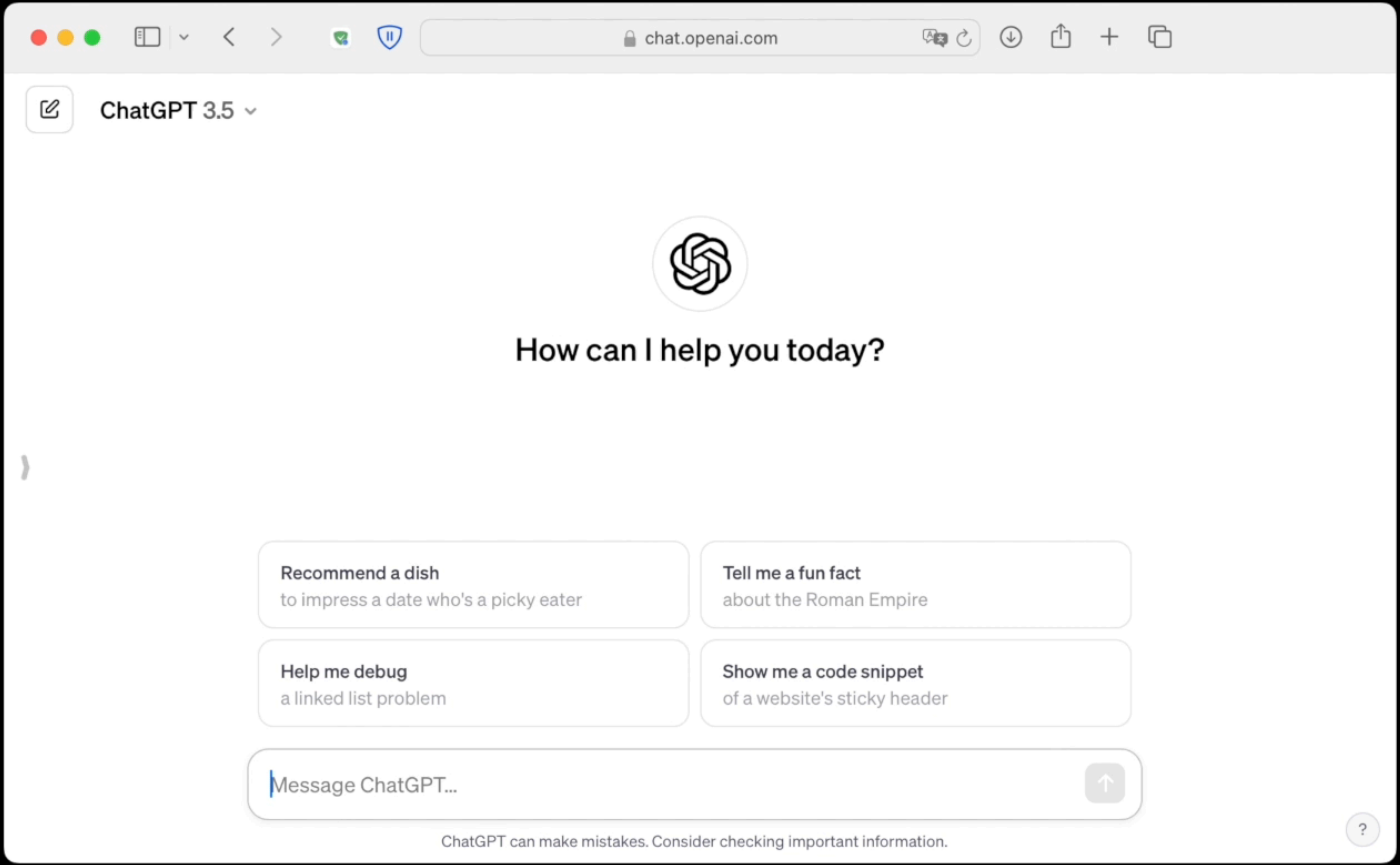
Van klanken tot woorden

Intro tot taalmodellen

Pieter Delobelle

June 11, 2024

ChatGPT



How can I help you today?

Recommend a dish
to impress a date who's a picky eater

Tell me a fun fact
about the Roman Empire

Help me debug
a linked list problem

Show me a code snippet
of a website's sticky header

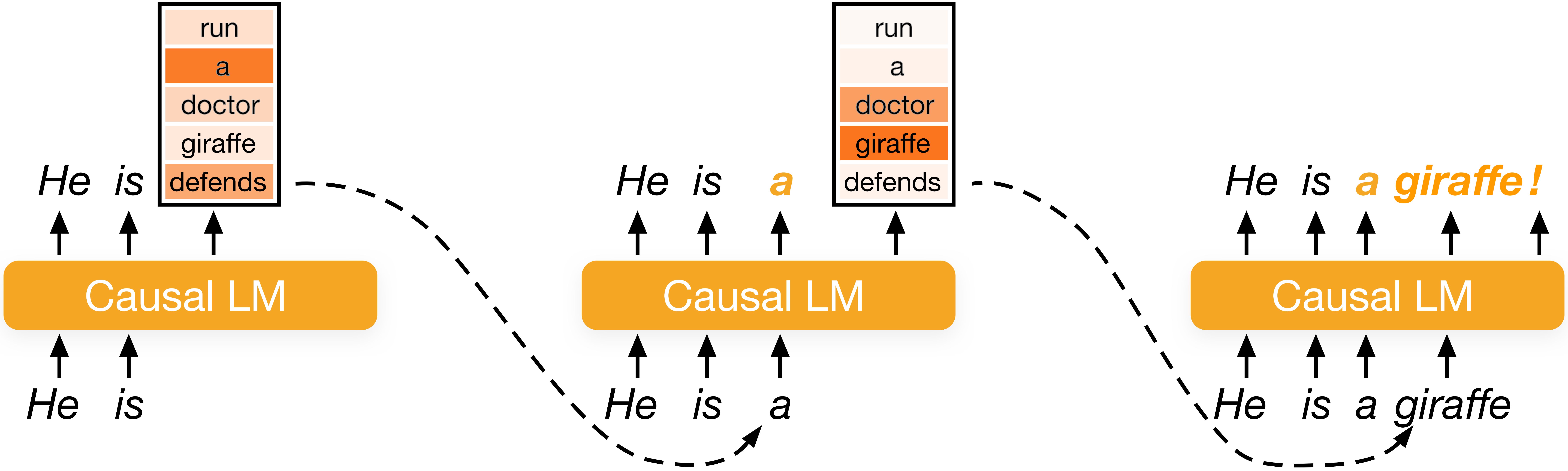
ChatGPT can make mistakes. Consider checking important information.

