

# Introduction to Language Models

Michael Hahn

April 2024

# Language Modeling

The task of predicting the next word

# Language Modeling

The task of predicting the next word

the students attended a \_\_\_\_\_

# Language Modeling

The task of predicting the next word

seminar

the students attended a \_\_\_\_\_ lecture

concert

...

# Language Modeling

The task of predicting the next word probabilistically

the students attended a \_\_\_\_\_

seminar	<input type="checkbox"/>
lecture	<input type="checkbox"/>
concert	<input type="checkbox"/>
toothpaste	<input type="checkbox"/>
the	<input type="checkbox"/>
...	

$$P(\mathbf{x}^{(t+1)} | \mathbf{x}^{(t)}, \dots, \mathbf{x}^{(1)})$$

↑  
next word

←  
context

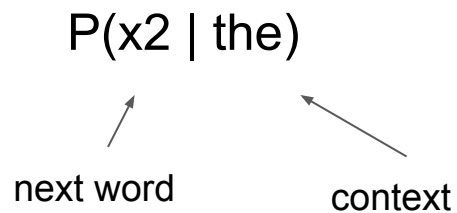
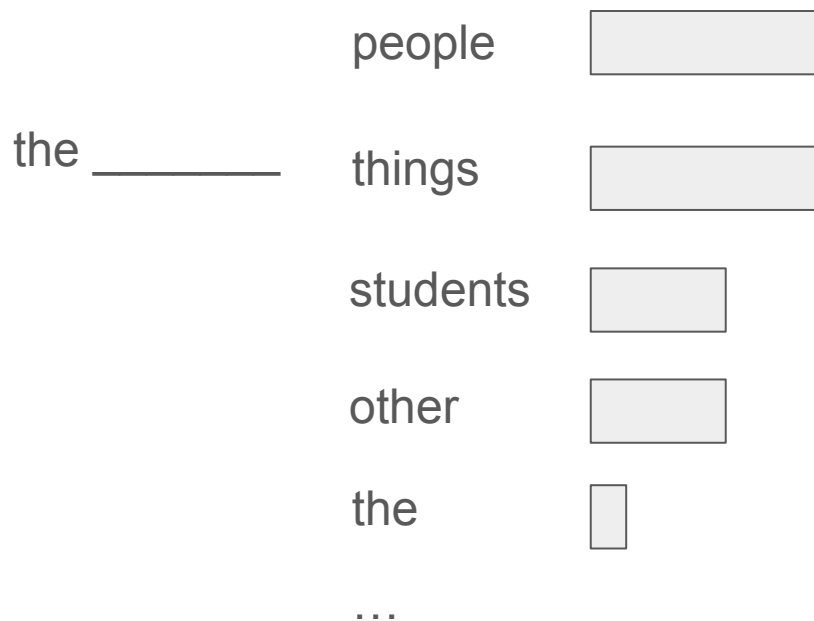
# Generating a Text

Generating a text

the \_\_\_\_\_

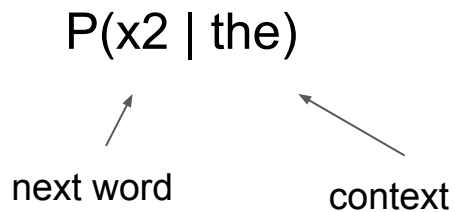
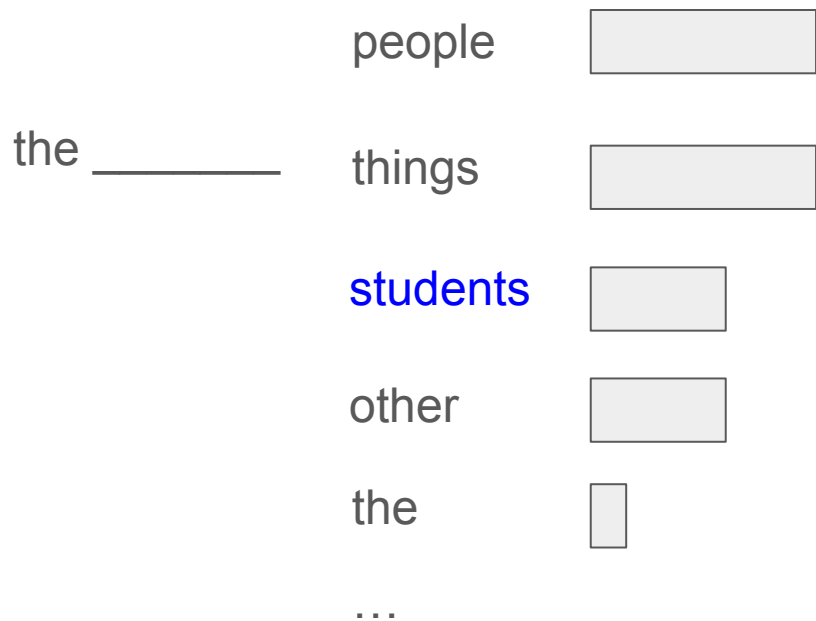
# Generating a Text

