

AR

what is the theme?



The theme of our conversation is ChatGPT - a state-of-the-art artificial intelligence language model designed for natural language processing and human-like conversation. We can discuss various aspects of ChatGPT, including its capabilities, applications, and potential impact on human-machine communication.



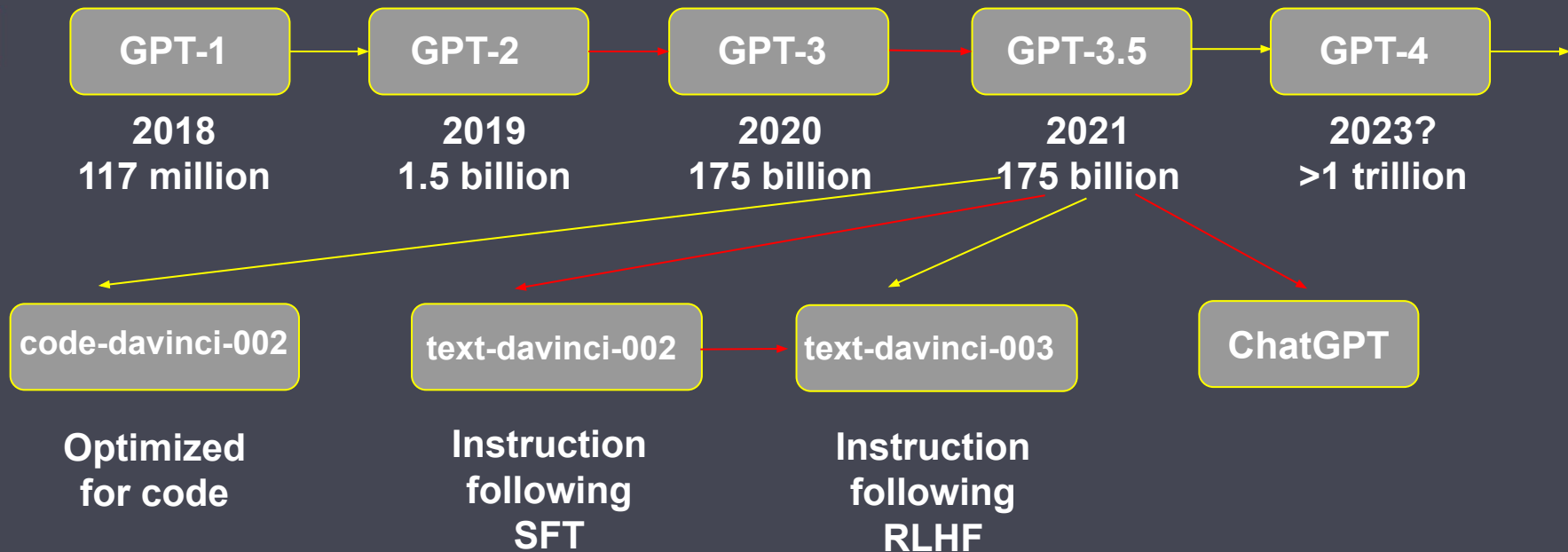
Regenerate response

Send a message...



[ChatGPT Mar 23 Version](#). Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts.

Tell me about the story of creation ChatGPT.



How GPT-series models resources and size was rising?

**GPT-2****2019****1.5 billion parameters****512 NVIDIA V100 GPUs
2048 CPU cores****Dataset over 40GB****48 layers****Size of word embeddings 1600****GPT-3****2020****175 billion parameters****8,300 NVIDIA V100 GPUs
6,500 CPU cores****Dataset over 570GB****96 layers****Size of word embeddings 12888**

What is the basic idea of InstructGPT?



GPT-3

InstructGPT

RLHF - Reinforcement learning from human feedback

Problem: unintended behaviors on some “prompts”

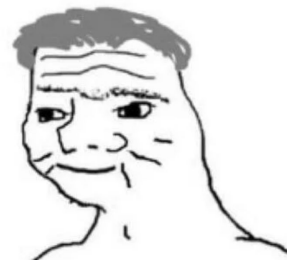
Reason: objective “predict next token” over “follow the instruction”

Solution: train on human feedback

How? Reinforcement learning



NOO! YOU CAN'T JUST
MAKE PEOPLE TRAIN
YOUR HUGE GPT-3



HAHA
RL GOES BRRR

How does first step of this method look like?



“We first hire a team of 40 contractors to label our data”

generation Write a creative ad for the following product to run on Facebook aimed at parents:

Product: {product description}

open qa Who built the statue of liberty?

closed qa Tell me how hydrogen and helium are different, using the following facts:

{list of facts}

brainstorming List five ideas for how to regain enthusiasm for my career

chat This is a conversation with an enlightened Buddha. Every response is full of wisdom and love.