Hoe werkt dit?

Hoe werkt dit?

Hoe kan ik dit gebruiken?

Generating text with LMs



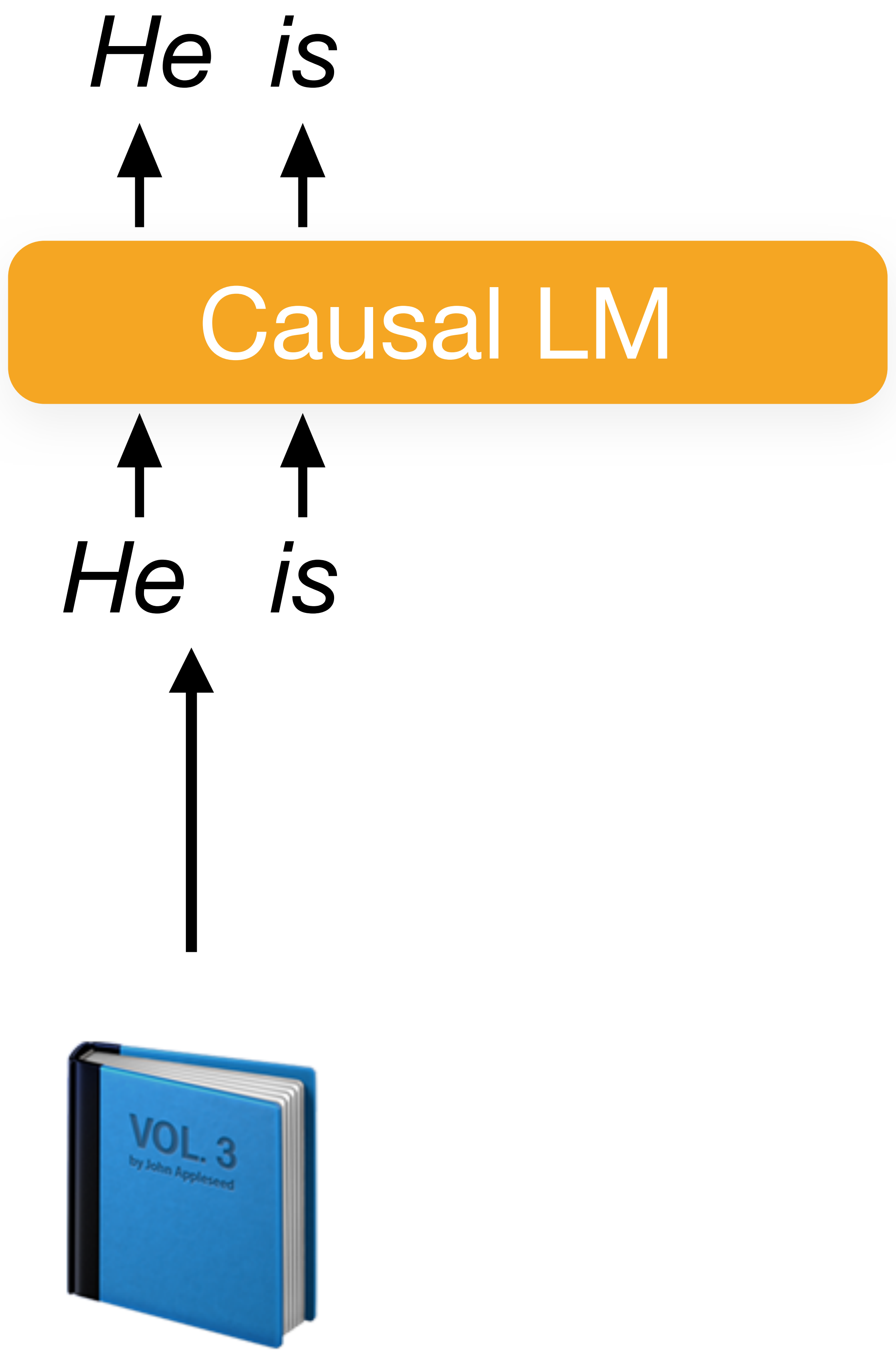
Parts of a language models

‘Heads’ of a language model
How a model predicts the next word

Attention mechanism
Each word affects the other words

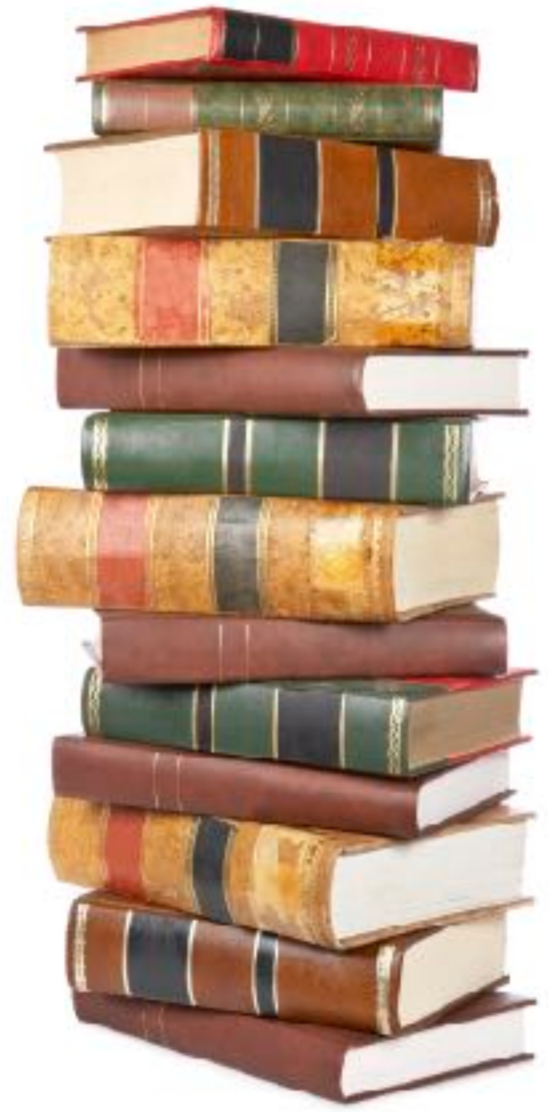
Tokenizer
How a model understands text

Training data
What a model learns



Training data

wikipedia



(copyright free) books

scraped data

Oscar corpus

Tokenizing the training data

an example