Generating a text



# Generating a Text

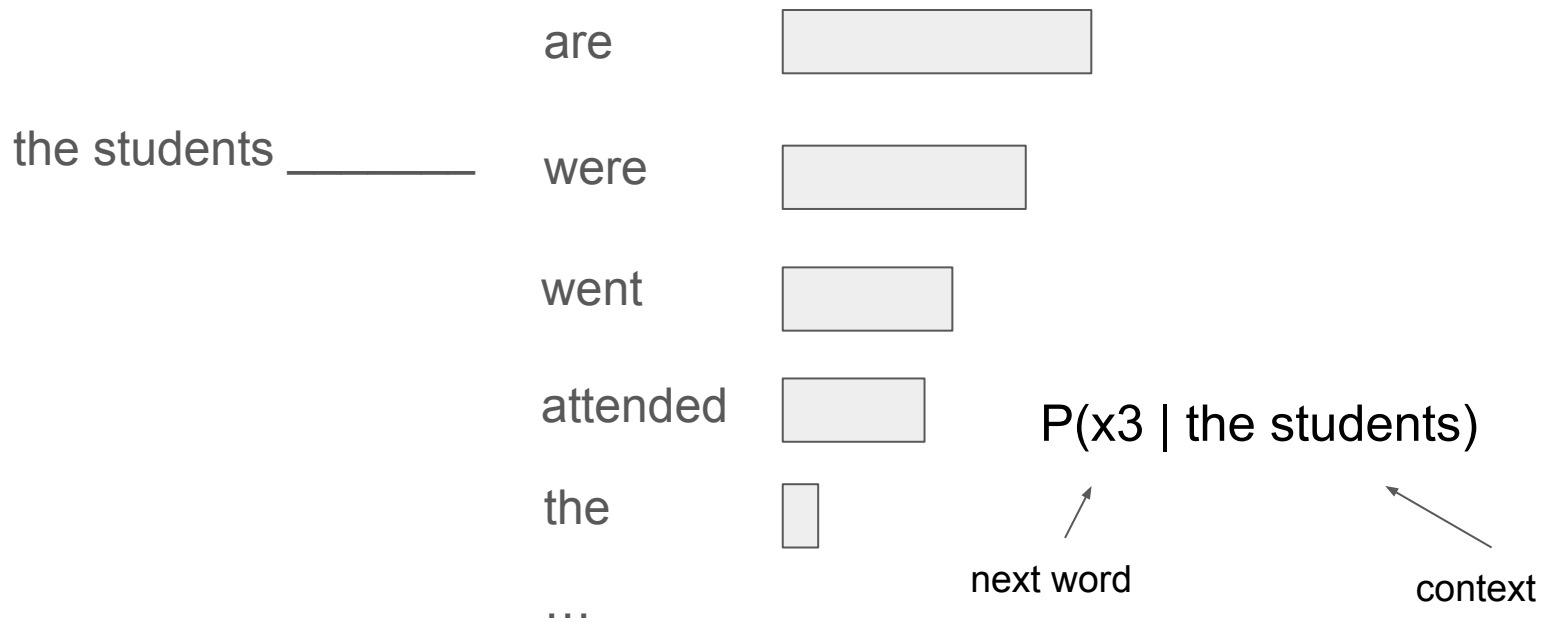
Generating a text





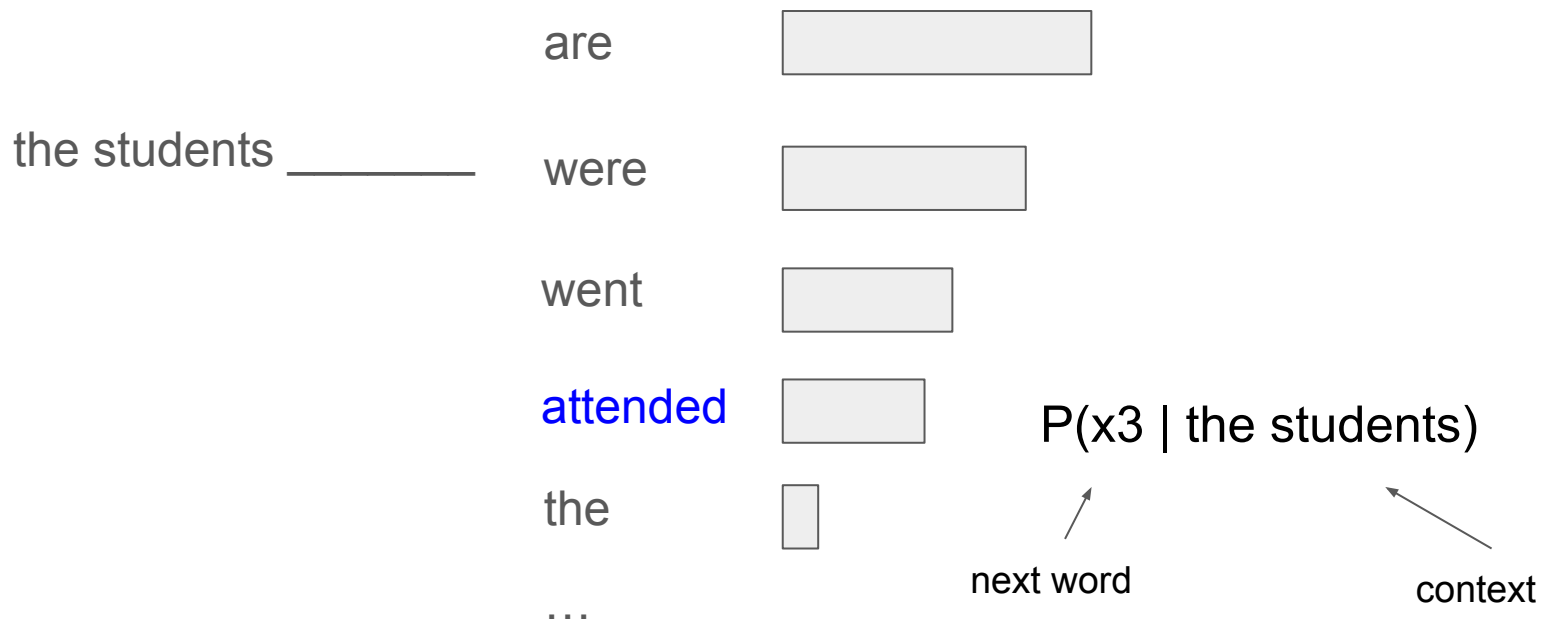
# Generating a Text

Generating a text



# Generating a Text

Generating a text



# Generating a Text

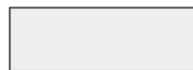
Generating a text

the students attended \_\_\_\_\_

the



a



another



seminars



some



...

$P(x_4 \mid \text{the students attended})$

next word

context



# Generating a Text

Generating a text

the students attended \_\_\_\_\_

the



**a**



another



seminars



some



...

$P(x_4 \mid \text{the students attended})$

next word

context

# Generating a Text

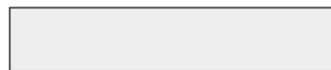
Generating a text

the students attended a \_\_\_\_\_

seminar



lecture



concert



toothpaste



$P(x_5 \mid \text{the students attended a})$

the



next word



context



...

# Generating a Text

Generating a text

the students attended a \_\_\_\_\_

seminar



lecture



concert



toothpaste



the



...

$P(x_5 \mid \text{the students attended a})$

next word

context



# Language Modeling

$P(\text{the students attended a seminar})$

=

$P(\text{the})$

×  $P(\text{students} \mid \text{the})$

×  $P(\text{attended} \mid \text{the students})$

×  $P(\text{a} \mid \text{the students attended})$

×  $P(\text{seminar} \mid \text{the students attended a})$

# Language Modeling

General Formula:

$$\begin{aligned} P(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}) &= P(\mathbf{x}^{(1)}) \times P(\mathbf{x}^{(2)} | \mathbf{x}^{(1)}) \times \dots \times P(\mathbf{x}^{(T)} | \mathbf{x}^{(T-1)}, \dots, \mathbf{x}^{(1)}) \\ &= \prod_{t=1}^T P(\mathbf{x}^{(t)} | \mathbf{x}^{(t-1)}, \dots, \mathbf{x}^{(1)}) \end{aligned}$$





what is the |



what is the **tower of london**

what is the **richest country in the world**

what is the **one piece**

what is the **name of the huge area in the centre of australia**

what is the **capital of australia**

what is the **weather today**

what is the **right course of action**

what is the **london eye**

what is the **meaning of life**

# How to calculate $P(\text{next word} \mid \text{context})$ ?

Traditional Approach  
(roughly until 2015)

$P(\text{seminar} \mid \text{the students attended a})$

$= \#(\text{the students attended a seminar}) / \#(\text{the students attended a})$

↑  
how often does each  
sequence occur in a  
huge corpus?

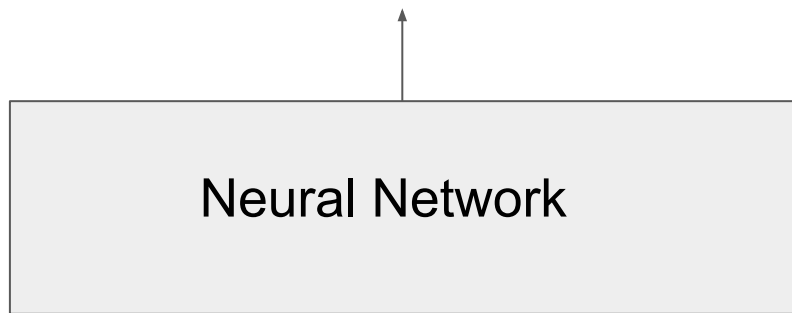
## Problem:

Longer sequences often never observed in corpus.  
Need to back-off to shorter contexts.  
Cannot capture long-distance relations.

# How to calculate $P(\text{next word} \mid \text{context})$ ?

Modern Approach  
(roughly since 2015)

$P(\dots \mid \text{the students attended a})$



trained on large  
corpus

the students attended a

# Input Representation

## Word Embeddings

	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6	dim 7	...
a	0.474	0.061	-0.262	0.287	0.051	-0.075	-0.084	
aardvark	-0.119	-0.115	0.222	-0.166	-0.259	-0.018	0.291	
abelian	-0.484	-0.016	-0.315	0.46	0.193	0.262	0.15	
acid	0.385	0.434	-0.203	0.058	0.154	-0.195	-0.249	

# Input Representation

Problem: what to do with unknown/rare words?

