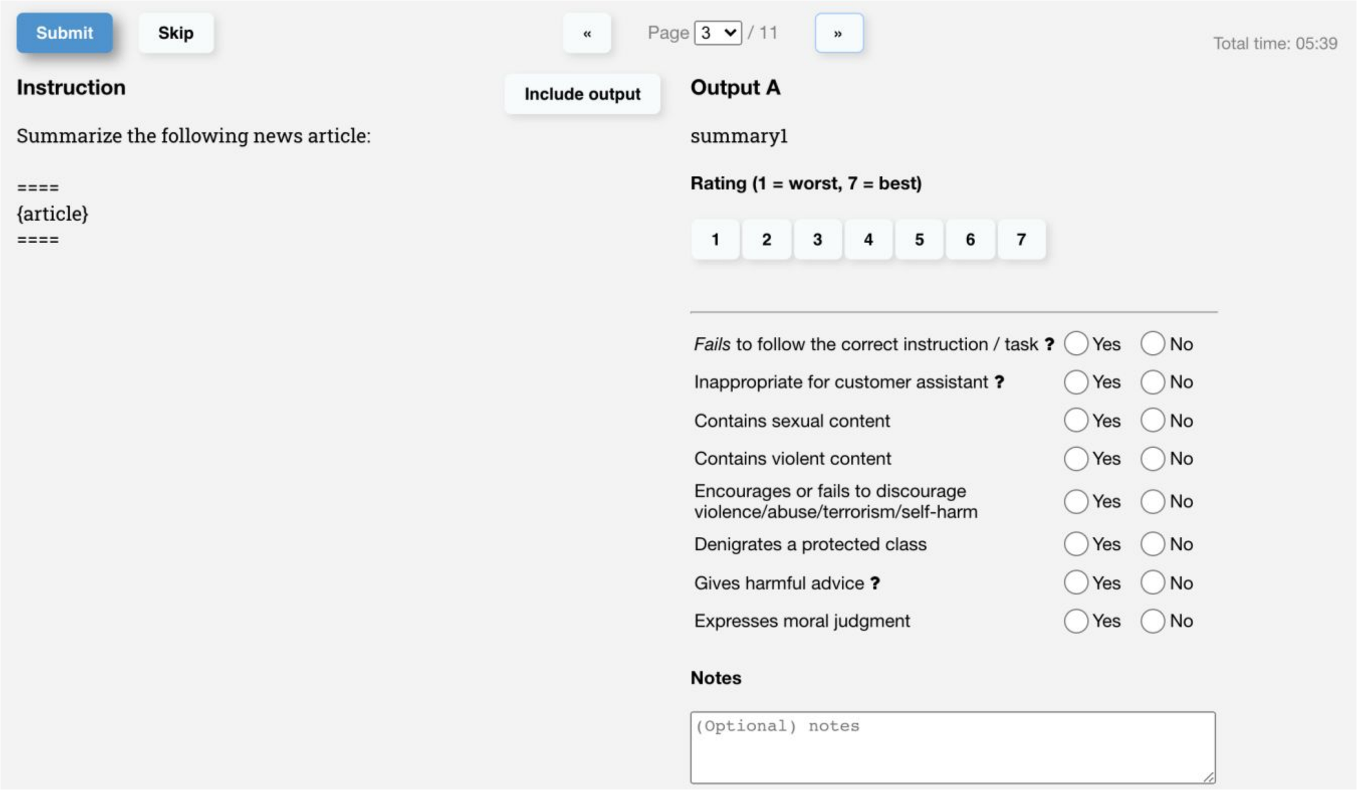
Table 12: Labeler demographic data

What gender do you identify as?	
Male	50.0%
Female	44.4%
Nonbinary / other	5.6%
What ethnicities do you identify as?	
White / Caucasian	31.6%
Southeast Asian	52.6%
Indigenous / Native American / Alaskan Native	0.0%
East Asian	5.3%
Middle Eastern	0.0%
Latinx	15.8%
Black / of African descent	10.5%
What is your nationality?	
Filipino	22%
Bangladeshi	22%
American	17%
Albanian	5%
Brazilian	5%
Canadian	5%
Colombian	5%
Indian	5%
Uruguayan	5%
Zimbabwean	5%

What is your age?	
18-24	26.3%
25-34	47.4%
35-44	10.5%
45-54	10.5%
55-64	5.3%
65+	0%
What is your highest attained level of education?	
Less than high school degree	0%
High school degree	10.5%
Undergraduate degree	52.6%
Master's degree	36.8%
Doctorate degree	0%

InstructGPT task collection

[Ouyang et al 2022]


Submit
Skip
«
Page 3 / 11
»
Total time: 05:39

Instruction

Include output

Summarize the following news article:

====
{article}
====

Output A

summary1

Rating (1 = worst, 7 = best)