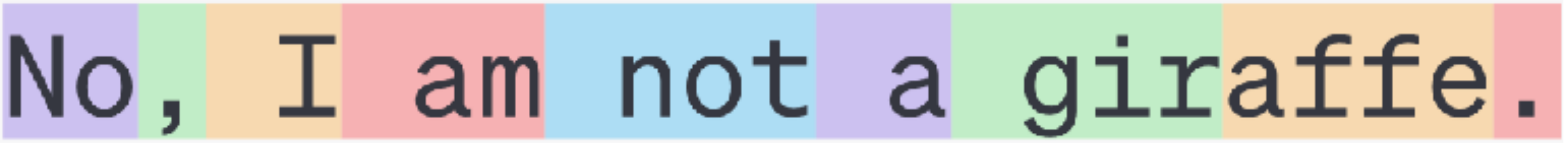
No, I am not a giraffe.

Tokenizing the training data

an example

No, I am not a giraffe.

↓



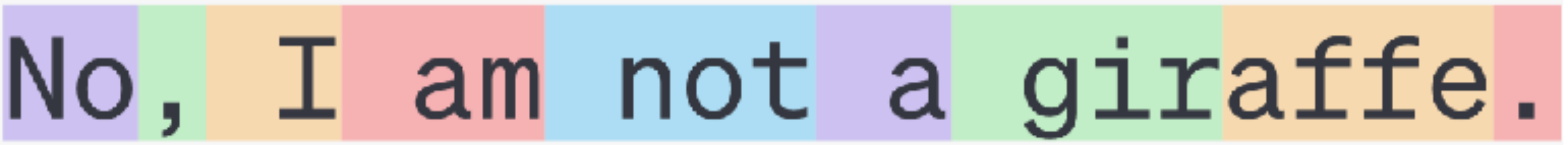
No, I am not a giraffe.

Tokenizing the training data

an example

No, I am not a giraffe.

↓



No, I am not a giraffe.

↓

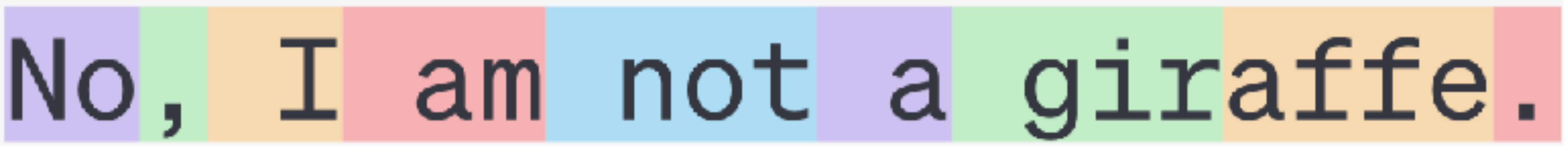
[2822, 11, 358, 1097, 539, 264, 41389, 38880, 13]

Tokenizing the training data

an example

No, I am not a giraffe.

↓

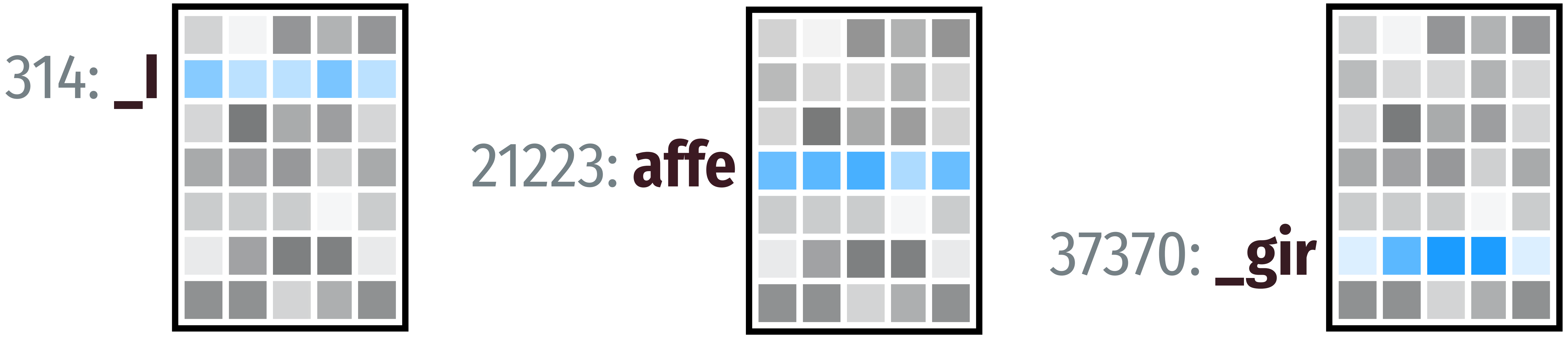


No, I am not a giraffe.