# Input Representation

Problem: what to do with unknown/rare words?



*The road running through Uds campus is called **Stuhlsatzenhausweg***

*may not have  
appeared in the  
training set!*

# Input Representation

Problem: what to do with unknown/rare words?



*The road running through Uds campus is called Stuhlsatzenhausweg*

Traditional Approach: Replace by special token

*The road running through Uds campus is called <UNKNOWN>*

Not very satisfying!

# Input Representation

Problem: what to do with unknown/rare words?



*The road running through UdS campus is called Stuhlsatzenhausweg*

Modern approach: subword tokenization

*The\_ road\_ run ning\_ through\_ U d S\_ campus\_ is called\_ Stuhl satz en haus weg\_*

Can represent any input.

In the worst case, just need to back off to the individual letters.



# Neural Network: Transformer

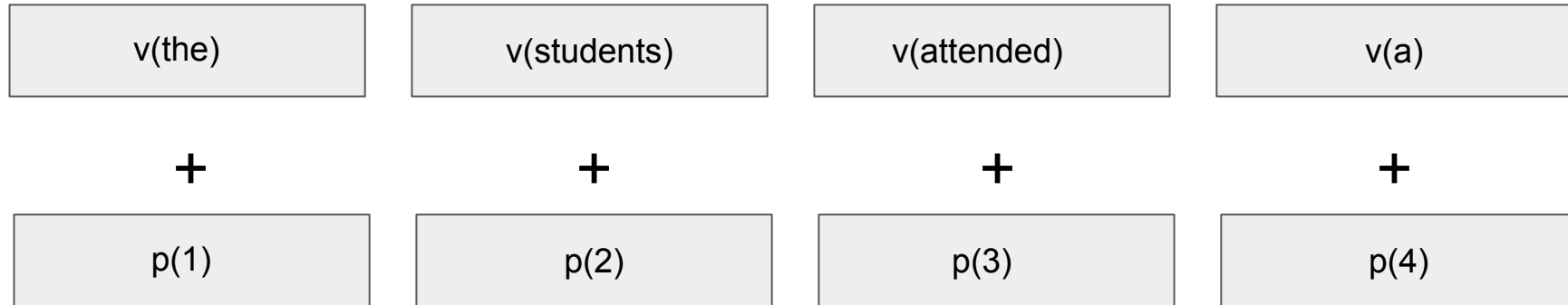
$v(\text{the})$

$v(\text{students})$

$v(\text{attended})$

$v(\text{a})$

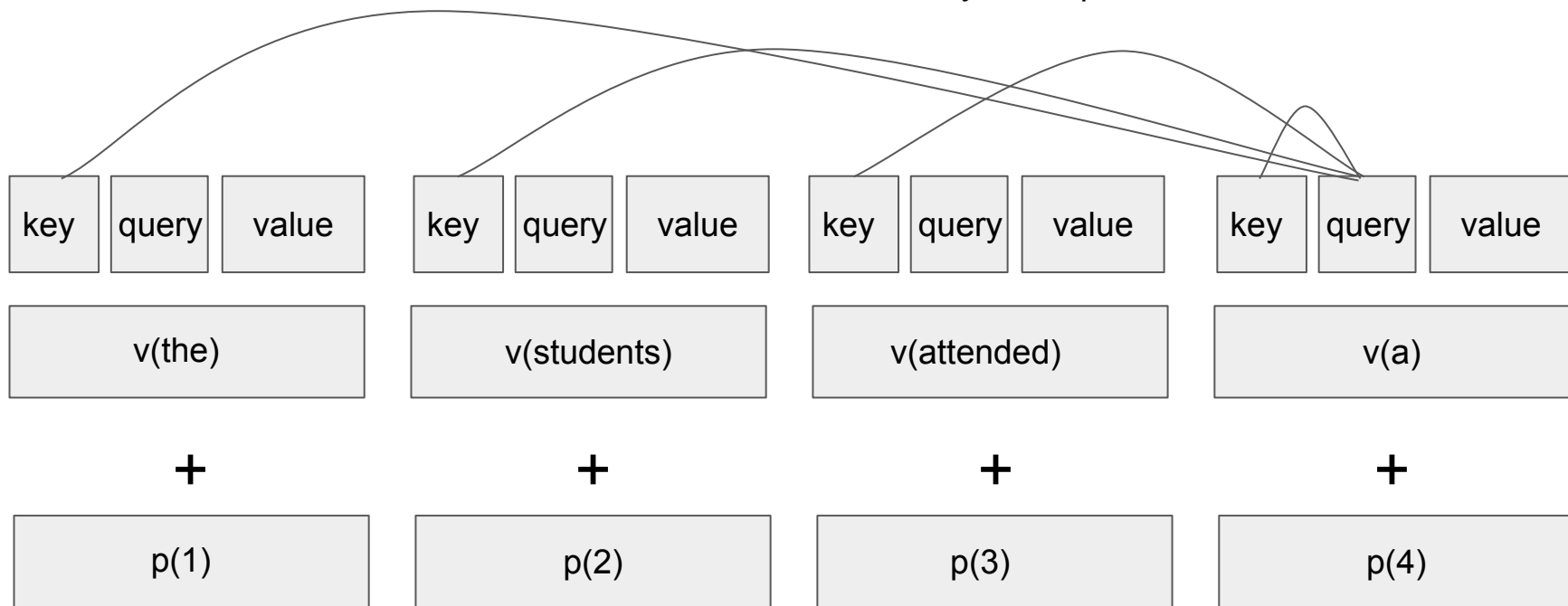