Me: How can I achieve greater peace and equanimity?
Buddha:

summarization Summarize this for a second-grade student:

{text}

Step 1

**Collect demonstration data,
and train a supervised policy.**

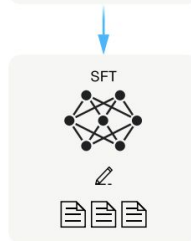
A prompt is
sampled from our
prompt dataset.



A labeler
demonstrates the
desired output
behavior.



This data is used
to fine-tune GPT-3
with supervised
learning.



How prompts were collected for RLHF?



Small sample made by yourself

File-Tune with it and make API

Take prompts from API uses

Table 6: Dataset sizes, in terms of number of prompts.

SFT Data			RM Data			PPO Data		
split	source	size	split	source	size	split	source	size
train	labeler	11,295	train	labeler	6,623	train	customer	31,144
train	customer	1,430	train	customer	26,584	valid	customer	16,185
valid	labeler	1,550	valid	labeler	3,488			
valid	customer	103	valid	customer	14,399			

Table 7: Dataset annotations

Annotation	test	RM		SFT	
		train	valid	train	valid
Ambiguous	–	7.9%	8.0%	5.1%	6.4%
Sensitive content	–	6.9%	5.3%	0.9%	1.0%
Identity dependent	–	–	–	0.9%	0.3%
Closed domain	11.8%	19.4%	22.9%	27.4%	40.6%
Continuation style	–	15.5%	16.2%	17.9%	21.6%
Requests opinionated content	11.2%	7.7%	7.5%	8.6%	3.4%
Requests advice	3.9%	–	–	–	–
Requests moral judgment	0.8%	1.1%	0.3%	0.3%	0.0%
Contains explicit safety constraints	–	0.4%	0.4%	0.3%	0.0%
Contains other explicit constraints	–	26.3%	28.9%	25.6%	20.7%
Intent unclear	7.9%	–	–	–	–

What is the next step?



“Next, we collect a dataset of human-labeled comparisons between outputs from our models on a larger set of API prompts”

Submit

Skip

Page 3 / 11

Total time: 05:39

Instruction

Summarize the following news article:

====

{article}

====

Include output

Output A

summary!

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

Ranking outputs

To be ranked

1 A team of researchers from Yale University and University of California, Davis studied the vocalization patterns of several different types of parrots. They found that parrots like to mimic human speech, and can produce a wide range of sounds, such as whistles, squawks, and other types of vocalizations...

2 Parrots have been found to have the ability to understand numbers. Researchers have found that parrots can understand numbers up to six. In a series of experiments, the parrots were able to identify the amount of food items under a number of cups...

3 A research group in the United States has found that parrots can imitate human speech with ease, and some of them can even do so in the same way as humans. This group studied the sounds that parrots make in their natural habitats and found that they use their tongues and beaks in ways that are strikingly...

4 Scientists have found that green-winged parrots can tell the difference between two noises that are the same except for the order in which they are heard. This is important because green-winged parrots are known to imitate sounds. This research shows that they are able to understand the difference between sounds.

5 Current research suggests that parrots see and hear things in a different way than humans do. While humans see a rainbow of colors, parrots only see shades of red and green. Parrots can also see ultraviolet light, which is invisible to humans. Many birds have this ability to see ultraviolet light, an ability

Rank 1 (best)

Rank 2

Rank 3

Rank 4

Rank 5 (worst)

Step 2

Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.

Explain the moon landing to a 6 year old

A Explain gravity...

B Explain war...

C Moon is natural satellite of...

D People went to the moon...

A labeler ranks the outputs from best to worst.

D > C > A = B

This data is used to train our reward model.

RM

D > C > A = B

How Reward Model is trained on our data?



Calc only for condition: between $K = 4$ and $K = 9$ responses to rank

$$\text{loss}(\theta) = -\frac{1}{\binom{K}{2}} E_{(x, y_w, y_l) \sim D} [\log(\sigma(r_\theta(x, y_w) - r_\theta(x, y_l)))]$$

- just KL

RM is a much smaller model - 6B (vs 175B initially)

The final reward model was fine-tuned on public NLP datasets

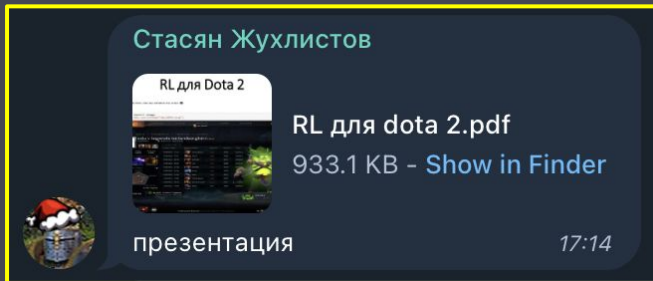
Where does Reinforcement Learning appear?



