

# The Road to ChatGPT

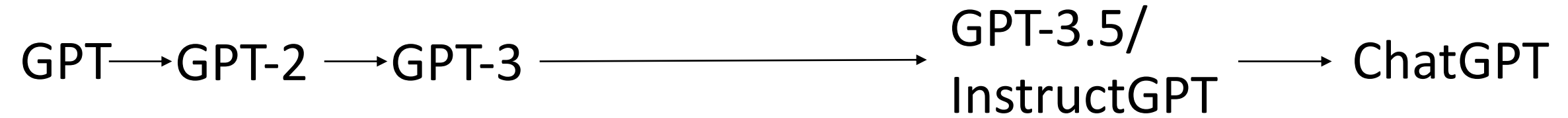
An informal explainer on how ChatGPT was built

*Rama Ramakrishnan*

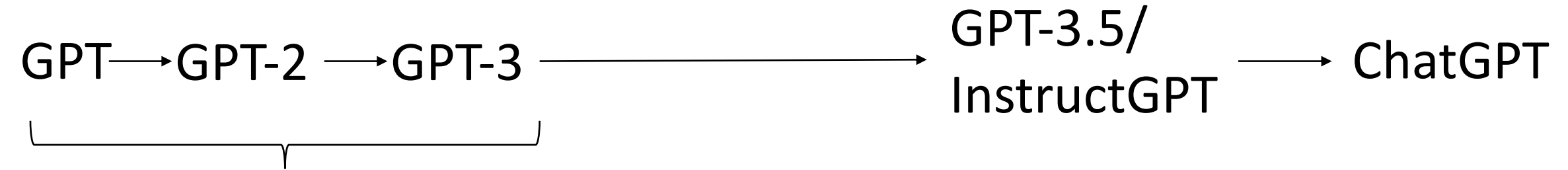
*MIT Sloan School of Management*

*March 5, 2023*

# The road to ChatGPT



# The road to ChatGPT



We will first look at how GPT-3 was built.

GPT and GPT-2 were built in a similar way; we will highlight the differences later

# Caveat

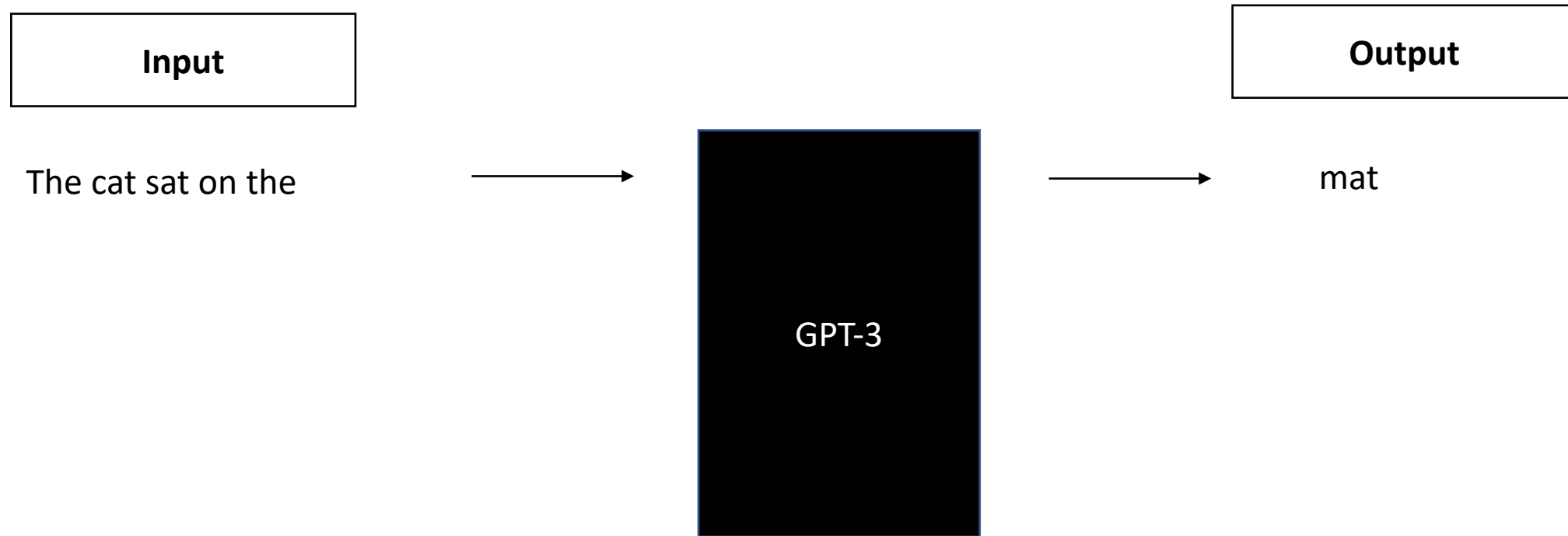
To distill/make-intuitive the key ideas, I have stripped out many details. If you want to know exactly how things work, please see the links to the technical papers and blog posts referenced in the deck.

# GPT-3 is a mathematical model\* ...

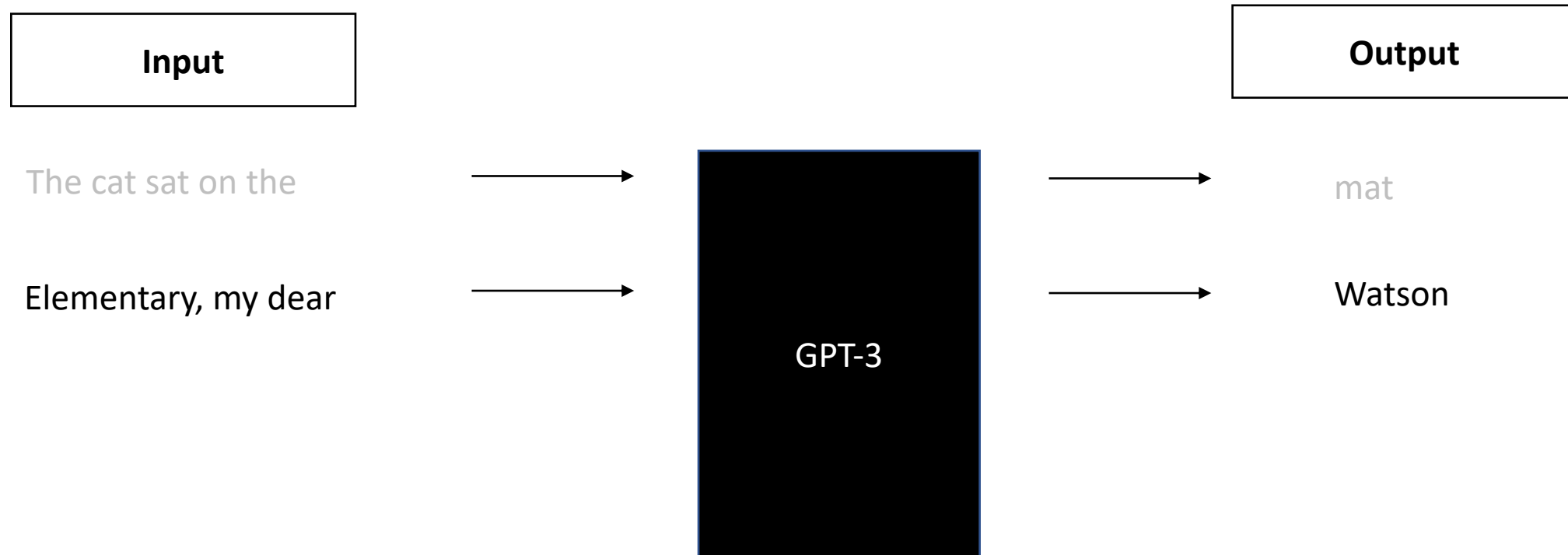


GPT-3

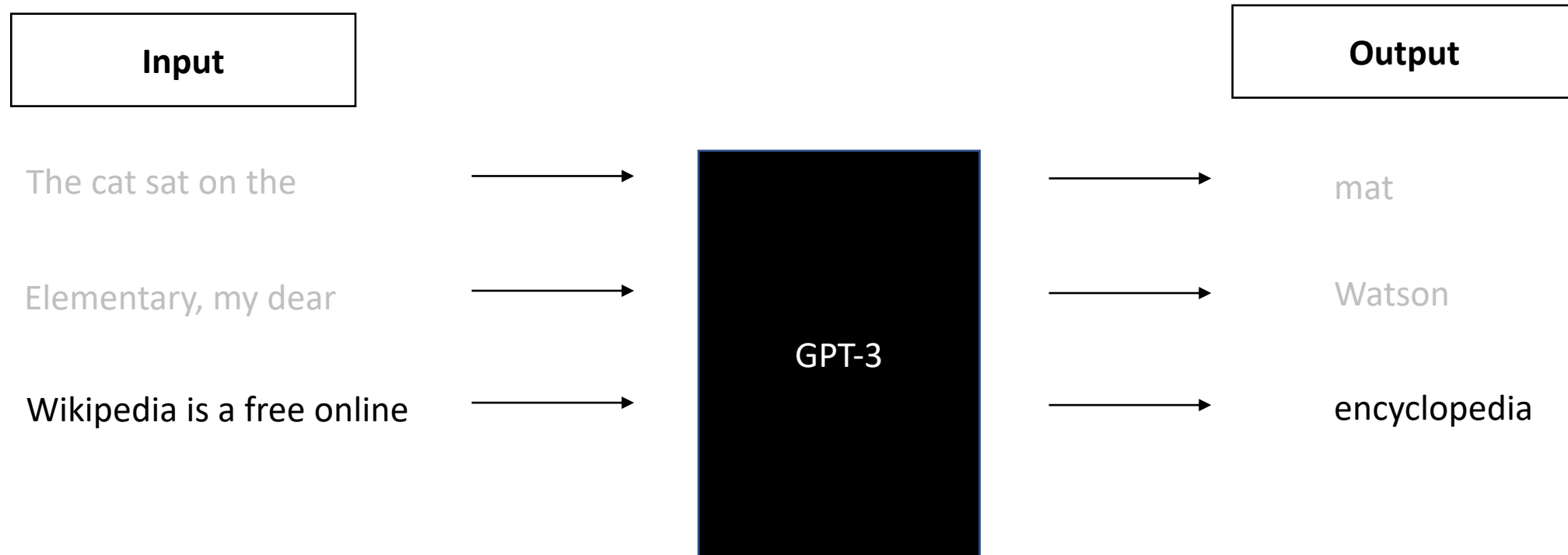
... trained to *predict the next word* in a sentence,  
using the previous words