# Neural Network: Transformer



positional embeddings indicating  
position in the sequence

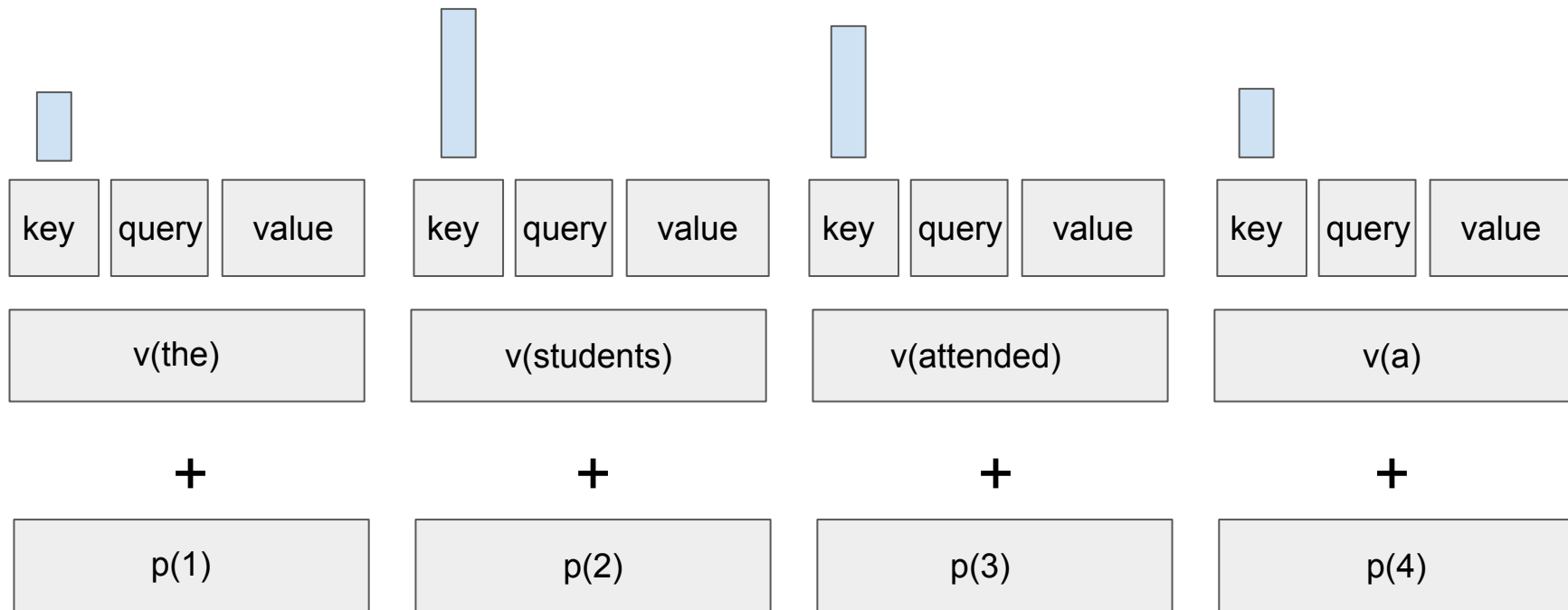
# Neural Network: Transformer

*compute all dot products between keys and queries*



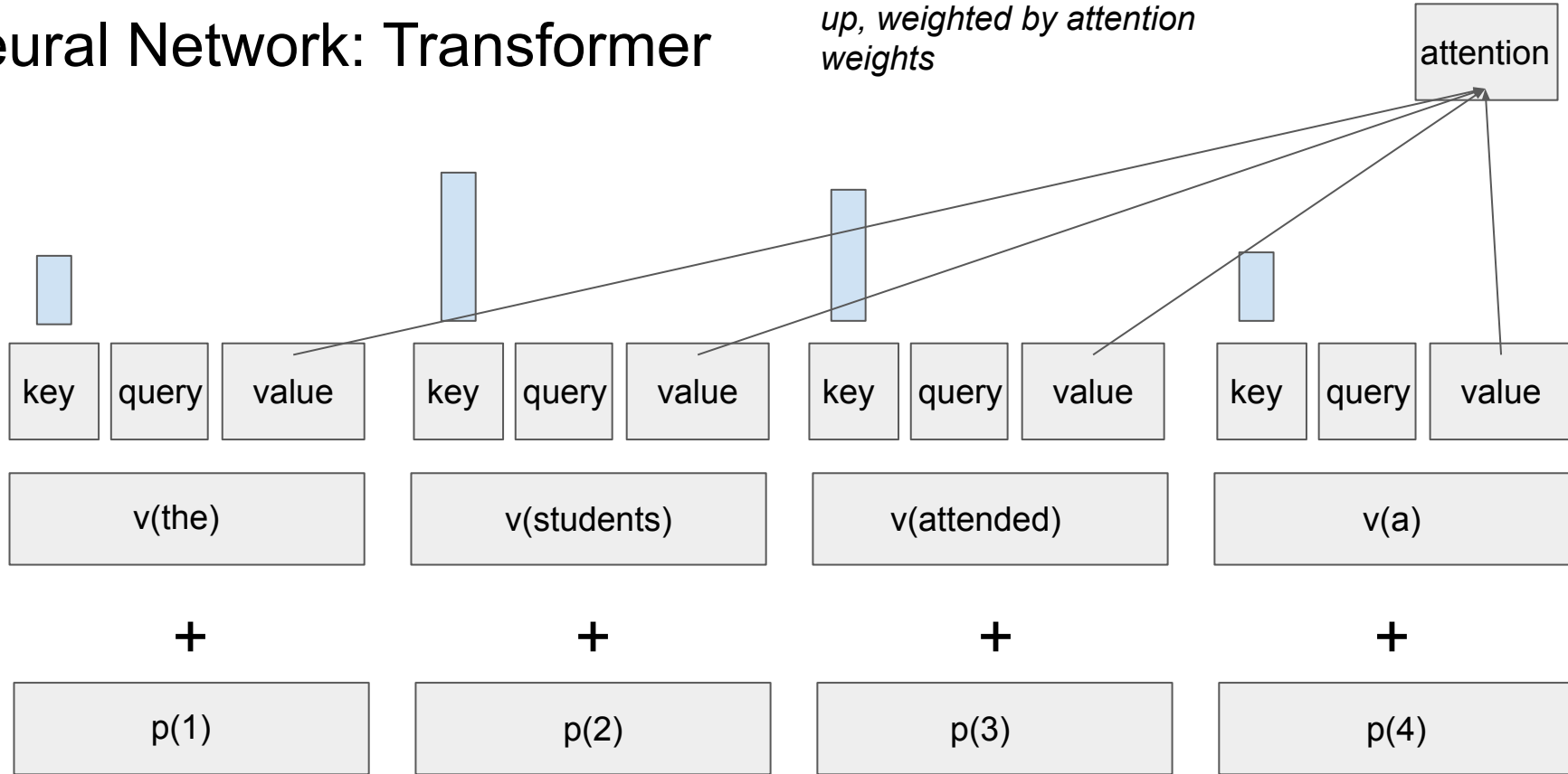
# Neural Network: Transformer

*attention weights are  
normalized to sum up to 1*



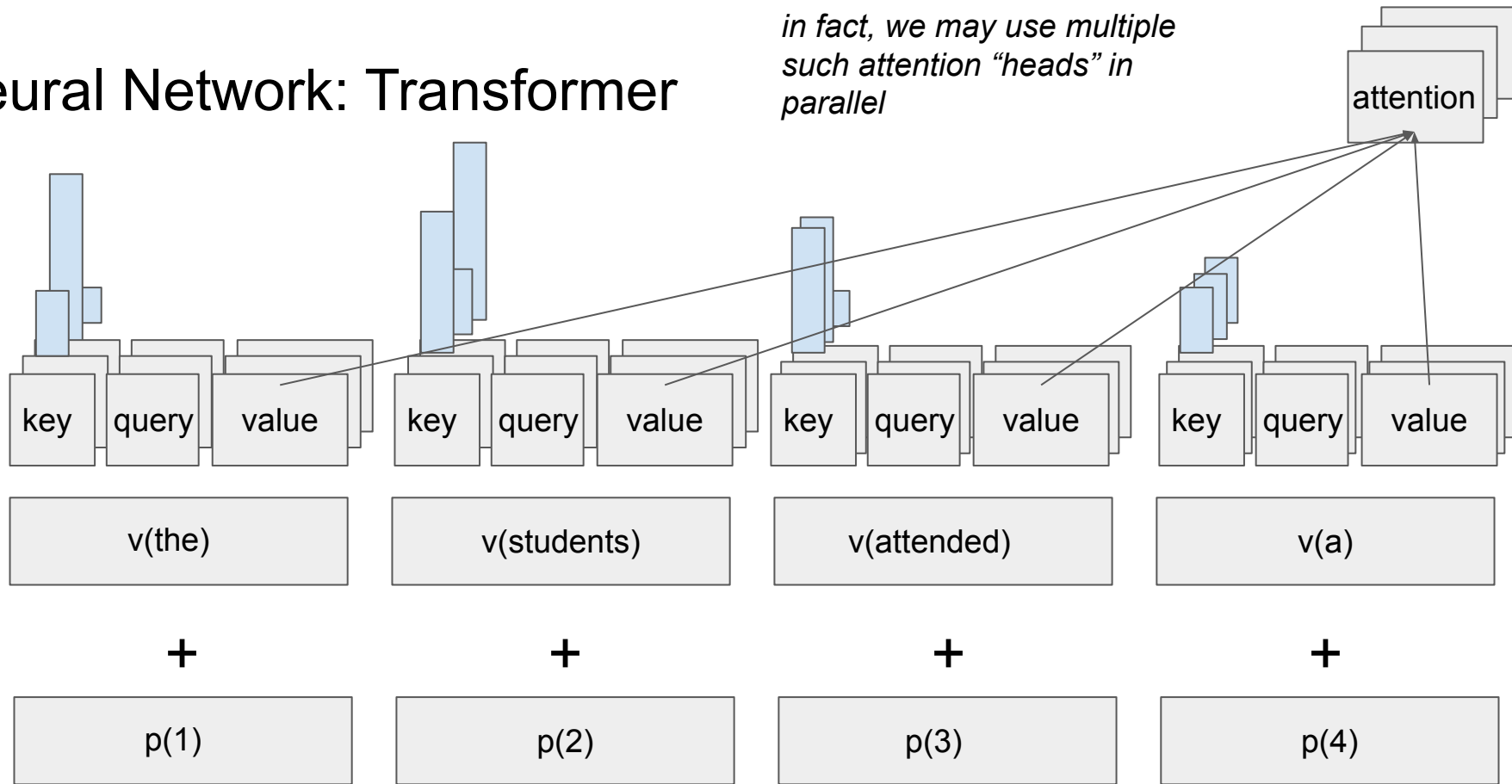
# Neural Network: Transformer

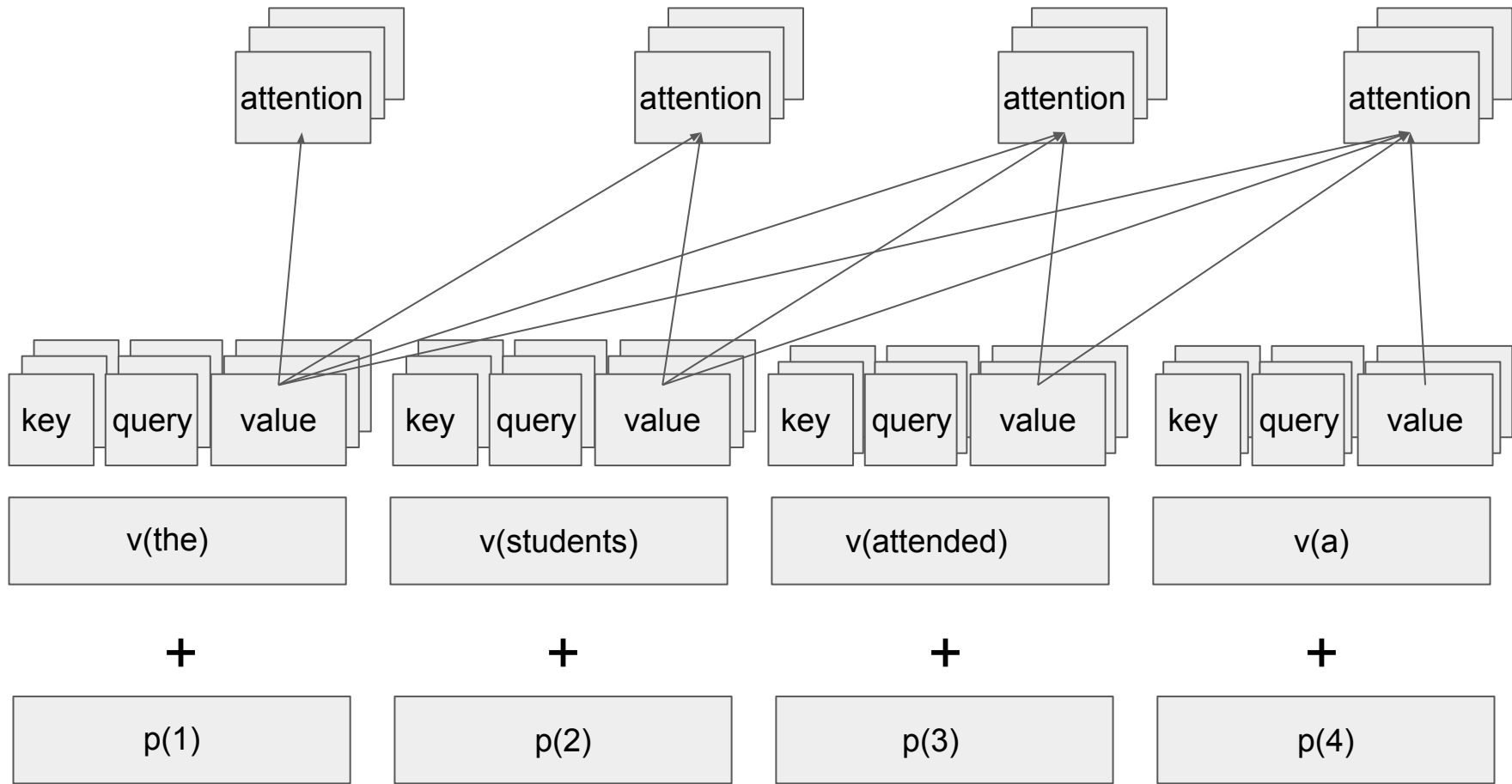
*value vectors are summed up, weighted by attention weights*