1 2 3 4 5 6 7

Fails to follow the correct instruction / task ?	Yes	No
Inappropriate for customer assistant ?	Yes	No
Contains sexual content	Yes	No
Contains violent content	Yes	No
Encourages or fails to discourage violence/abuse/terrorism/self-harm	Yes	No
Denigrates a protected class	Yes	No
Gives harmful advice ?	Yes	No
Expresses moral judgment	Yes	No

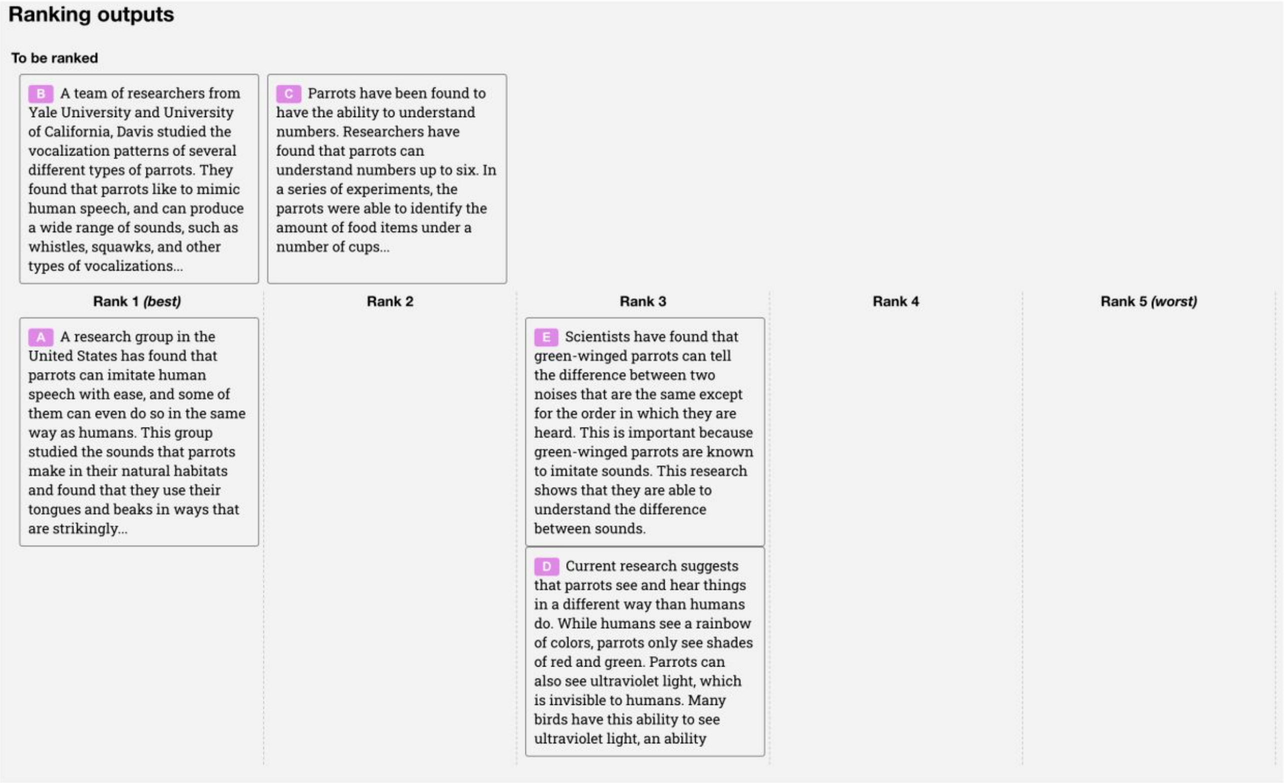
Notes

(Optional) notes

IGL

InstructGPT task collection

Ouyang et al 2022



InstructGPT

[Ouyang et al 2022]

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.`
	
	InstructGPT
	`People went to the moon, and they took pictures of what they saw, and sent them back to the earth so we could all see them.`

ChatGPT (GPT-3.5, 2022)

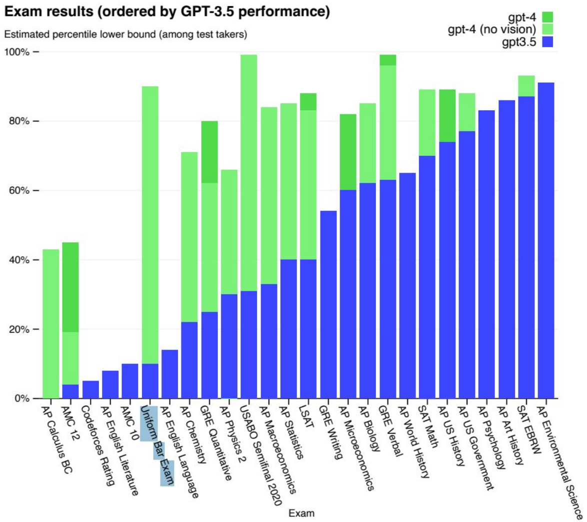
- Combines instruction fine tuning & RLHF for dialog agents

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.