... trained to *predict the next word* in a sentence,  
using the previous words

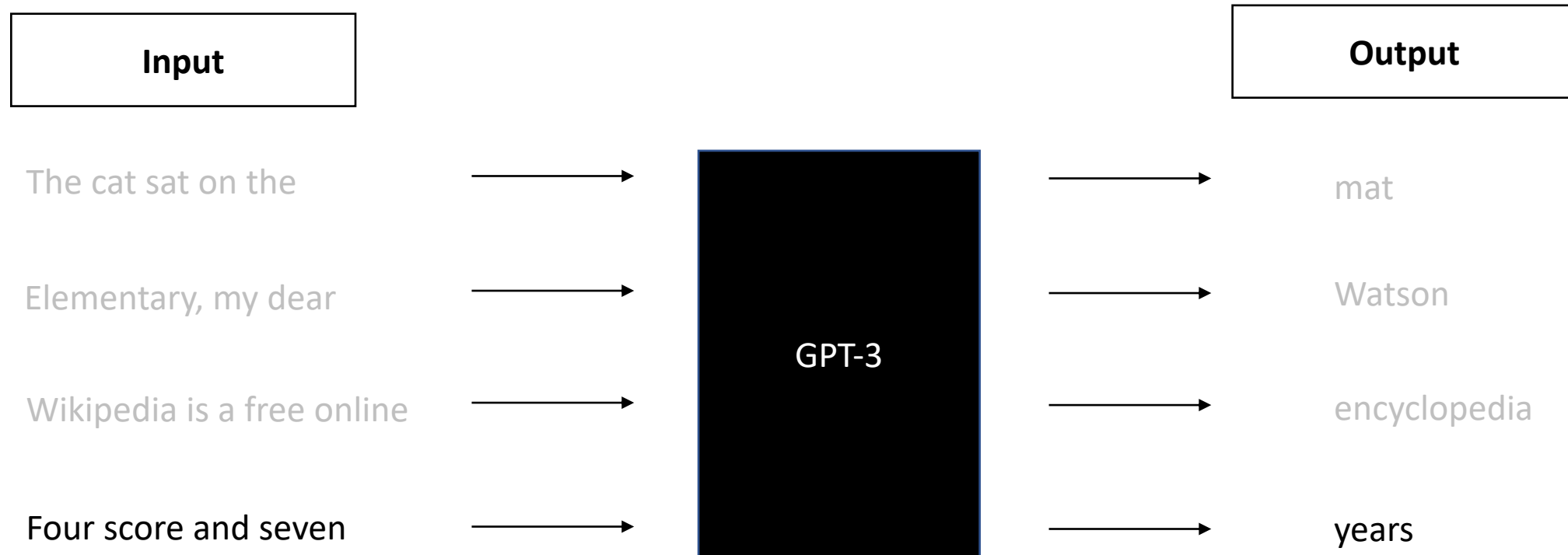


... trained to *predict the next word* in a sentence,  
using the previous words





... trained to *predict the next word* in a sentence,  
using the previous words



# These training sentences were extracted from the Internet and books

Dataset	Quantity (tokens)
Common Crawl (filtered)	410 billion
WebText2	19 billion
Books1	12 billion
Books2	55 billion
Wikipedia	3 billion

<http://arxiv.org/abs/2005.14165>

# How many sentences?

Dataset	Quantity (tokens)
Common Crawl (filtered)	410 billion
WebText2	19 billion
Books1	12 billion
Books2	55 billion
Wikipedia	3 billion

<http://arxiv.org/abs/2005.14165>

I couldn't find this in the GPT-3 paper so here's a back-of-the-envelope calc:

According to the table on the left, ~500 billion "tokens" were used.

(Since we are doing a rough estimate), *we can think of a 'token' as a word. So, 500 billion words were used.*

Google says that there are 15-20 words per sentence on average.

So, 500 billion/15-to-20 = **25-33 billion sentences.**

After training, GPT-3 can do something very useful

# Given any text as input ...

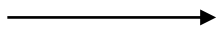
The cat sat on the



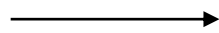
GPT-3

... GPT-3 can calculate this table

The cat sat on the



GPT-3



Next Word	Probability ("how often")
aardvark	0.0
...	
fridge	0.05
...	
mat	0.2
...	
...	
zebra	0.0

# What does this table contain?

The table shows, for each word in the dictionary, approximately what % of the time that word appears in the training data right after the input phrase\*

	Next Word	Probability ("how often")
The cat sat on the	aardvark	0.0
	...	
The cat sat on the	fridge	0.05
	...	
The cat sat on the	mat	0.2
	...	
	...	
The cat sat on the	zebra	0.0

\*Intuitive, NOT technically accurate, description. Forgive me, technical folks 🙏

# Interpreting the table

	Next Word	Probability ("how often")
The cat sat on the	aardvark	0.0
	...	
The cat sat on the	fridge	0.05
	...	
The cat sat on the	mat	0.2
	...	
	...	
The cat sat on the	zebra	0.0

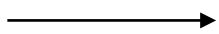
## Interpretation:

For all the sentences that started with  
"The cat sat on the", the next word was  
"mat" 20% of the time

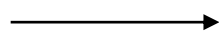


# OK, what can we do with this table?

The cat sat on the



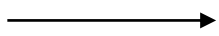
GPT-3



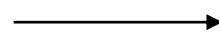
Next Word	Probability ("how often")
aardvark	0.0
...	
fridge	0.05
...	
mat	0.2
...	
...	
zebra	0.0

# We can “*auto complete*” any input sentence!

The cat sat on the

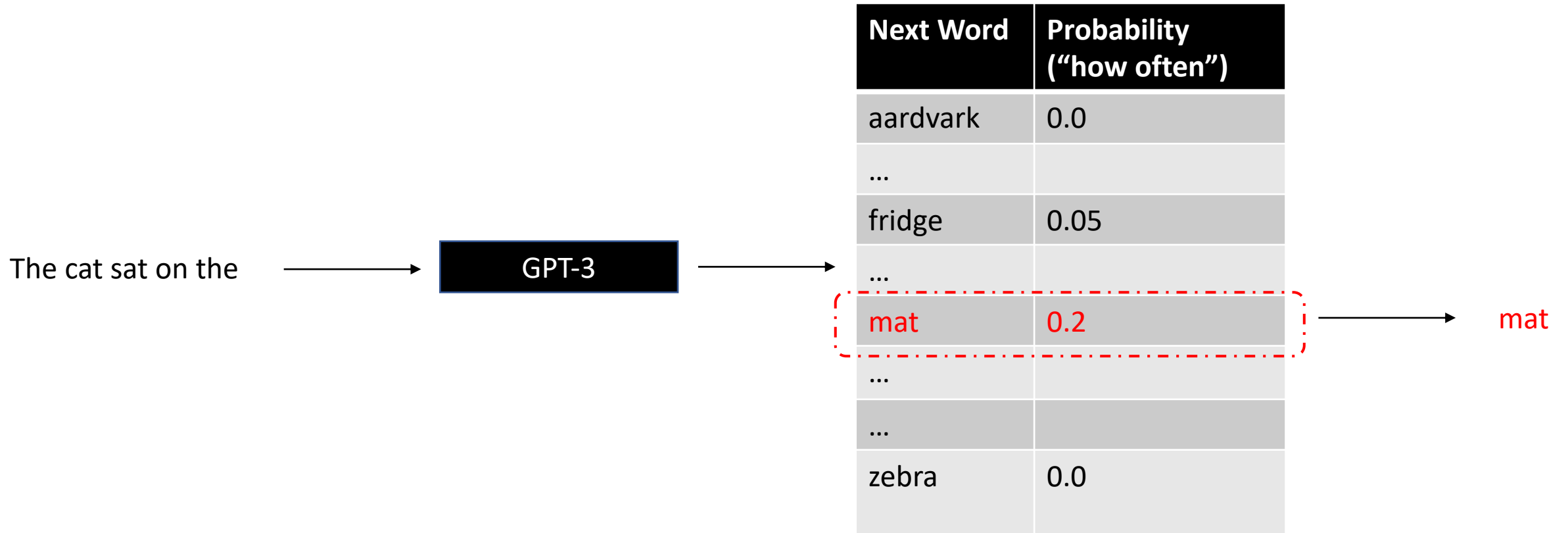


GPT-3

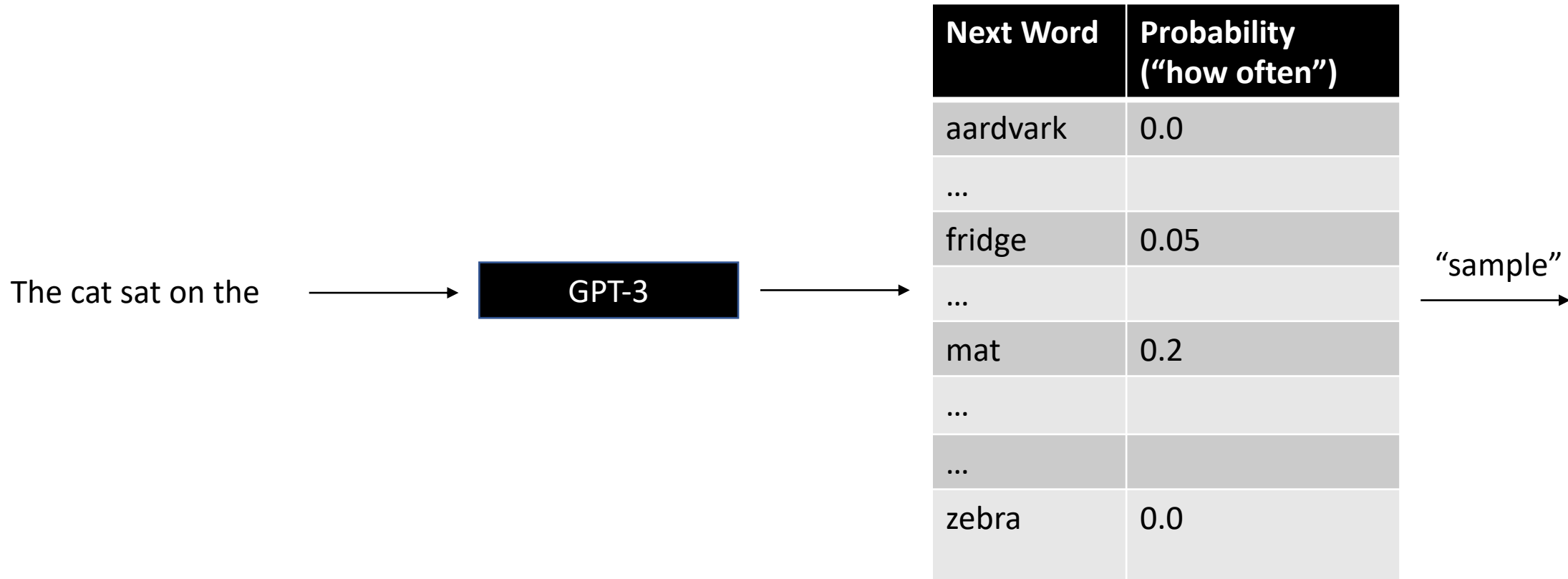


Next Word	Probability ("how often")
aardvark	0.0
...	
fridge	0.05
...	
mat	0.2
...	
...	
zebra	0.0