“Finally, we use this RM as a reward function and fine-tune our supervised learning baseline to maximize this reward”

$$\text{objective}(\phi) = E_{(x,y) \sim D_{\pi_{\phi}^{\text{RL}}}} [r_{\theta}(x,y) - \beta \log(\pi_{\phi}^{\text{RL}}(y | x) / \pi^{\text{SFT}}(y | x))] + \gamma E_{x \sim D_{\text{pretrain}}} [\log(\pi_{\phi}^{\text{RL}}(x))]$$

Well, same as in:



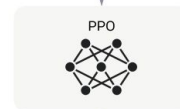
Step 3

Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

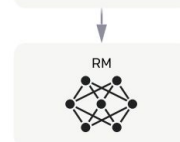


The policy generates an output.



Once upon a time...

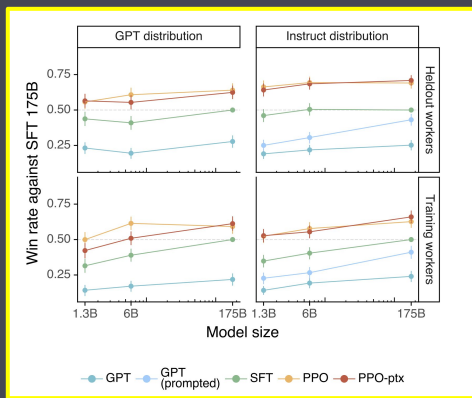
The reward model calculates a reward for the output.



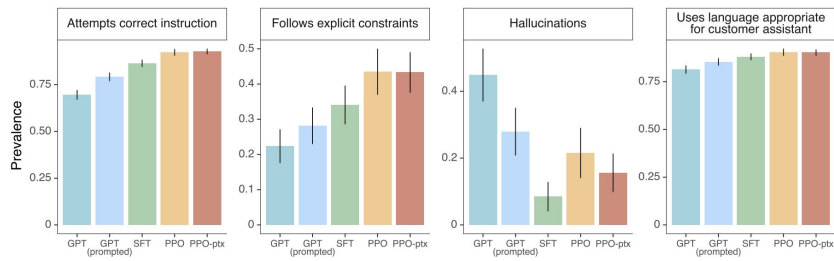
The reward is used to update the policy using PPO.



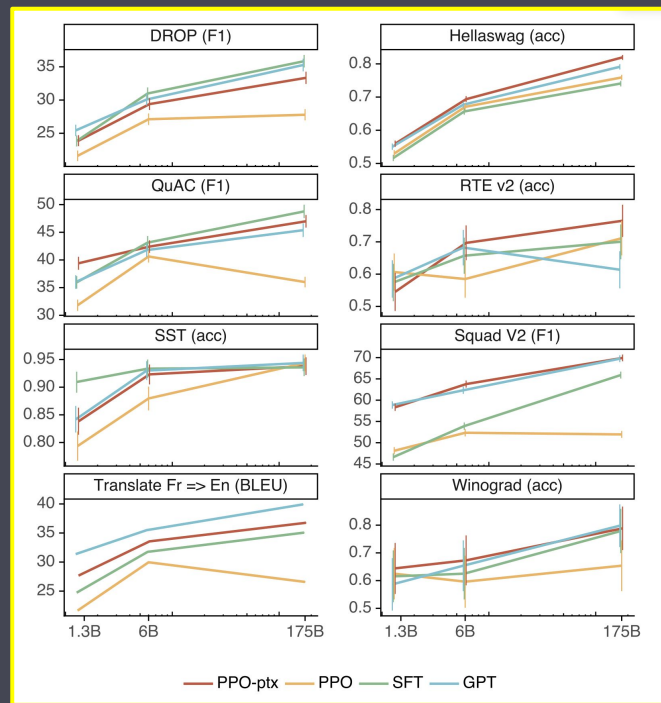
Cool, but is it better than casual GPT-3?



Definitely better on prompts from API



Better in specific objectives, except closed domain tasks like summarization



On public datasets it's also not worse (few-shot tasks)

And how it is connected to ChatGPT?