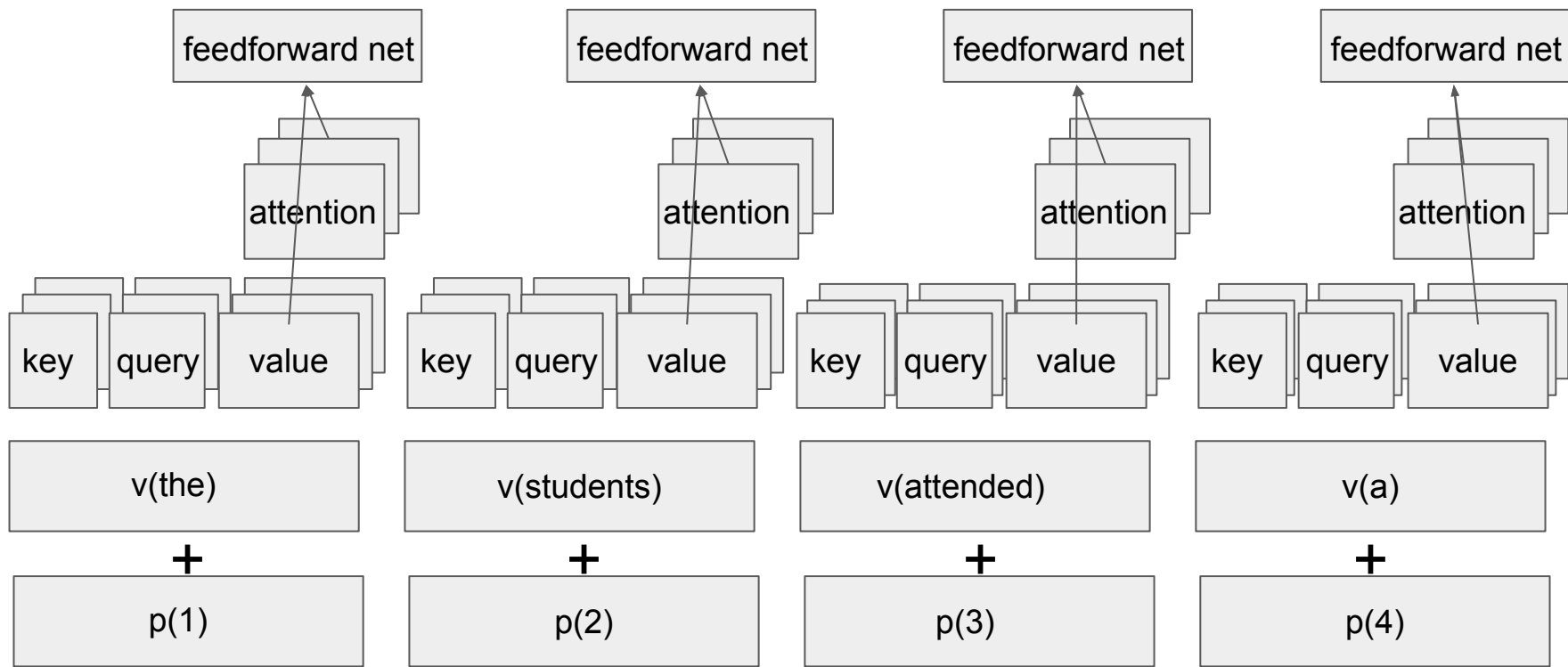


# Neural Network: Transformer

*in fact, we may use multiple such attention "heads" in parallel*

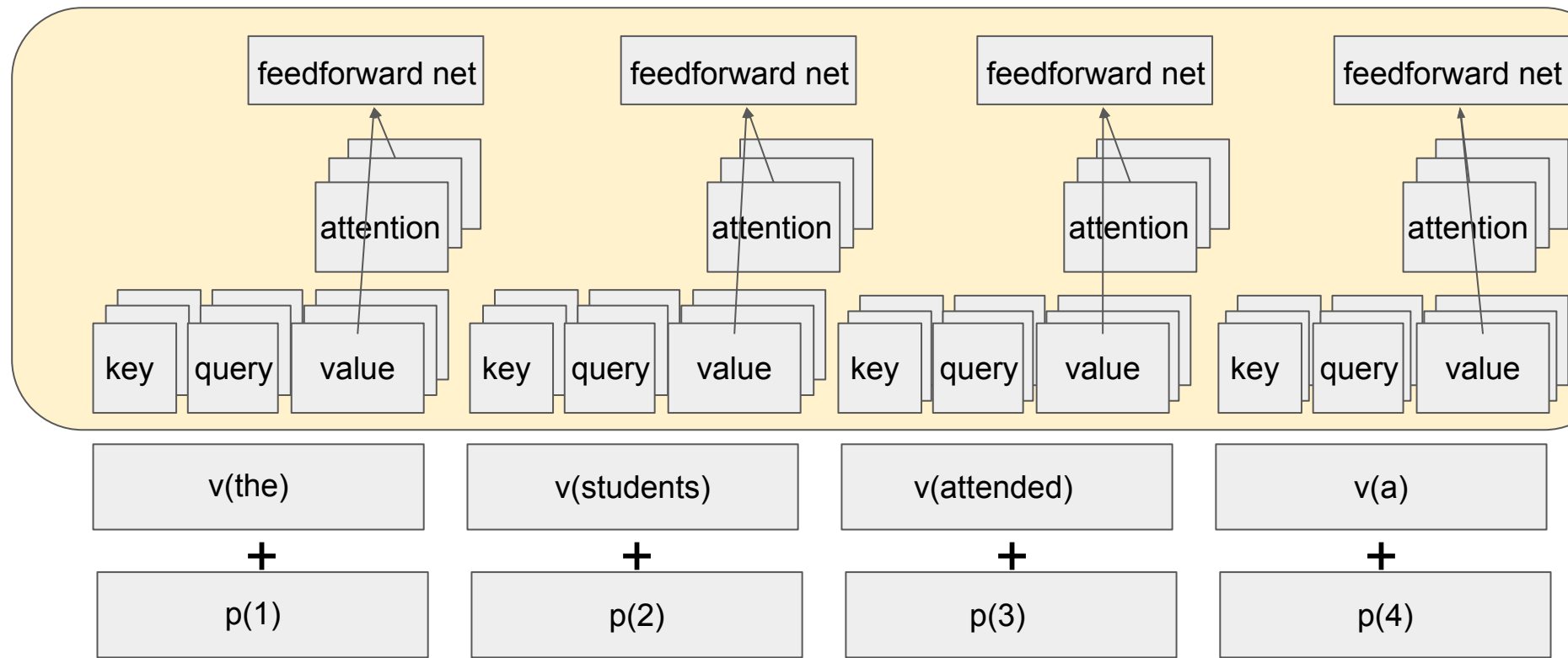








*this is one transformer layer*



Layer N

...

Layer 3

Layer 2

Layer 1

v(the)

v(students)

v(attended)

v(a)

+

+

+

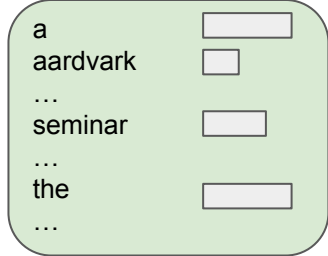
+

p(1)

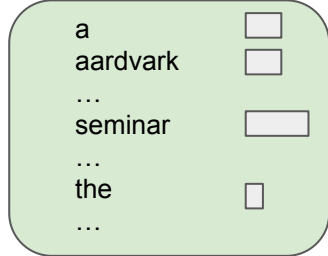
p(2)

p(3)

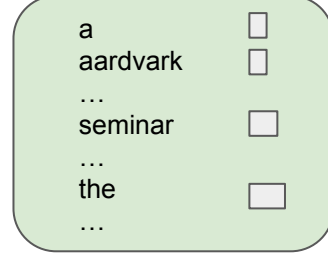
p(4)



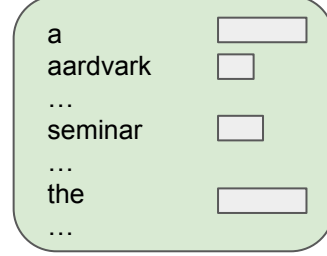
feedforward net



feedforward net



feedforward net



feedforward net

Layer N

...

Layer 1

$v(\text{the})$

+

$p(1)$

$v(\text{students})$

+

$p(2)$

$v(\text{attended})$

+

$p(3)$

$v(a)$

+

$p(4)$

# GPT-3: Prompting & In-Context Learning

Circulation revenue has increased by 5%  
in Finland. // Positive

Panostaja did not disclose the purchase  
price. // Neutral

Paying off the national debt will be  
extremely painful. // Negative

The company anticipated its operating  
profit to improve. // \_\_\_\_\_



# GPT-3: Prompting & In-Context Learning

Circulation revenue has increased by 5% in Finland. // Positive

Panostaja did not disclose the purchase price. // Neutral

Paying off the national debt will be extremely painful. // Negative

The company anticipated its operating profit to improve. // \_\_\_\_\_



Circulation revenue has increased by 5% in Finland. // Finance

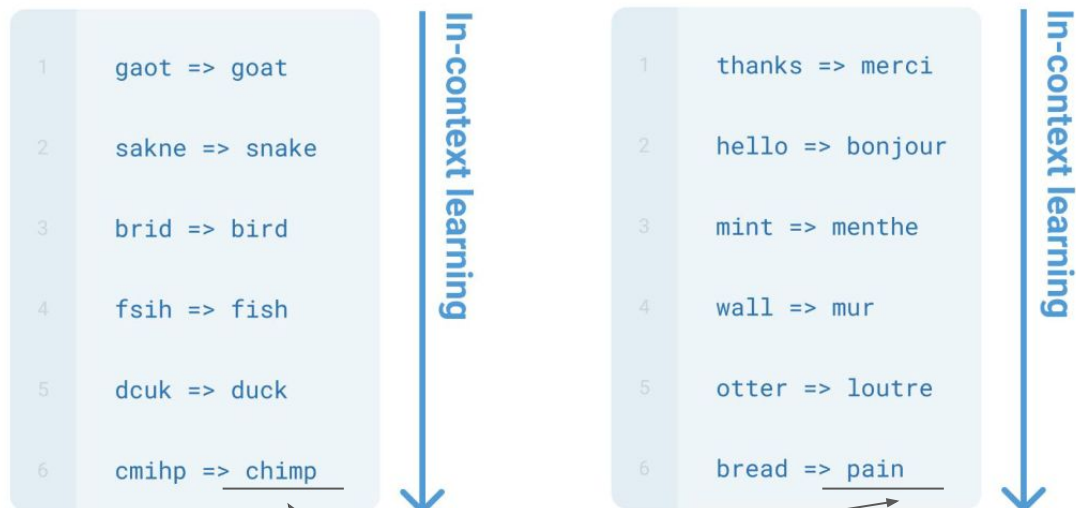
They defeated ... in the NFC Championship Game. // Sports

Apple ... development of in-house chips. // Tech

The company anticipated its operating profit to improve. // \_\_\_\_\_



# GPT-3: Prompting & In-Context Learning



*LM correctly completes the prompt!*

(Brown et al., 2020)

# Chain-of-thought prompting

## Standard Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

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

# Chain-of-thought prompting

## Standard Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

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

Model Output

A: The answer is 27. ❌



# Chain-of-thought prompting

## Standard Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

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

Model Output

A: The answer is 27. ❌

## Chain of Thought Prompting

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?