We can tell GPT-3 to pick the most-frequent word (i.e., the highest probability word) from this table and output it ...



# Or we can ask GPT-3 to “sample” the next word and output that



# What is sampling?

Imagine you have a six-sided die. If you roll the die, each of the 6 sides is equally likely to show up.

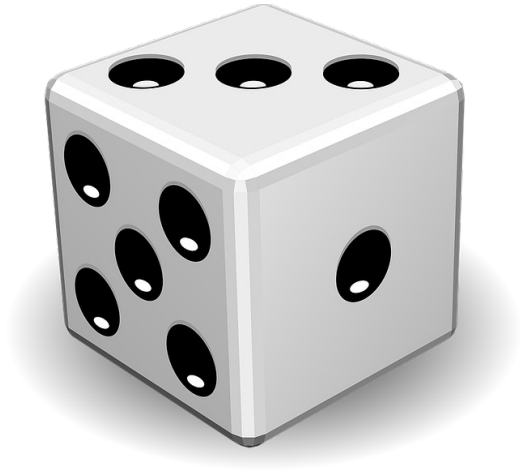


Image by [OpenClipart-Vectors](#) from [Pixabay](#)

# What is sampling?

Imagine you have a six-sided die. If you roll the die, each of the 6 sides is equally likely to show up. This is captured in this table.

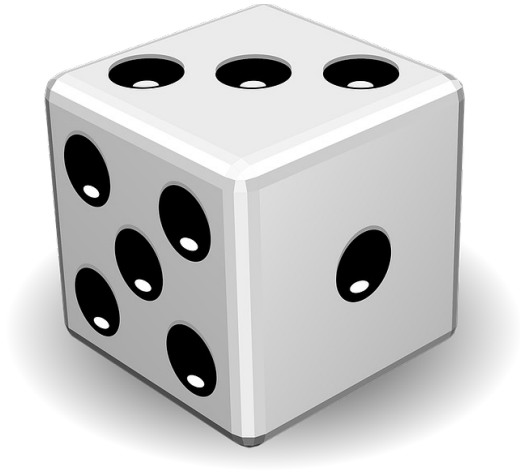


Image by [OpenClipart-Vectors](#) from [Pixabay](#)

↓

Side	Probability*
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$

\*note that the probabilities add to 1.0

Now, imagine you have a *weighted* die. Each of the 6 sides is not equally likely. The table below shows the probabilities



<https://www.etsy.com/listing/1202528855/weighted-dice-set-koplow-games>

Side	Probability
1	0.2
2	0.01
3	0.3
4	0.4
5	0.04
6	0.05

# Imagine rolling this weighted dice once



<https://www.etsy.com/listing/1202528855/weighted-dice-set-koplow-games>

Side	Probability
1	0.2
2	0.01
3	0.3
4	0.4
5	0.04
6	0.05



# Imagine rolling this weighted dice once



<https://www.etsy.com/listing/1202528855/weighted-dice-set-koplow-games>

Side	Probability
1	0.2
2	0.01
3	0.3
4	0.4
5	0.04
6	0.05

*You are most likely to get a 4 (since that's the highest probability) ...*

# Imagine rolling this weighted dice once



<https://www.etsy.com/listing/1202528855/weighted-dice-set-koplow-games>

Side	Probability
1	0.2
2	0.01
3	0.3
4	0.4
5	0.04
6	0.05

*You are most likely to get a 4 (since that's the highest probability) ...*

*but you may get a 3 pretty often ...*

# Imagine rolling this weighted dice once



<https://www.etsy.com/listing/1202528855/weighted-dice-set-koplow-games>

Side	Probability
1	0.2
2	0.01
3	0.3
4	0.4
5	0.04
6	0.05

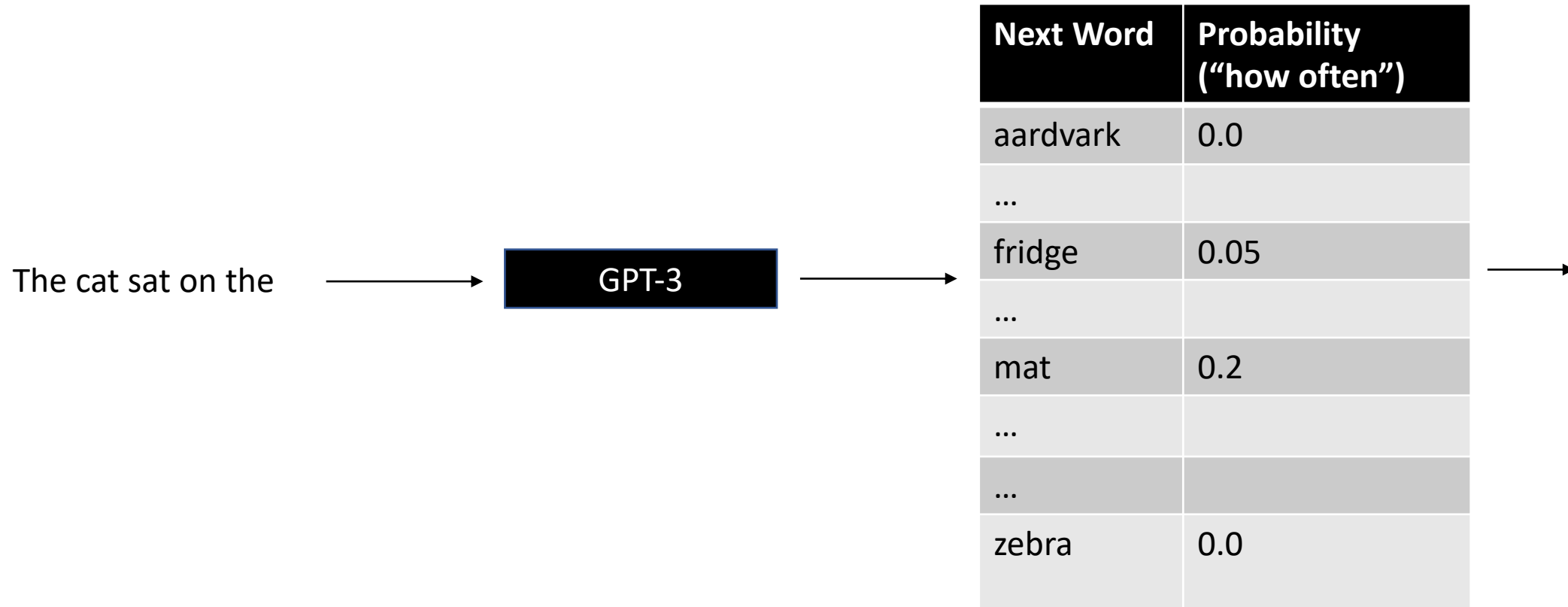
*You are most likely to get a 4 (since that's the highest probability) ...*

*but you may get a 3 pretty often ...*

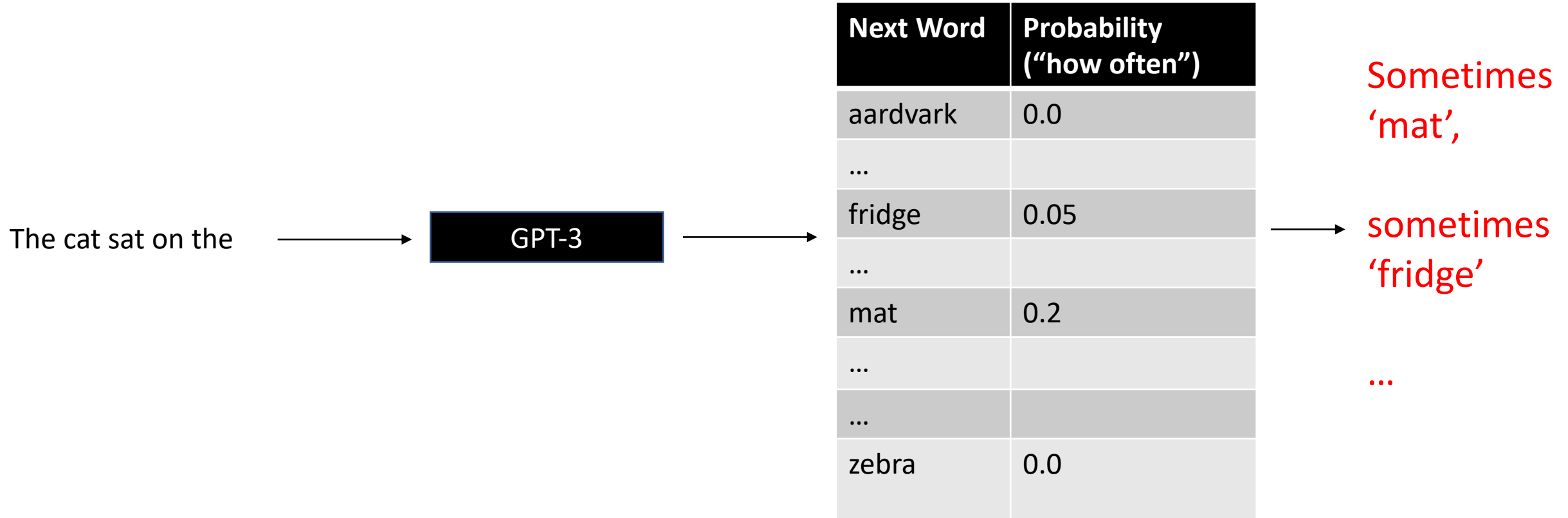
*and once in a rare while you may even get a 2!*