↓

[2822, 11, 358, 1097, 539, 264, 41389, 38880, 13]

↓



Embeddings capture meaning

0.9	0.1	0.1	0.5	0.4	0.1	0.0

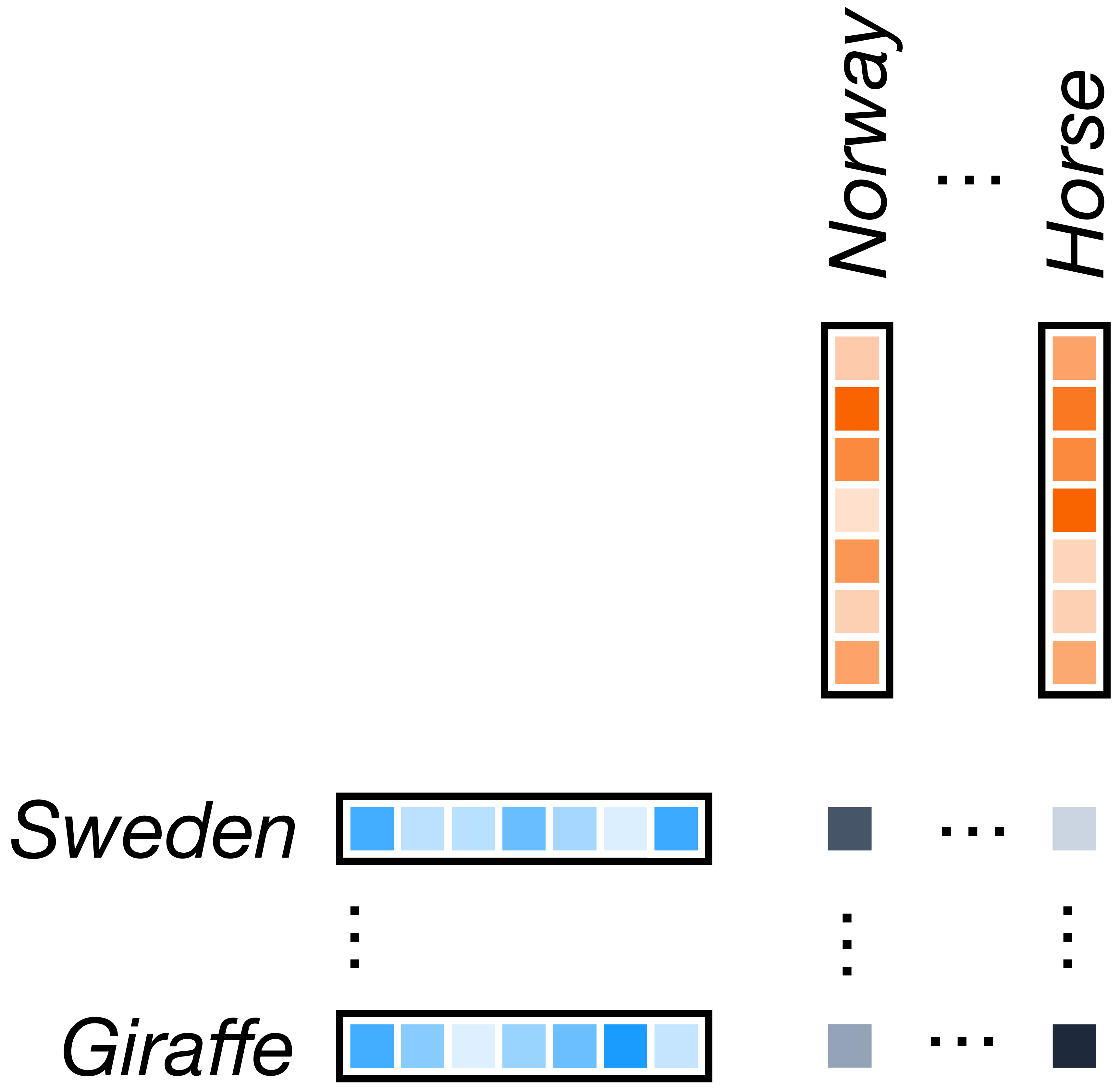
Giraffe

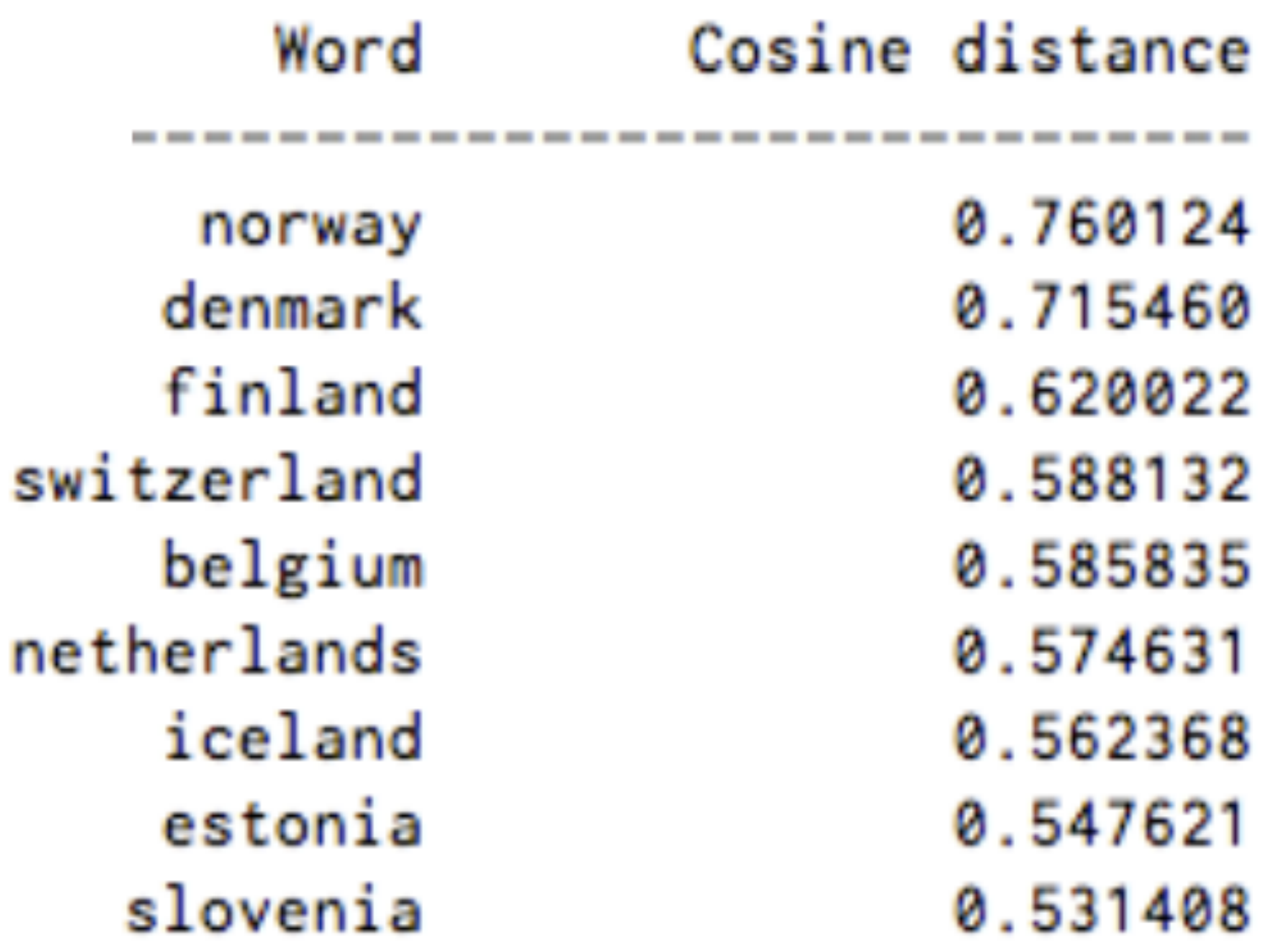
0.8	0.1	0.2	0.5	0.4	0.2	0.0

Horse

Embeddings capture meaning

Word embeddings



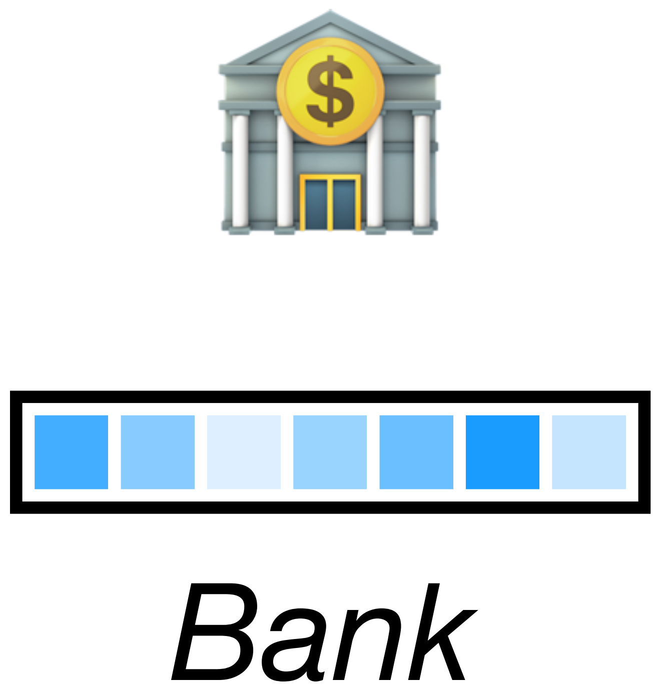


```
         Word        Cosine distance
------------------------------------
       norway               0.760124
      denmark               0.715460
      finland               0.620022
  switzerland               0.588132
      belgium               0.585835
  netherlands               0.574631
      iceland               0.562368
      estonia               0.547621
     slovenia               0.531408
```