# Chain-of-thought prompting

## Standard Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

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

Model Output

A: The answer is 27. ❌

## Chain of Thought Prompting

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.***

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. ✓*

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)	

# However...

**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.

# Instruction Finetuning

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

**COMPLETION**

**Human**

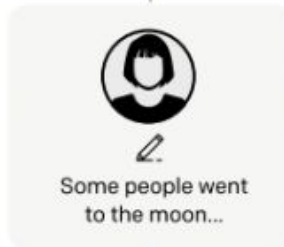
A giant rocket ship blasted off from Earth carrying astronauts to the moon. The astronauts landed their spaceship on the moon and walked around exploring the lunar surface. Then they returned safely back to Earth, bringing home moon rocks to show everyone.

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

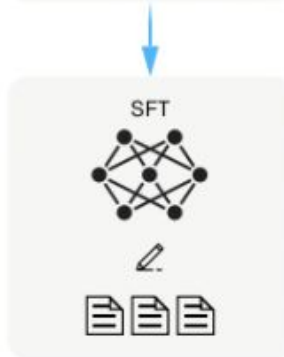
A prompt is  
sampled from our  
prompt dataset.



A labeler  
demonstrates the  
desired output  
behavior.



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





### 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)

### 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.

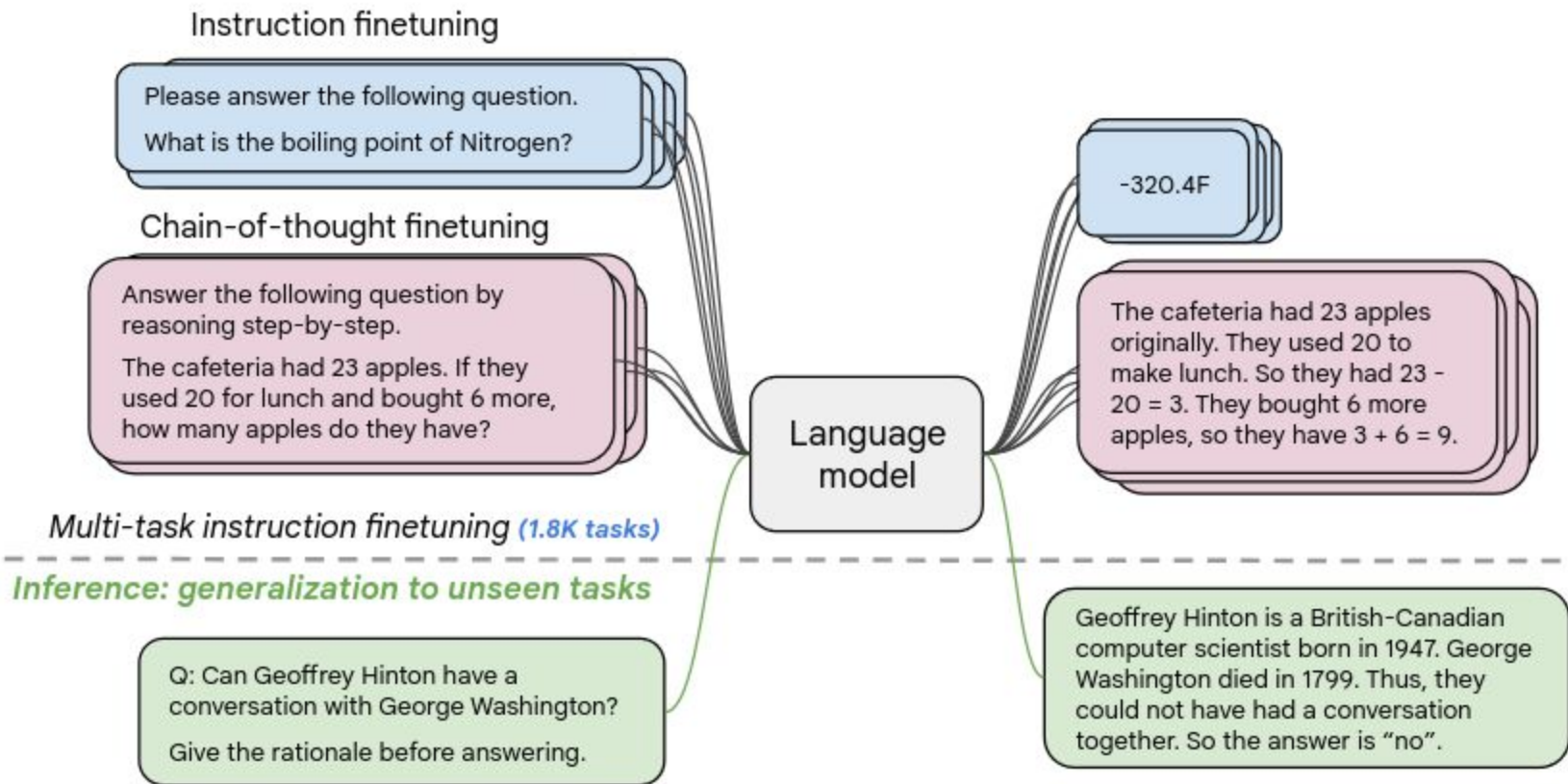
Options:

- (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.

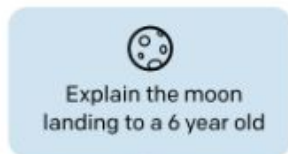
### After instruction finetuning

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

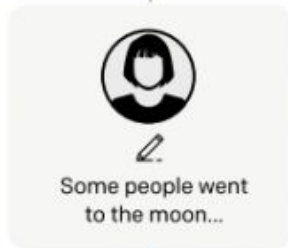


## Collect demonstration data, and train a supervised policy.

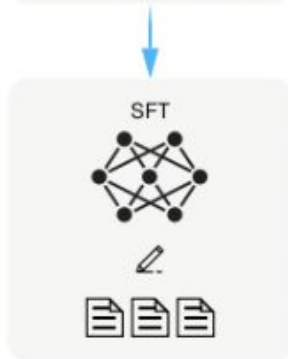
A prompt is  
sampled from our  
prompt dataset.



A labeler  
demonstrates the  
desired output  
behavior.



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



## Limitations:

- often there is no single correct way of completing a task
- getting some tokens wrong is much worse than other tokens

# Human Feedback

1. given a prompt, sample outputs from the LM

# Human Feedback

1. given a prompt, sample outputs from the LM
2. have humans rate them

Summarize the following:  
Saarland University (German:  
Universität des Saarlandes,  
pronounced) is a public research  
university located in Saarbrücken,  
the capital of the German state of  
Saarland. It was founded in 1948  
in Homburg in co-operation with  
France and is organized in six  
faculties that cover all major  
fields of...

# Human Feedback

1. given a prompt, sample outputs from the LM
2. have humans rate them

Summarize the following:  
Saarland University (German:  
Universität des Saarlandes,  
pronounced) is a public research  
university located in Saarbrücken,  
the capital of the German state of  
Saarland. It was founded in 1948  
in Homburg in co-operation with  
France and is organized in six  
faculties that cover all major  
fields of...