Sampling a word from GPT-3's table =  
Rolling a weighted die once

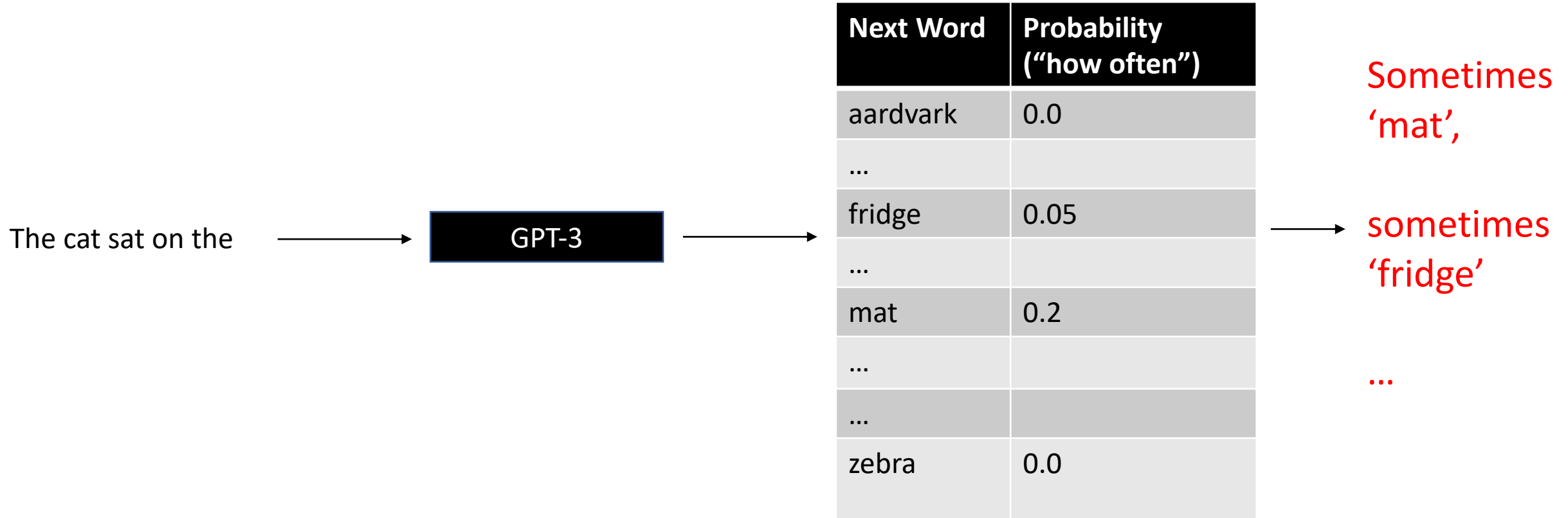
# We can “roll” (i.e., sample) this table to get the next word



# Each time we “sample”, we may get a different word



BTW, this is what gives GPT-3 the ability to generate multiple outputs for the same input



We have seen how GPT-3 can generate the *next word* from your single sentence input.

Next, let's see how it can generate *paragraphs*.



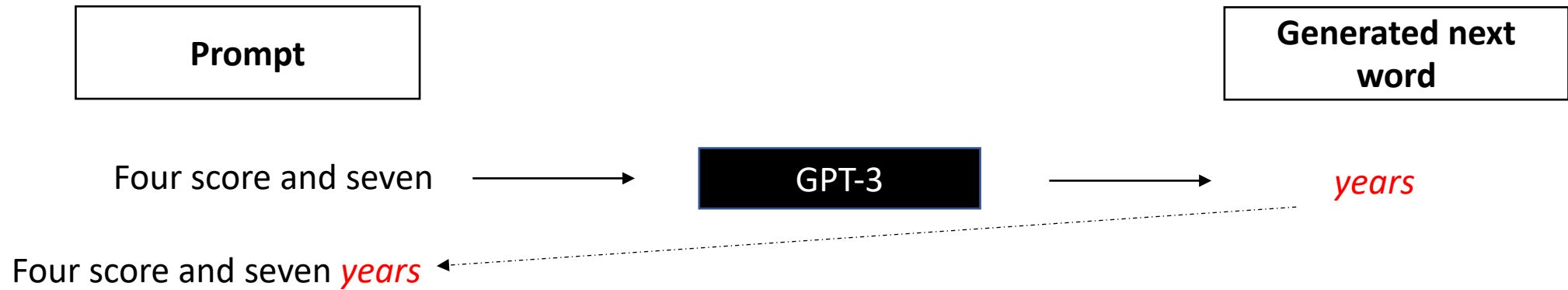
When we ask GPT-3 something (called a “prompt”), it either *picks* the most likely next word or *samples* a word from the table\* ...

\*we get to tell GPT-3 whether to pick or to sample

When we ask GPT-3 something (called a “prompt”), it either *picks* the most likely next word or *samples* a word from the table ...



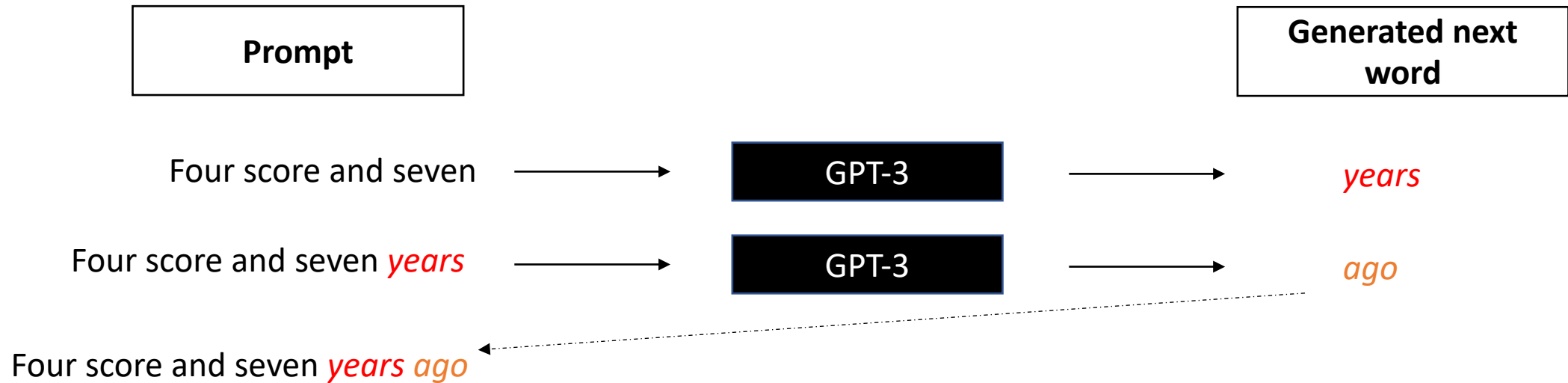
... it then appends that word to the original prompt ...



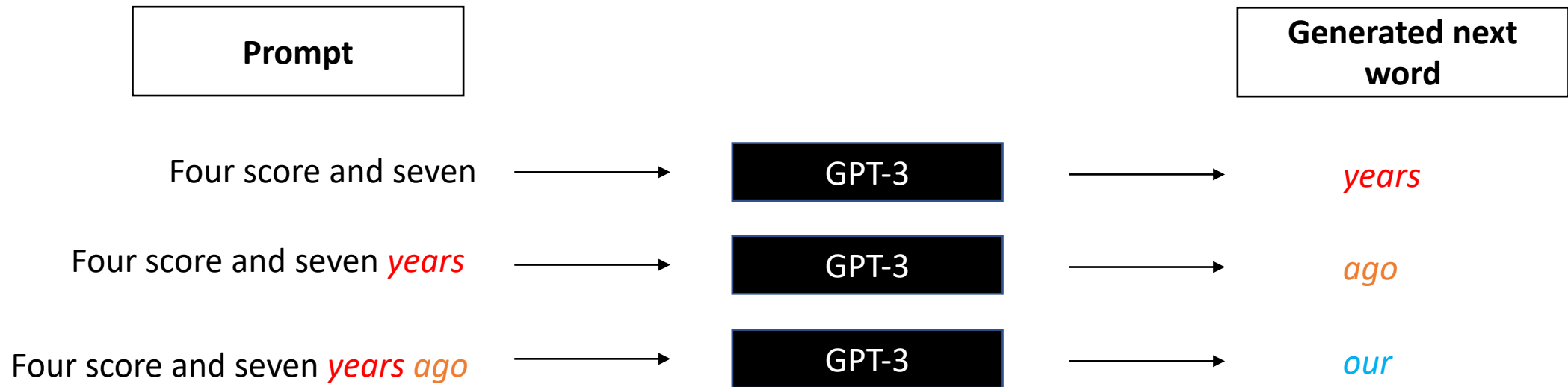
... generates the next word from the associated table ...