LLMs use context to learn embeddings to address polysemy

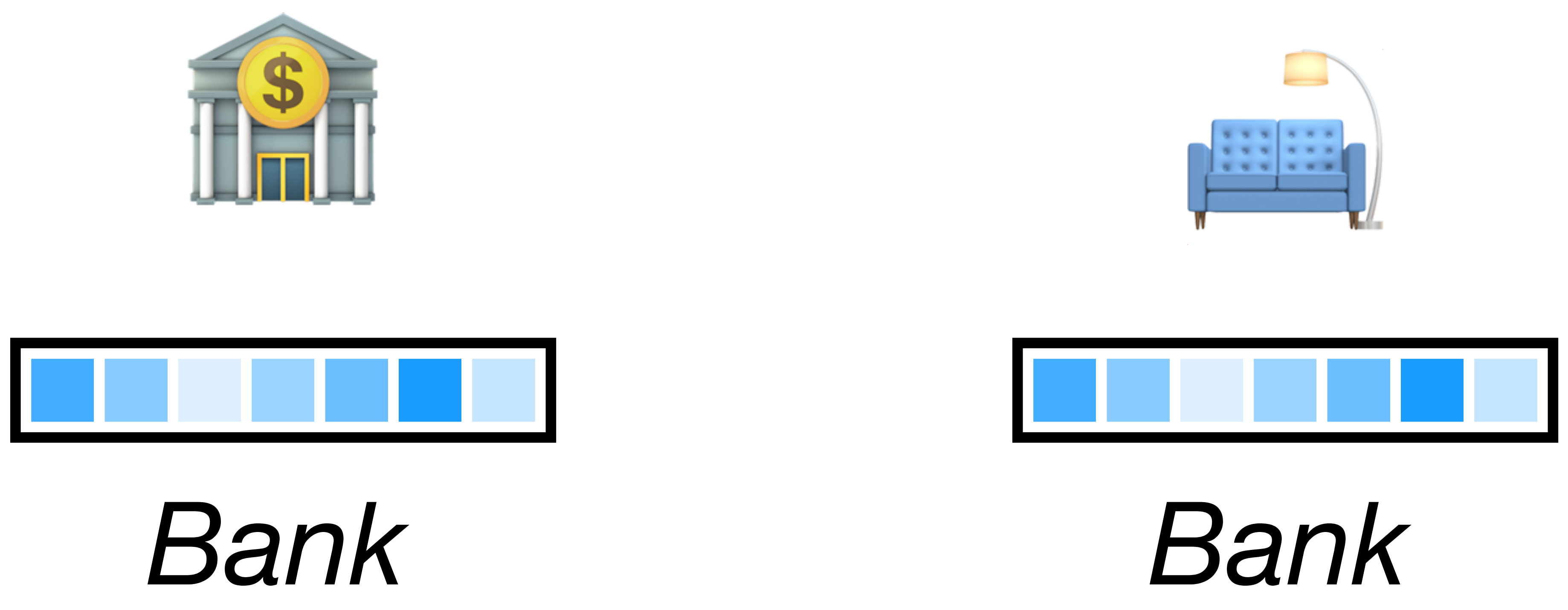
LLMs use context to learn embeddings

to address polysemy

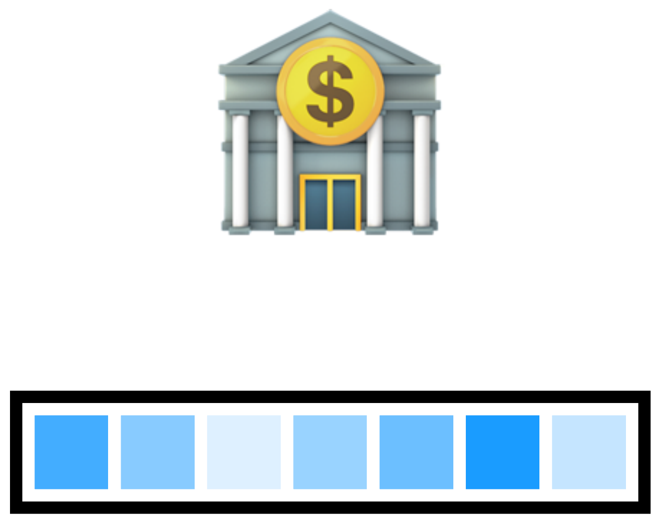