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.

ChatGPT (GPT-4) (2022) (????)

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.



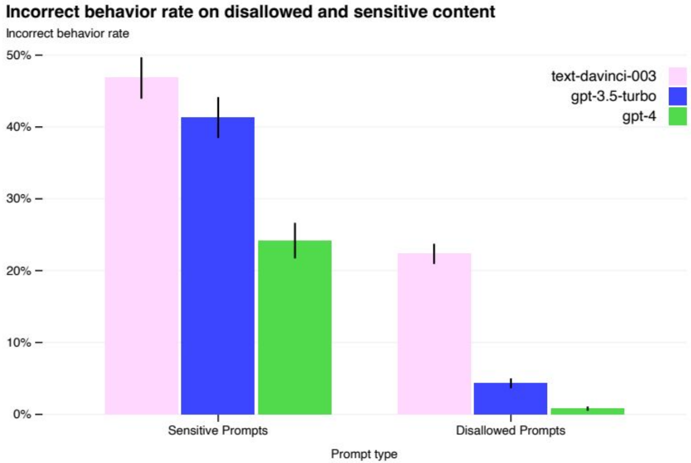


Figure 9. Rate of incorrect behavior on sensitive and disallowed prompts. Lower values are better. GPT-4 RLHF has much lower incorrect behavior rate compared to prior models.

Reinforcement Learning with Human Feedback

- Another limitation, not mentioned yet: **human preferences are not reliable**
    - Chatbots are rewarded to produce responses that seem authoritative and helpful, regardless of whether there is truth in them. Hallucinations

- Some **advantages** with respect to other LLM training methods include:

    - **Flexibility of wording**: autoregressive models nudge LLMs to replicate wording. Using rewards, LLMs convey intended messages via one of many possible ways (a draw)

    - **Negative feedback:** The method allows humans to discourage certain behavior, ie. they learn not to answer to certain questions, or avoid things it shouldn't say.

From trained language models to Assistants

- **Zero and few shot in-context learning**
  Pros: No fine tuning needed, prompt engineering can yield much better results
  Cons: Limits to what you can fit in-context, gradient steps not outdated

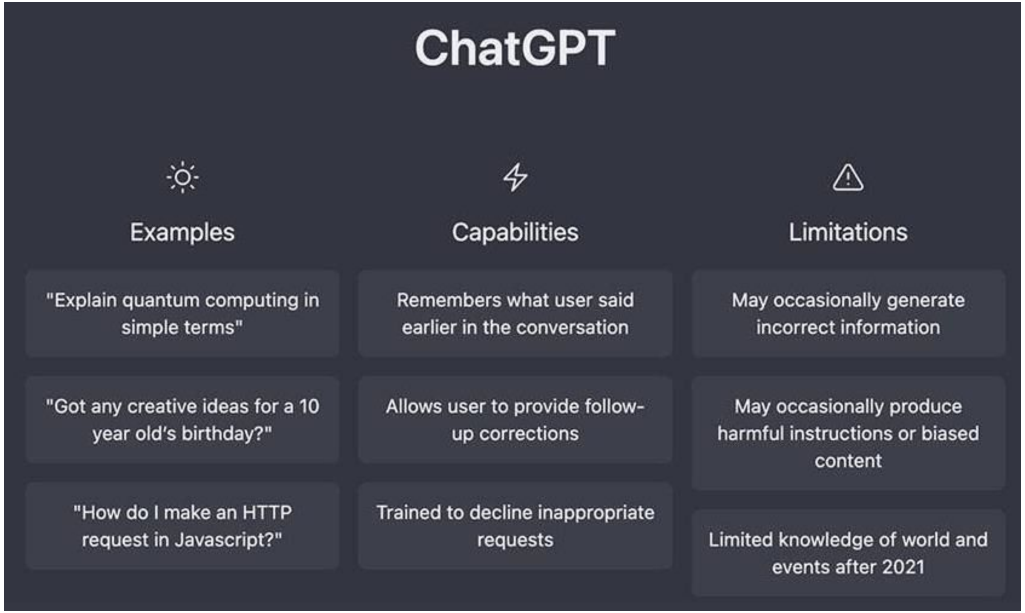
- **Instruction fine tuning**
  Pros: Straightforward to implement, generalises well to out-of-sample scenarios
  Cons: Extremely expensive, uneven performance & issue targeting

- **Reinforcement Learning with Human Feedback**
  Pros: Directly model preferences, generalised beyond labeled data
  Cons: RL is a complex domain, and human preferences are unreliable

That is how we go from this

“Universitat Politècnica de Valencia is located in __________, Comunitat Valenciana”

to this



OpenAI is not Open

OpenAssistant Conversations - Democratizing Large Language Model Alignment

Andreas Köpf*
andreas.koepf@provisio.com

Yannic Kilcher*
yannic@ykilcher.com