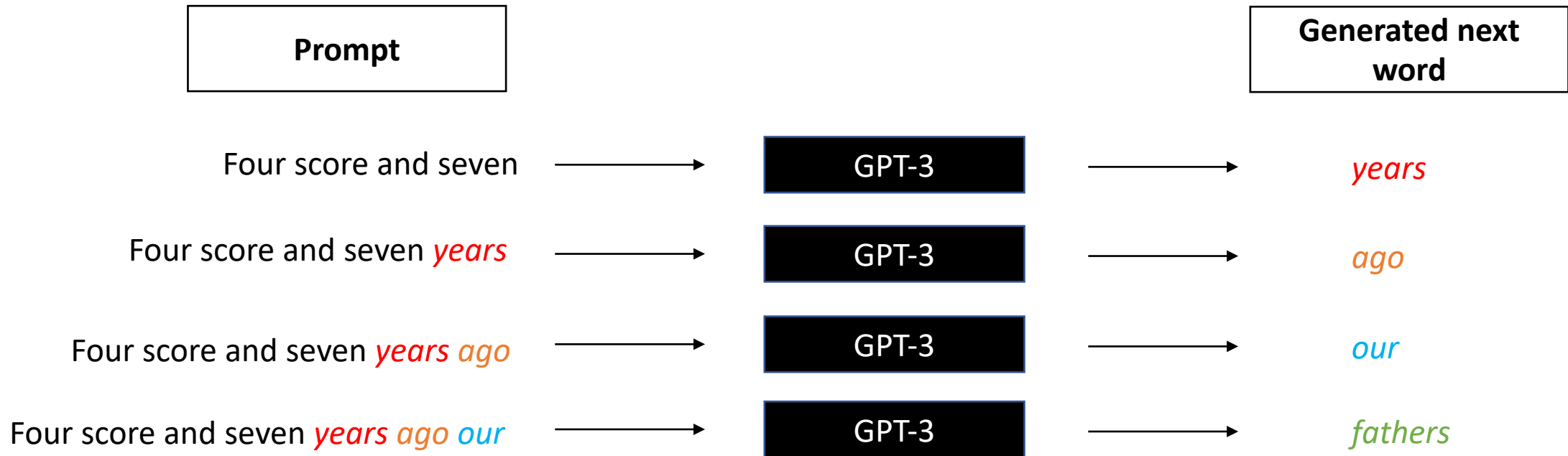
... appends *that* word to the input ...



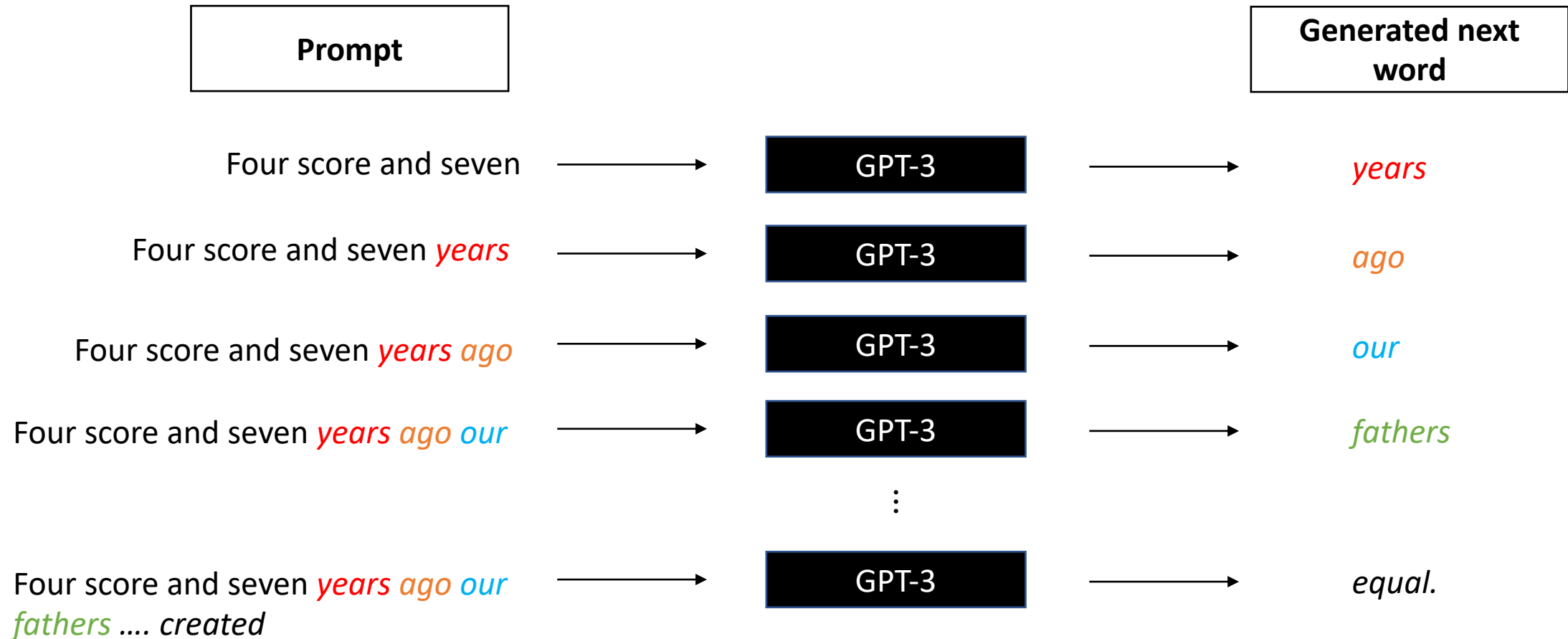
... generates the next word from the associated table and ...



... continues in this fashion for a while



... till it reaches a stopping condition\*



\*we have some control over this



This “repeatedly generate the next word” approach is how GPT-3 can generate paragraphs of text from a single-sentence input



[https://en.wikipedia.org/wiki/Gettysburg\\_Address](https://en.wikipedia.org/wiki/Gettysburg_Address)

# Sidebar: Why is it called GPT?

GPT stands for Generative Pretrained Transformer

- *Generative* – because it can generate the next word
- *Pretrained* – because it was pretrained on a lot of text from the Internet and thus ready to respond to questions
- *Transformer* – because the underlying Deep Neural Network uses a particular kind of network architecture called the Transformer

# Sidebar: What about GPT and GPT-2?

- Like GPT-3, GPT and GPT-2 were trained in a very similar “predict the next word” fashion
- Two key differences
  - GPT-3 was trained on much more data
  - GPT-3’s underlying neural network was much bigger

Model	Number of parameters*
GPT	117 million
GPT-2	1.5 billion
GPT-3	175 billion

\*roughly: how big is the underlying neural network

# Sidebar: What about GPT and GPT-2?

- Like GPT-3, GPT and GPT-2 were trained in a very similar “predict the next word” fashion
- Two key differences
  - GPT-3 was trained on much more data
  - GPT-3’s underlying neural network was much bigger
- To the researchers’ surprise, GPT-3 turned out to have *unanticipated* abilities (that GPT and GPT-2 didn’t have)

Model	Number of parameters*
GPT	117 million
GPT-2	1.5 billion
GPT-3	175 billion

\*roughly: how big is the underlying neural network

# What can GPT-3 do?

You can find numerous examples online. Here are a couple.

It can generate amazing *continuations* from a starting prompt



# It can generate amazing *continuations* from a starting prompt

Prompt

## The importance of being on twitter

by Jerome K. Jerome  
London, Summer 1897

It

---

Generated by GPT-3

is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling-cage. I called it an anomaly, and it is.

I spoke to the sexton, whose cottage, like all sexton's cottages, is full of antiquities and interesting relics of former centuries. I said to him, "My dear sexton, what does all this twittering mean?" And he replied, "Why, sir, of course it means Twitter." "Ah!" I said, "I know about that. But what is Twitter?"

"It is a system of short and pithy sentences strung together in groups, for the purpose of conveying useful information to the initiated, and entertainment and the exercise of wits to the initiated, and entertainment and the exercise of wits to the rest of us."

# It can quickly learn how to “auto complete” something new if you list a few examples in the prompt itself

<b>Prompt</b>
---------------

Poor English input: I eated the purple berries.

Good English output: I ate the purple berries.

Poor English input: Thank you for picking me as your designer. I'd appreciate it.

Good English output: Thank you for choosing me as your designer. I appreciate it.

Poor English input: The mentioned changes have done. or I did the alteration that you requested. or I changed things you wanted and did the modifications.

Good English output: The requested changes have been made. or I made the alteration that you requested. or I changed things you wanted and made the modifications.

Poor English input: I'd be more than happy to work with you in another project.



# It can quickly learn how to “auto complete” something new if you list a few examples in the prompt itself

## Prompt

Poor English input: I eated the purple berries.

Good English output: I ate the purple berries.

Poor English input: Thank you for picking me as your designer. I'd appreciate it.

Good English output: Thank you for choosing me as your designer. I appreciate it.

Poor English input: The mentioned changes have done. or I did the alteration that you requested. or I changed things you wanted and did the modifications.

Good English output: The requested changes have been made. or I made the alteration that you requested. or I changed things you wanted and made the modifications.

Poor English input: I'd be more than happy to work with you in another project.

Good English output: I'd be more than happy to work with you on another project.

## GPT-3 “Auto complete”

# It can quickly learn how to “auto complete” something new if you list a few examples in the prompt itself

**Prompt**

Poor English input: I eated the purple berries.

Good English output: I ate the purple berries.

Poor English input: Thank you for picking me as your designer. I'd appreciate it.

Good English output: Thank you for choosing me as your designer. I appreciate it.

Poor English input: The mentioned changes have done. or I did the alteration that you requested. or I changed things you wanted and did the modifications.

Good English output: The requested changes have been made. or I made the alteration that you requested. or I changed things you wanted and made the modifications.

Poor English input: I'd be more than happy to work with you in another project.

Good English output: I'd be more than happy to work with you on another project.

*GPT-3 has learned to fix the error!*

**GPT-3**  
**“Auto complete”**

But it is not very good at following user  
*instructions*

Let's make this request\* to GPT-3:

*help me write a short note to introduce myself to my neighbor*

# GPT-3's answer

help me write a short note to introduce myself to my neighbor.

what is a good introduction to a resume.

good introduction to a resume.

best photos of good resume introduction example sample.

# GPT-3's answer

help me write a short note to introduce myself to my neighbor.

what is a good introduction to a resume

good introduction to a resume.



best photos of good resume introduction example sample.

In addition to unhelpful answers, it can produce offensive or factually incorrect output as well

You can find numerous examples online.