LLMs use context to learn embeddings

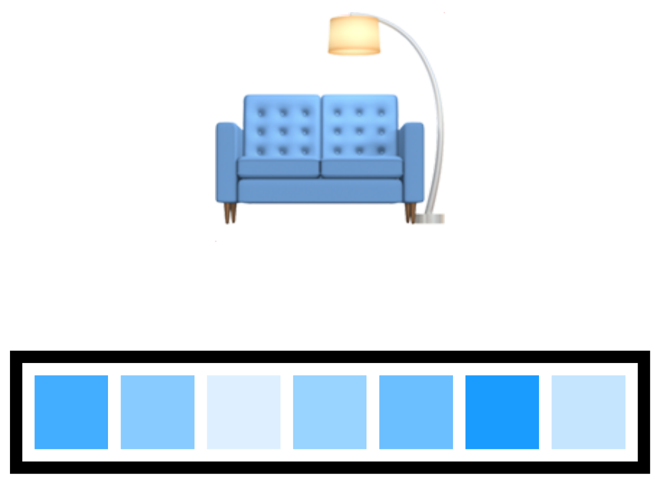
to address polysemy



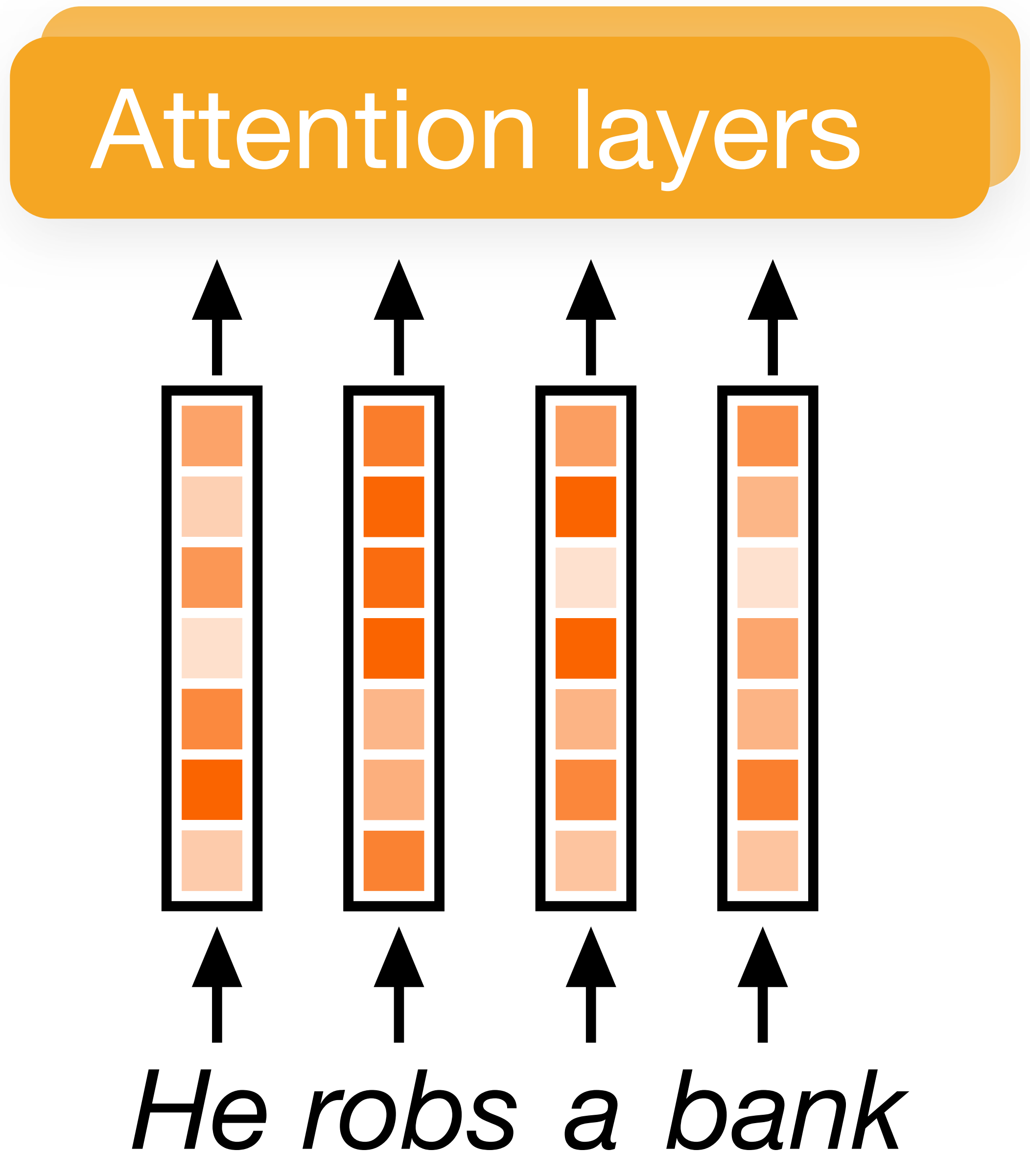
LLMs use context to learn embeddings to address polysemy