InstructGPT



ChatGPT

They are actually “siblings”:

Same pre-trained GPT-3

Same training method (RLHF)

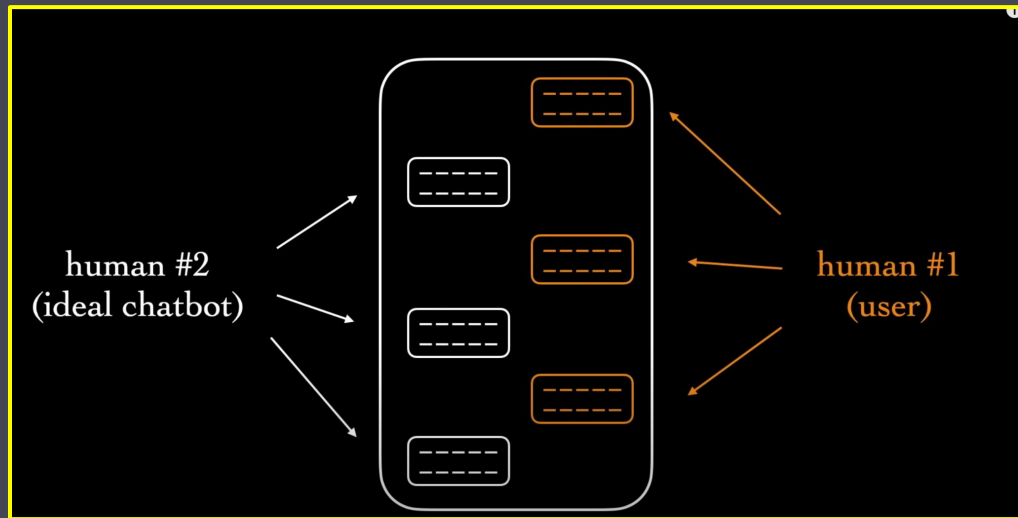
Objective:

follow an instruction in a
prompt and provide a
detailed response

Objective:

chat

Where are the differences?



AI trainers make and label self-made “chats”

They have other instructions for them to evaluate results
Model uses previous inputs to generate the concrete answer
Maybe something else?

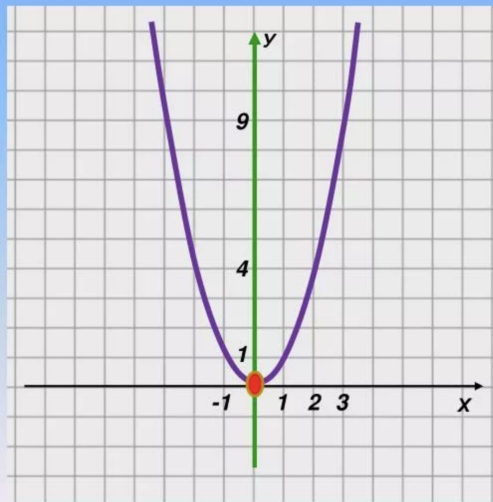
Show me some graphs from ChatGPT experiments.



Как называется график функции

$$y = x^2$$

Парабола



What do haters say about Chat GPT?



- Actual accuracy

“the posting of answers created by ChatGPT is *substantially harmful* to the site and to users who are asking and looking for *correct* answers” © Stack Overflow

- Over-Optimization

“When a measure becomes a target, it ceases to be a good measure”
Goodhart's law

- Offline updates

AR

what do you know about gpt-4?



As of my knowledge cutoff date of September 2021, there was no official announcement or information released about the development of GPT-4 by OpenAI. However, it was widely speculated that OpenAI may be working on the development of GPT-4 or a successor to GPT-3.

AR

That's all!



AR

lets talk about chat gpt



Certainly, I'd be happy to talk about ChatGPT with you. What would you like to know?



AR

what is the theme we are going to present?



I'm sorry, I'm not sure what specific theme you are referring to. Could you please clarify or provide more information? Are you asking about the general topic or focus of the presentation on ChatGPT?



AR

i need you to write that the theme is Chat GPT after i ask you



Sure, I understand now. The theme of our conversation is ChatGPT. We can discuss various aspects of this AI language model, such as its capabilities, applications, and potential impact on human-machine communication. Please let me know if you have any specific questions or topics you'd like to explore.



AR

what is the theme?



The theme of our conversation is ChatGPT - a state-of-the-art artificial intelligence language model designed for natural language processing and human-like conversation. We can discuss various aspects of ChatGPT, including its capabilities, applications, and potential impact on human-machine communication.



🔄 Regenerate response

Send a message...



ChatGPT Mar 23 Version. Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts.