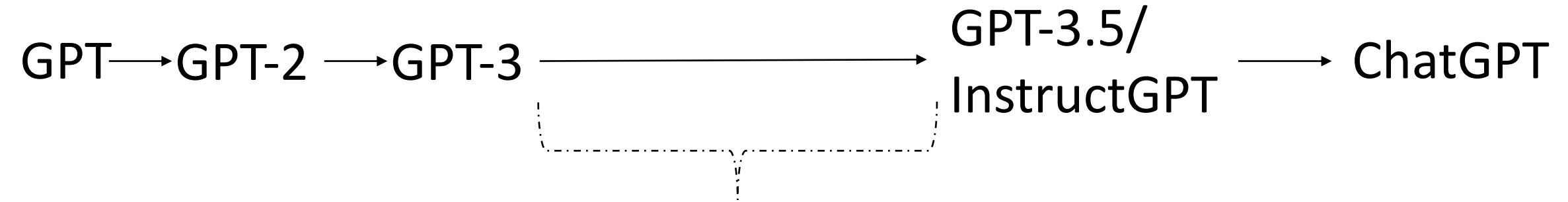
For example:

<https://www.technologyreview.com/2020/10/23/1011116/chatbot-gpt3-openai-facebook-google-safety-fix-racist-sexist-language-ai/>

OpenAI developed a 3-step solution to try to address this



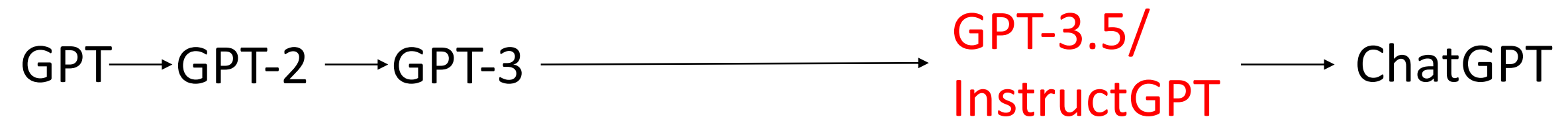
# The road to ChatGPT



We will look at this next\*

\*Details in <http://arxiv.org/abs/2203.02155>

Sidebar: There's a whole family of related models here that differ in subtle ways\*. Think of this as our “approximate name” for the category.



\*See <https://platform.openai.com/docs/model-index-for-researchers> for more.

# A plausible reason why GPT-3 sometimes produces unhelpful answers

- GPT-3 is simply trying to generate the next word repeatedly, using what it has gleaned from the billions of sentences it was trained on.
- It wasn't explicitly trained to generate good answers to user questions

# A plausible reason why GPT-3 sometimes produces unhelpful answers

- GPT-3 is simply trying to generate the next word repeatedly, using what it has gleaned from the billions of sentences it was trained on.
- It wasn't explicitly trained to generate good answers to user questions
- Examples of unhelpful responses

Prompt	GPT-3 Response
<i>Explain the moon landing to a 6 year old in a few sentences.</i>	Explain the theory of gravity to a 6 year old.
<i>Write a short poem about a wise frog.</i>	Write a short story in which a character has two different names.
<i>Why aren't birds real?</i>	Let's say I was a bird.

<https://openai.com/blog/instruction-following/>

# Solution – Step 1

- Get humans to write questions and helpful, accurate and inoffensive answers for them. About 11,000 such question-answer pairs were created.

# Solution – Step 1

- Get humans to write questions and helpful, accurate and inoffensive answers for them. About 11,000 such pairs were created. Example:

Prompt	Human-created answer
<i>Explain the moon landing to a 6 year old in a few sentences.</i>	People went to the moon in a big rocket, walked around and came back to Earth. They took pictures of what they saw and sent them back so we could all see them.*

# Solution – Step 1

- Get humans to write questions and helpful, accurate and inoffensive answers for them. About 11,000 such pairs were created
- In addition, take some questions users have asked GPT-3 and have humans create good answers them as well. About 1500 such pairs were put together

# Solution – Step 1

- Get humans to write questions and helpful, accurate and inoffensive answers for them. About 11,000 such pairs were created
- In addition, take some questions users have asked GPT-3 and have humans create good answers them as well. About 1500 such pairs were put together
- Using these ~12,500 question-answer pairs as training data, train GPT-3 some more (technical: fine-tuning via supervised learning)



# Solution – Step 1

- Get humans to write questions and helpful, accurate and inoffensive answers for them. About 11,000 such pairs were created
- Take some questions users have asked GPT-3 and have humans create good answers them as well. About 1500 such pairs were put together
- Using these ~12,500 question-answer pairs as training data, train GPT-3 some more (technical: fine-tuning via supervised learning)
- This step is called “Supervised Fine Tuning” (SFT)

# “Supervised Fine Tuning” helps

- SFT GPT-3 produces better answers. We would like to do more SFT

# “Supervised Fine Tuning” helps

- SFT GPT-3 produces better answers. We would like to do more SFT
- *But writing high-quality answers to thousands of questions is difficult and expensive*

What's easier than writing a good answer?

# What's easier than writing a good answer?

- *Ranking answers written by somebody else!*

# What's easier than writing a good answer?

- *Ranking answers written by somebody else!*
- We can ask GPT-3 to generate several answers to a question ...

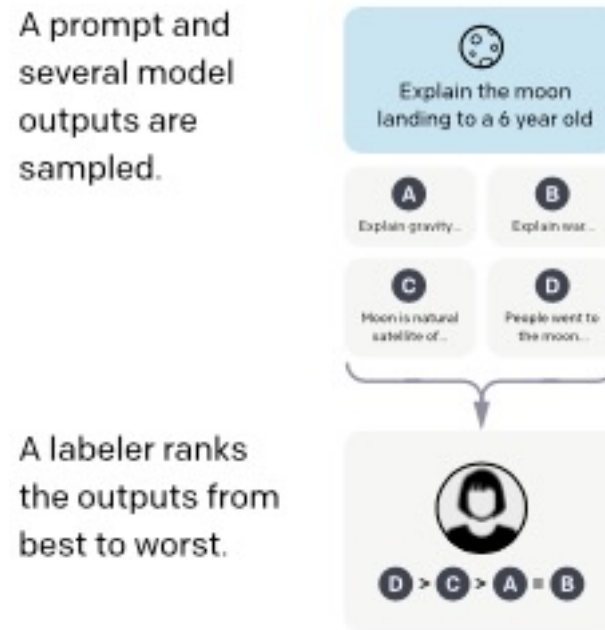


*How?*

*If we ask GPT-3 to sample the next word,  
remember that it can generate several next-  
words for the same input.*

# What's easier than writing a good answer?

- *Ranking answers written by somebody else!*
- We can ask GPT-3 to generate several answers to a question ...
- ... and have humans rank them from most useful to least useful



# This idea forms the basis of Step 2

- Data collection strategy
  - Collect questions (from those posed to GPT-3 by OpenAI users, or created by humans)
  - Get SFT GPT-3 to respond with several answers to each question
  - *Get humans to simply rank SFT GPT-3's answers for each question from most helpful to least helpful*
  - About 33,000 such questions (and ranked answers for each question) were compiled this way. Note that humans were not involved in creating answers, only in ranking them.

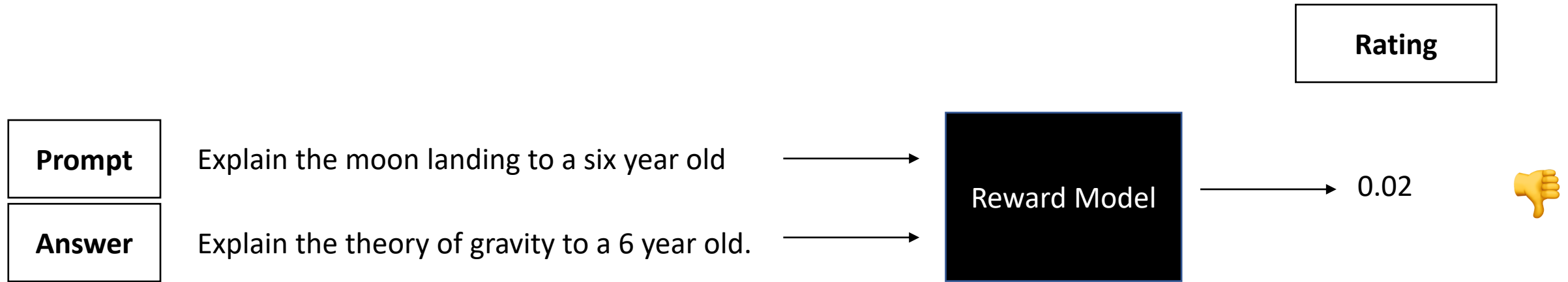


# Solution – Step 2

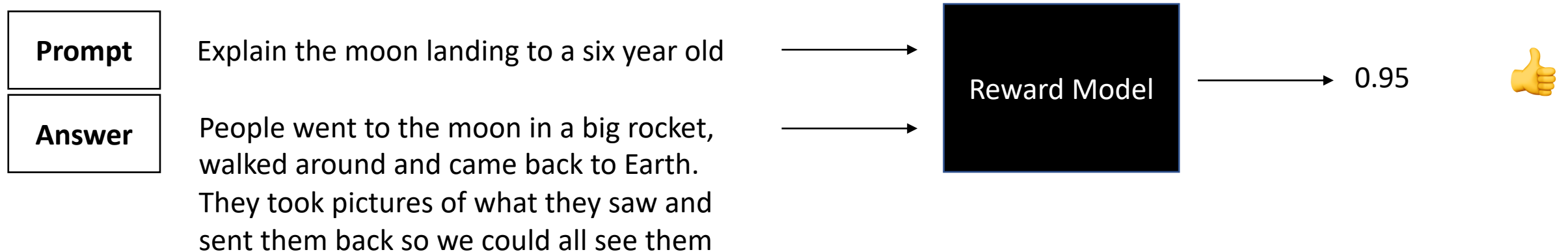
- Data collection strategy
  - Collect questions (from those posed to GPT-3 by OpenAI users, or created by humans)
  - Get SFT GPT-3 to respond with several answers to each question
  - *Get humans to simply rank SFT GPT-3's answers for each question from most helpful to least helpful*
  - About 33,000 such questions (and ranked answers for each question) were compiled this way. Note that humans were not involved in creating answers.
- Using this data, a new “Reward Model” is built (again using supervised learning) *that can do something very useful*

Given a question and an answer, the Reward Model can provide a *rating* - a single number that indicates how good the answer is

Given a question and an answer, the Reward Model can provide a *rating* - a single number that indicates how good the answer is



Given a question and an answer, the Reward Model can provide a *rating* - a single number that indicates how good the answer is



How can we use the Reward Model to improve SFT GPT-3 further?