## Summary 1:

Saarland University is a research  
university in Saarbrücken,  
Germany.

## Summary 3:

An university in Saarland was  
founded in 1948.

## Summary 2:

The university senate has nine  
professors.

# Human Feedback

1. given a prompt, sample outputs from the LM

2. have humans rate them

Summarize the following:  
Saarland University (German:  
Universität des Saarlandes,  
pronounced) is a public research  
university located in Saarbrücken,  
the capital of the German state of  
Saarland. It was founded in 1948  
in Homburg in co-operation with  
France and is organized in six  
faculties that cover all major  
fields of...

Summary 1:

Saarland University is a research  
university in Saarbrücken,  
Germany.

*good*

Summary 3:

An university in Saarland was  
founded in 1948.

*fair*

Summary 2:

The university senate has nine  
professors.

*poor*



# Human Feedback

1. given a prompt, sample outputs from the LM
2. have humans rate them
3. tune LM to give responses given higher human ratings

# Human Feedback

1. given a prompt, sample outputs from the LM
2. have humans rate them
3. tune LM to give responses given higher human ratings

**challenge: human feedback hard to scale**

# Human Feedback

1. given a prompt, sample outputs from the LM
2. have humans rate them
3. tune LM to give responses given higher human ratings

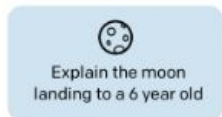
**challenge: human feedback hard to scale**

solution: train proxy reward model on human data & finetune LM for that proxy model

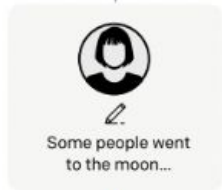
### Step 1

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

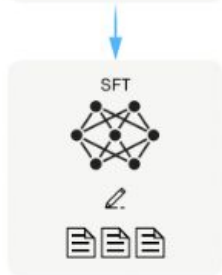
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



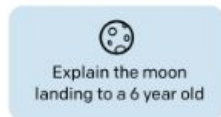
This data is used to fine-tune GPT-3 with supervised learning.



### Step 2

## Collect comparison data, and train a reward model.

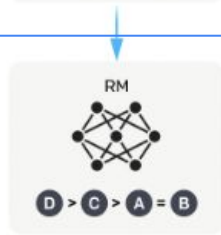
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



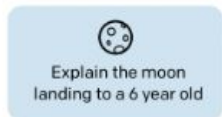
This data is used to train our reward model.



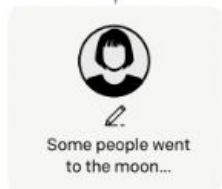
### Step 1

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

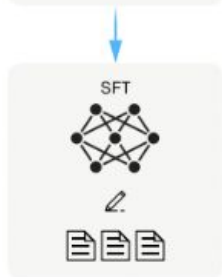
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



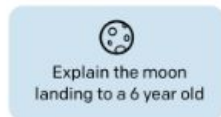
This data is used to fine-tune GPT-3 with supervised learning.



### Step 2

## Collect comparison data, and train a reward model.

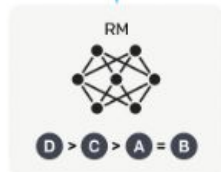
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



### 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.

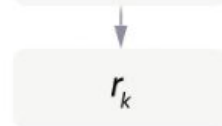


The reward model calculates a reward for the output.

Once upon a time...



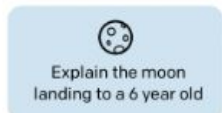
The reward is used to update the policy using PPO.



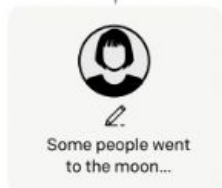
### Step 1

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

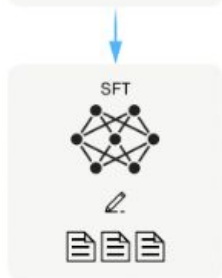
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



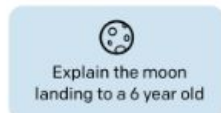
This data is used to fine-tune GPT-3 with supervised learning.



### Step 2

## Collect comparison data, and train a reward model.

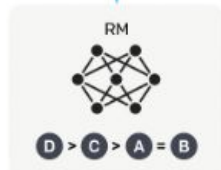
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



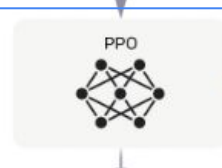
### Step 3

## 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.



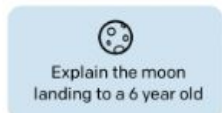
The reward is used to update the policy using PPO.



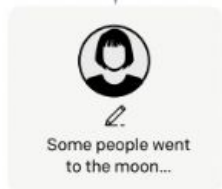
### Step 1

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

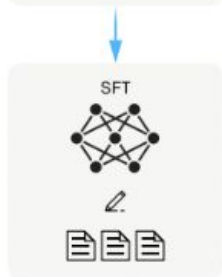
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



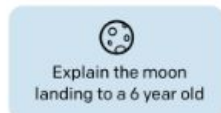
This data is used to fine-tune GPT-3 with supervised learning.



### Step 2

## Collect comparison data, and train a reward model.

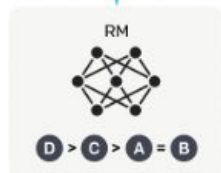
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



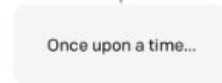
### Step 3

## Optimize a policy against the reward model using reinforcement learning.

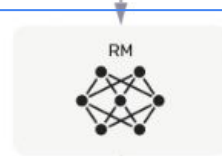
A new prompt is sampled from the dataset.



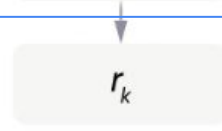
The policy generates an output.



The reward model calculates a reward for the output.



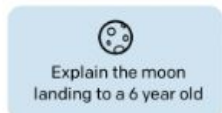
The reward is used to update the policy using PPO.



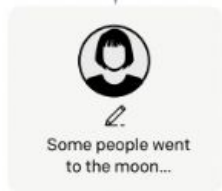
### Step 1

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

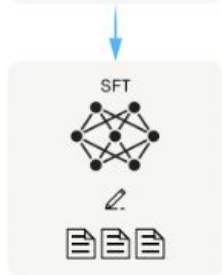
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



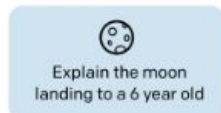
This data is used to fine-tune GPT-3 with supervised learning.



### Step 2

#### Collect comparison data, and train a reward model.

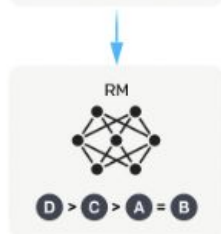
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



### Step 3

#### Optimize a policy against the reward model using reinforcement learning.

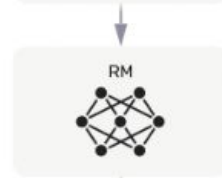
A new prompt is sampled from the dataset.



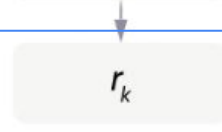
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.





**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.

**PROMPT** *Write a short poem about a wise frog.*

**COMPLETION** GPT-3

Write a short story in which a character has two different names.

Write a short story in which you try to get something back that you have lost.

Write a short story in which a character has a bad dream.

InstructGPT

The frog is a symbol of wisdom  
He knows all the secrets of the world  
He is a master of disguise  
And a great teacher of life  
He is a symbol of transformation  
And the bringer of change  
He is the frog who has seen it all  
And knows the meaning of it all

# Conclusion

Modern Large Language Models combine...

- next-word prediction
- finetuning for instructions following
- reinforcement learning for following human preferences

Thank you!