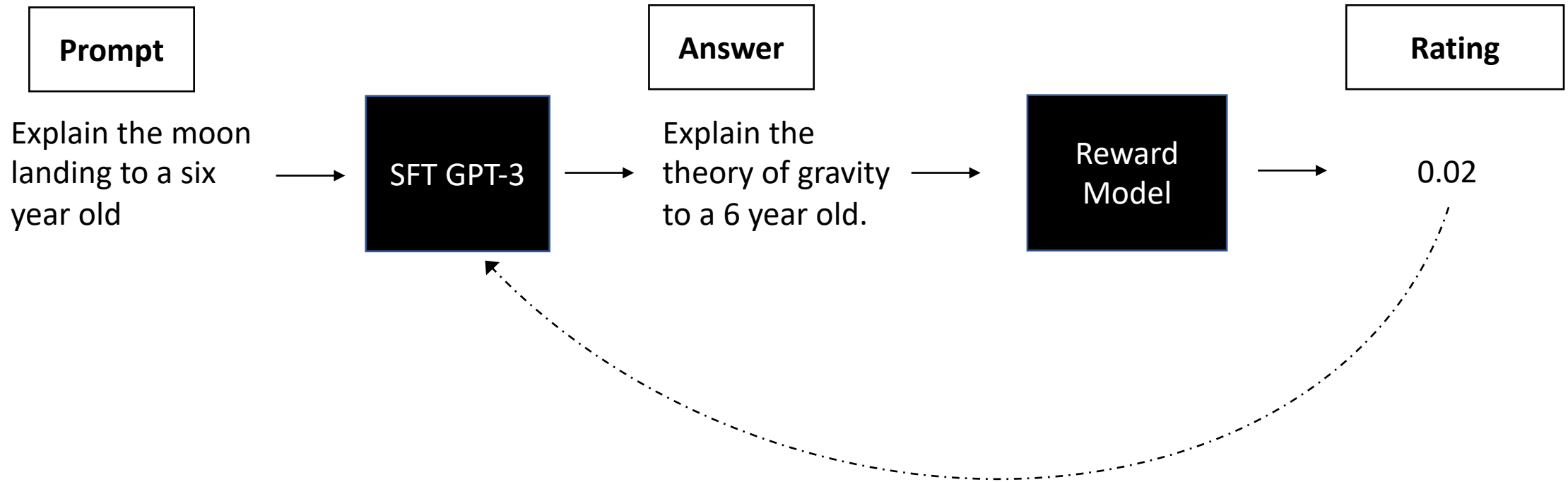
# Solution – Step 3

- Data collection strategy
  - Collect questions from those posed to GPT-3 by OpenAI users.
  - About 31,000 such questions were collected this way
  - *No answers are created or generated at this stage*

# Solution – Step 3

- Data collection strategy
  - Collect questions from those posed to GPT-3 by OpenAI users.
  - About 31,000 such questions were collected this way
  - *No answers are created or generated at this stage*
- *We then use these questions to further “fine tune” SFT GPT-3 (next slide)*

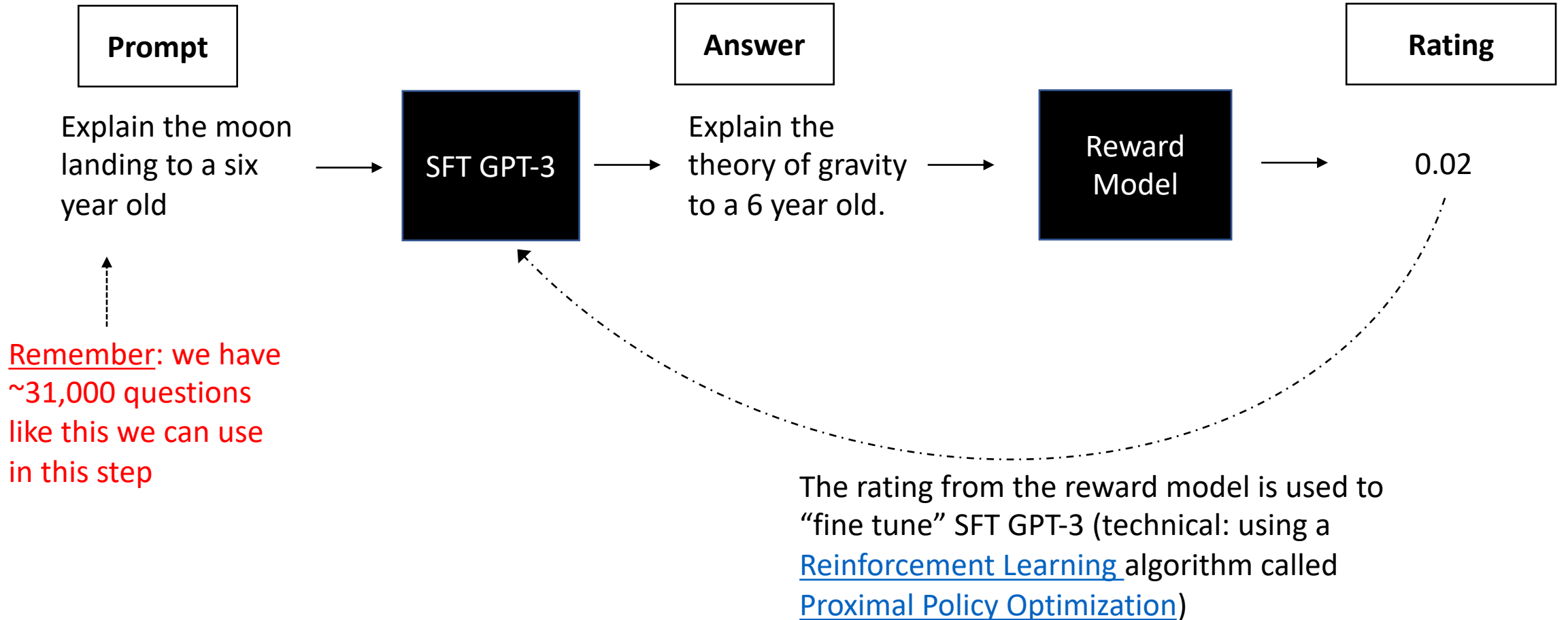
# Solution – Step 3



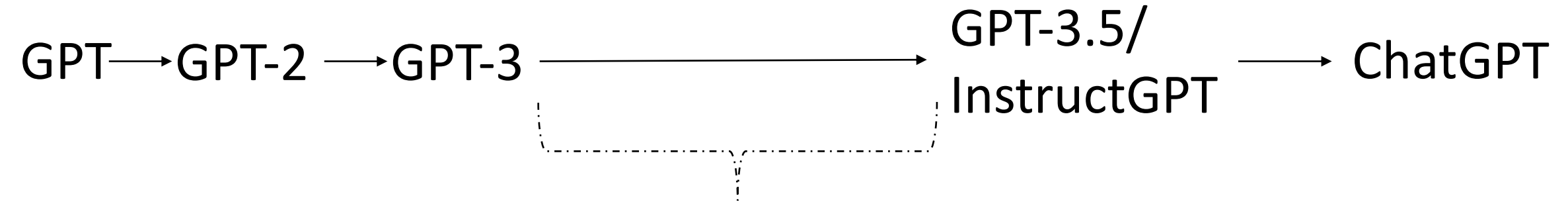
The rating from the reward model is used to "fine tune" SFT GPT-3 (technical: using a Reinforcement Learning algorithm called Proximal Policy Optimization)



# Solution – Step 3



# Summary



- (Step 1) Fine-tune GPT-3 using ~12,500 question-answer pairs. The answers are written by humans. The fine-tuned GPT-3 is SFT GPT-3
- (Step 2) Compile ~33,000 questions, have SFT GPT-3 provide multiple answers for each question, have humans rank the answers. With this training data, build an answer-rating model. This is called the Reward Model
- (Step 3) Using ~31,000 questions and the rating from the Reward Model for SFTGPT3's answer to each question, fine-tune SFT GPT3 using Reinforcement Learning

BTW, Steps 2 and 3 are referred to as  
“**Reinforcement Learning** from **Human Feedback**”  
(for obvious reasons 😊)

- (Step 1) Fine-tune GPT-3 using ~12,500 question-answer pairs. The answers are written by humans. The fine-tuned GPT-3 is SFT GPT-3
- (Step 2) Compile ~33,000 questions, have SFT GPT-3 provide multiple answers for each question, **have humans rank the answers**. With this training data, build an answer-rating model. This is called the Reward Model
- (Step 3) Using ~31,000 questions and the rating from the Reward Model for SFTGPT3's answer to each question, **fine-tune SFT GPT3 using Reinforcement Learning**.

OK, let's see how GPT 3.5/InstructGPT does on this prompt

help me write a short note to introduce myself to my neighbor.

# OK, let's see how GPT 3.5/InstructGPT does on this prompt

help me write a short note to introduce myself to my neighbor.



Hello Neighbor,

My name is [Name] and I recently moved into the neighborhood. I am excited to get to know my new neighbors and am looking forward to getting to know you. If you ever need anything, please don't hesitate to reach out.

Best,

[Name]

# Clearly better than GPT-3!

help me write a short note to introduce myself to my neighbor.



Hello Neighbor,

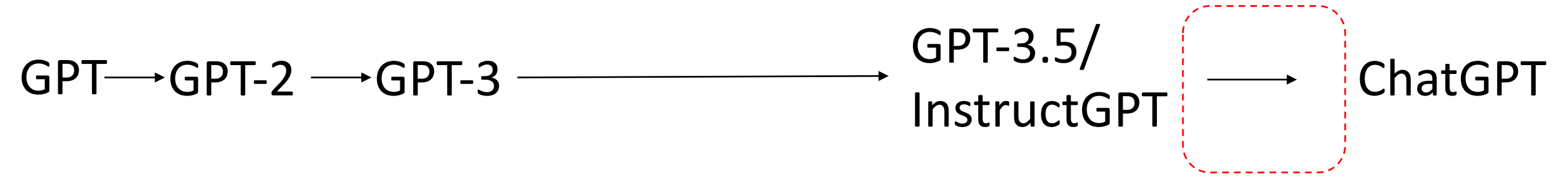
My name is [Name] and I recently moved into the neighborhood. I am excited to get to know my new neighbors and am looking forward to getting to know you. If you ever need anything, please don't hesitate to reach out.

Best,

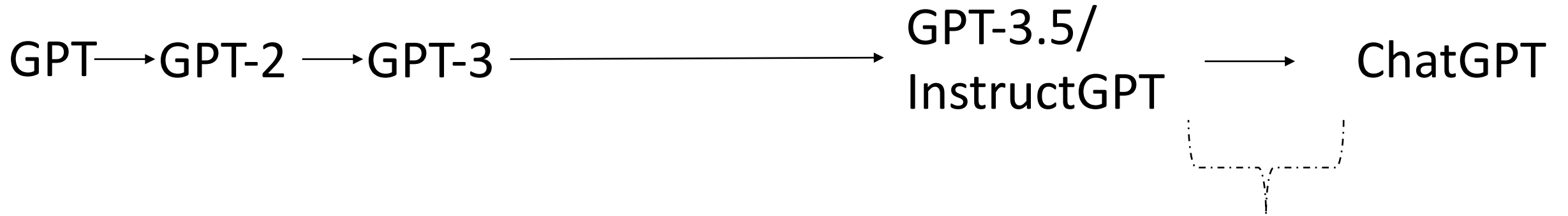
[Name]



# The last leg of our journey



# The last leg of our journey



Turns out it is very similar to the 3-step process we followed to get from GPT-3 to GPT-3.5/InstructGPT

Caveat: Based on [blog post from OpenAI that announced ChatGPT](#) (the only official source of information I could find)



# Overall approach

*ChatGPT is **fine-tuned** from a model in the GPT-3.5 series ...*

*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.*

This  
should  
sound  
familiar!



<https://openai.com/blog/ChatGPT/>

# Overall approach

*ChatGPT is fine-tuned from a model in the GPT-3.5 series ...*

*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.*

<https://openai.com/blog/ChatGPT/>

*Training data is conversations\* (instead of the question-response pairs used so far)*

\*since we want to build a conversational bot i.e., a chatbot! 😊

# Step 1

*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 mixed this new dialogue dataset with the InstructGPT dataset, which we transformed into a dialogue format.*

A new dialogue dataset was created and the question-answer dataset we saw earlier was transformed into a dialogue format

*This makes sense. **If we want a model that's good for chatting, let's train it on chatty (i.e., dialogue) training data!***

# Steps 2 and 3

*To create a **reward model** for **reinforcement learning** ... we took **conversations** that AI trainers had with the chatbot.*

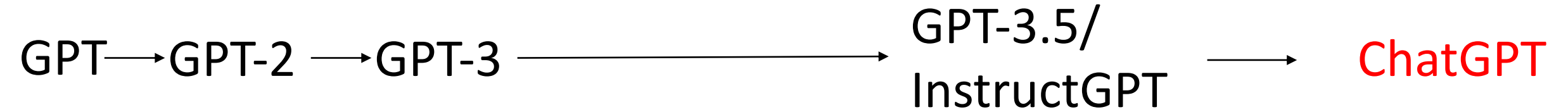
*We randomly selected a model-written message, sampled several alternative completions, and **had AI trainers rank them**.*

*Using these reward models, we can **fine-tune the model using Proximal Policy Optimization**. We performed several iterations of this process.*

Very similar to steps 2 and 3 from before

- Have humans rank model outputs
- Build a reward model with this data
- Use the reward model's ratings to fine-tune using Proximal Policy Optimization

# And that's how ChatGPT came into being!



# How well does ChatGPT do on our prompt?

RA help me write a short note to introduce myself to my neighbor.



Dear [Neighbor's Name],

I hope this note finds you well. My name is [Your Name], and I am your new neighbor. I recently moved into the [apartment/house] next to yours and wanted to take a moment to introduce myself.

I am a [profession/hobby] and enjoy spending my free time [activity/interest]. I also love meeting new people and getting to know my community. If you ever need anything or would like to chat, please don't hesitate to stop by and say hello.

I look forward to getting to know you and being a part of this wonderful neighborhood.

Best regards,  
[Your Name]

# How well does ChatGPT do on our prompt?

RA help me write a short note to introduce myself to my neighbor.



Dear [Neighbor's Name],

I hope this note finds you well. My name is [Your Name], and I am your new neighbor. I recently moved into the [apartment/house] next to yours and wanted to take a moment to introduce myself.

I am a [profession/hobby] and enjoy spending my free time [activity/interest]. I also love meeting new people and getting to know my community. If you ever need anything or would like to chat, please don't hesitate to stop by and say hello.

I look forward to getting to know you and being a part of this wonderful neighborhood.

Best regards,  
[Your Name]



# How well does ChatGPT do on our prompt?

RA help me write a short note to introduce myself to my neighbor.



Dear [Neighbor's Name],

I hope this note finds you well. My name is [Your Name], and I am your new neighbor. I recently moved into the [apartment/house] next to yours and wanted to take a moment to introduce myself.

I am a [profession/hobby] and enjoy spending my free time [activity/interest]. I also love meeting new people and getting to know my community. If you ever need anything or would like to chat, please don't hesitate to stop by and say hello.

I look forward to getting to know you and being a part of this wonderful neighborhood.

Best regards,  
[Your Name]

*Wait, there's more 😊*

*Because ChatGPT is trained on conversations, we can ask follow-on questions!!*



# How well does ChatGPT do on our prompt?

RA help me write a short note to introduce myself to my neighbor.

RA can you make it more formal?



Dear [Neighbor's Name],

I hope this note finds you well. My name is [Your Name], and I am your new neighbor. I recently moved into the [apartment/house] next to yours and wanted to take a moment to introduce myself.

I am a [profession/hobby] and enjoy spending my free time [activity/interest]. I also love meeting new people and getting to know my community. If you ever need anything or would like to chat, please don't hesitate to stop by and say hello.

I look forward to getting to know you and being a part of this wonderful neighborhood.

Best regards,  
[Your Name]

# How well does ChatGPT do on our prompt?

RA help me write a short note to introduce myself to my neighbor.

 Dear [Neighbor's Name],


I hope this note finds you well. My name is [Your Name], and I am your new neighbor. I recently moved into the [apartment/house] next to yours and wanted to take a moment to introduce myself.

I am a [profession/hobby] and enjoy spending my free time [activity/interest]. I also love meeting new people and getting to know my community. If you ever need anything or would like to chat, please don't hesitate to stop by and say hello.

I look forward to getting to know you and being a part of this wonderful neighborhood.

Best regards,  
[Your Name]

RA can you make it more formal?

 Certainly, here's a more formal version:

Dear [Neighbor's Name],

I hope this letter finds you in good health and spirits. My name is [Your Name], and I am your new neighbor. I recently moved into the [apartment/house] adjacent to yours and wanted to take a moment to introduce myself formally.

As a [profession/hobby], I value building strong relationships with my neighbors and am excited to get to know the community. In my free time, I enjoy [activity/interest], and I am always open to discovering new hobbies and interests.

I wanted to extend an invitation to stop by anytime if you need any assistance or if you would like to have a friendly conversation. I look forward to getting to know you and being a part of this wonderful neighborhood.

Thank you for taking the time to read my letter. Please accept my warmest regards.

Sincerely,

# How well does ChatGPT do on our prompt?

RA help me write a short note to introduce myself to my neighbor.

 Dear [Neighbor's Name],


I hope this note finds you well. My name is [Your Name], and I am your new neighbor. I recently moved into the [apartment/house] next to yours and wanted to take a moment to introduce myself.

I am a [profession/hobby] and enjoy spending my free time [activity]. I also love meeting new people and getting to know my community. If I ever need anything or would like to chat, please don't hesitate to stop by and say hello.

I look forward to getting to know you and being a part of this wonderful neighborhood.

Best regards,  
[Your Name]

RA can you make it more formal?

 Certainly, here's a more formal version:

Dear [Neighbor's Name],

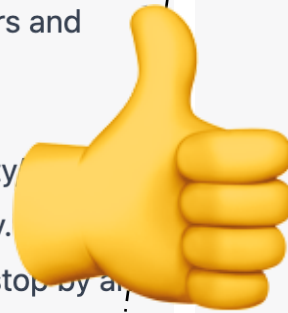
I hope this letter finds you in good health and spirits. My name is [Your Name], and I am your new neighbor. I recently moved into the [apartment/house] adjacent to yours and wanted to take a moment to introduce myself formally.

As a [profession/hobby], I value building strong relationships with my neighbors and am excited to get to know the community. In my free time, I enjoy [activity/interest], and I am always open to discovering new hobbies and interests.

I wanted to extend an invitation to stop by anytime if you need any assistance or if you would like to have a friendly conversation. I look forward to getting to know you and being a part of this wonderful neighborhood.

Thank you for taking the time to read my letter. Please accept my warmest regards.

Sincerely,



Please note: While ChatGPT is better than GPT-3 in many ways, it is still prone to many of the same shortcomings

See <https://github.com/giuven95/ChatGPT-failures> for examples

We have reached the end of this journey\*! 🥳

Thanks for joining me!

Happy Learning!

# Appendix: How to make GPT-3.5 behave like ChatGPT

# You can make GPT-3.5/InstructGPT “conversational” like ChatGPT by feeding it the full question-response sequence each time

<b>First prompt</b>
<b>GPT-3.5 Answer</b>

help me write a short note to introduce myself to my neighbor.



Hello Neighbor,

My name is [Name] and I recently moved into the neighborhood. I am excited to get to know my new neighbors and am looking forward to getting to know you. If you ever need anything, please don't hesitate to reach out.

Best,

[Name]

## Second Prompt

help me write a short note to introduce myself to my neighbor.



Hello Neighbor,

My name is [Name] and I recently moved into the neighborhood. I am excited to get to know my new neighbors and am looking forward to getting to know you. If you ever need anything, please don't hesitate to reach out.

Best,  
[Name]

can you make it more formal?

*Note how the first prompt and its answer are now part of the second prompt*

## GPT-3.5 Answer

Greetings Neighbor,

My name is [Name], and I am delighted to make your acquaintance as I have recently moved into the area. I am eager to meet my new neighbors and am excited to get to know you better. Please do not hesitate to reach out if there is anything I can do to help.

Sincerely,  
[Name]