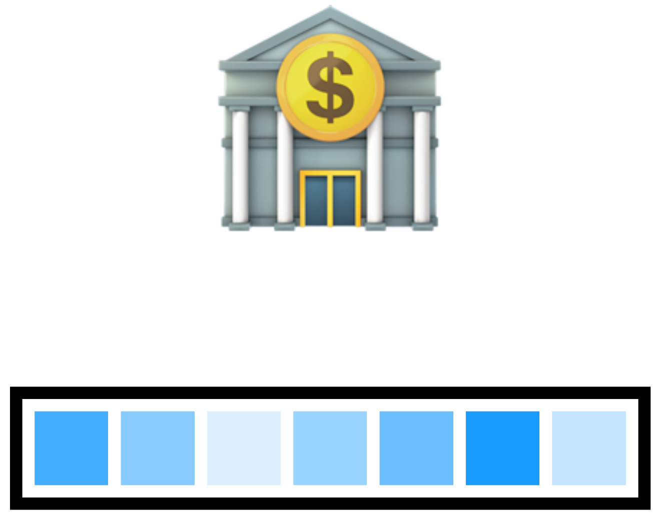
Bank



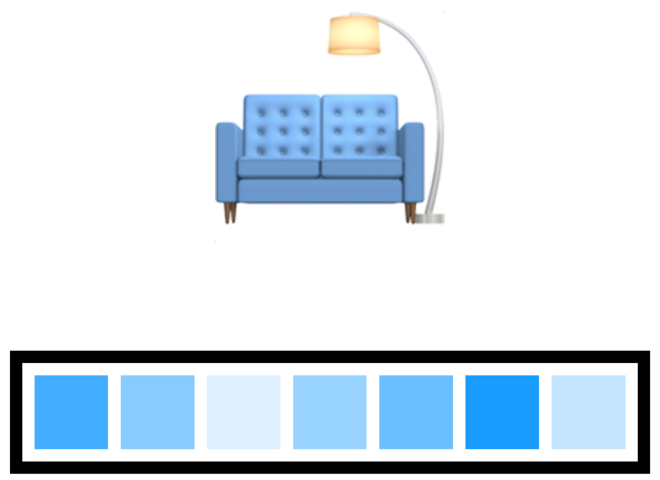
Bank



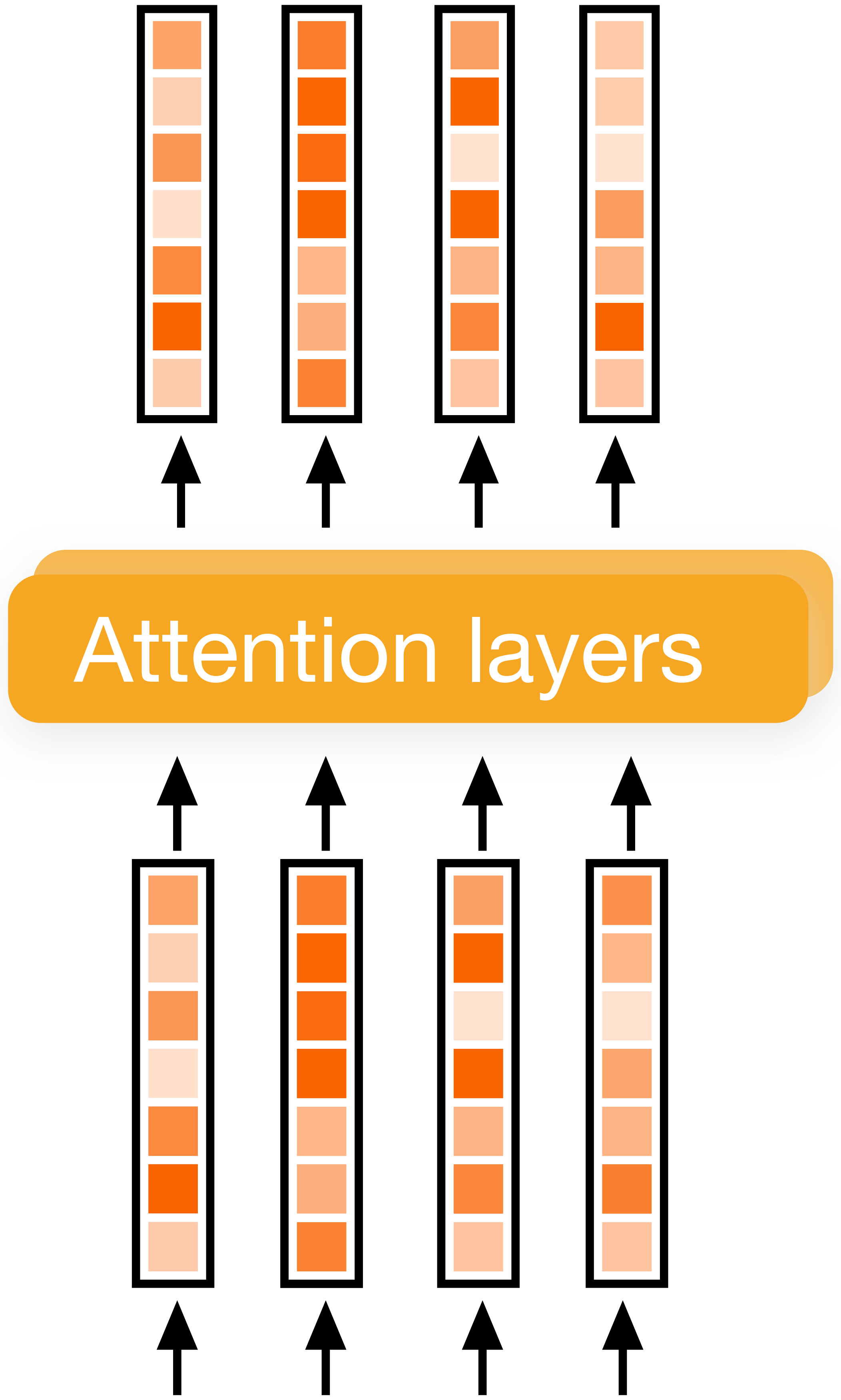
LLMs use context to learn embeddings to address polysemy



Bank



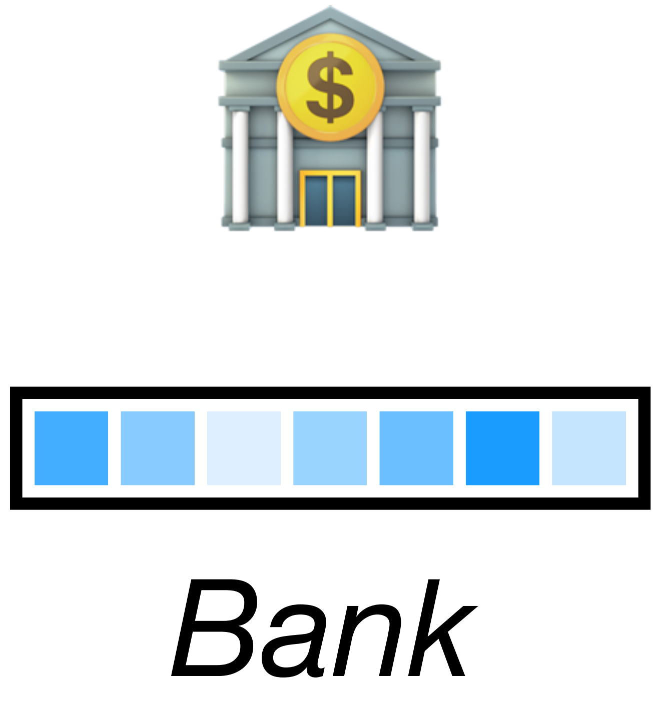
Bank

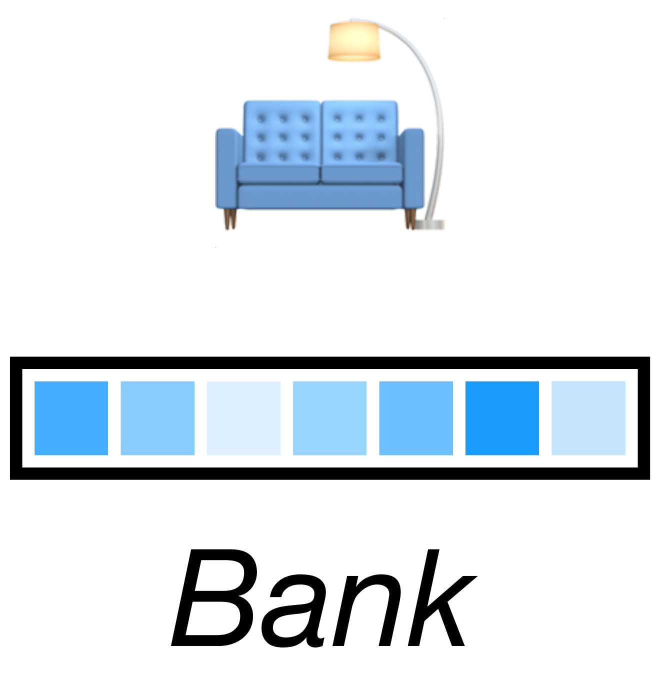


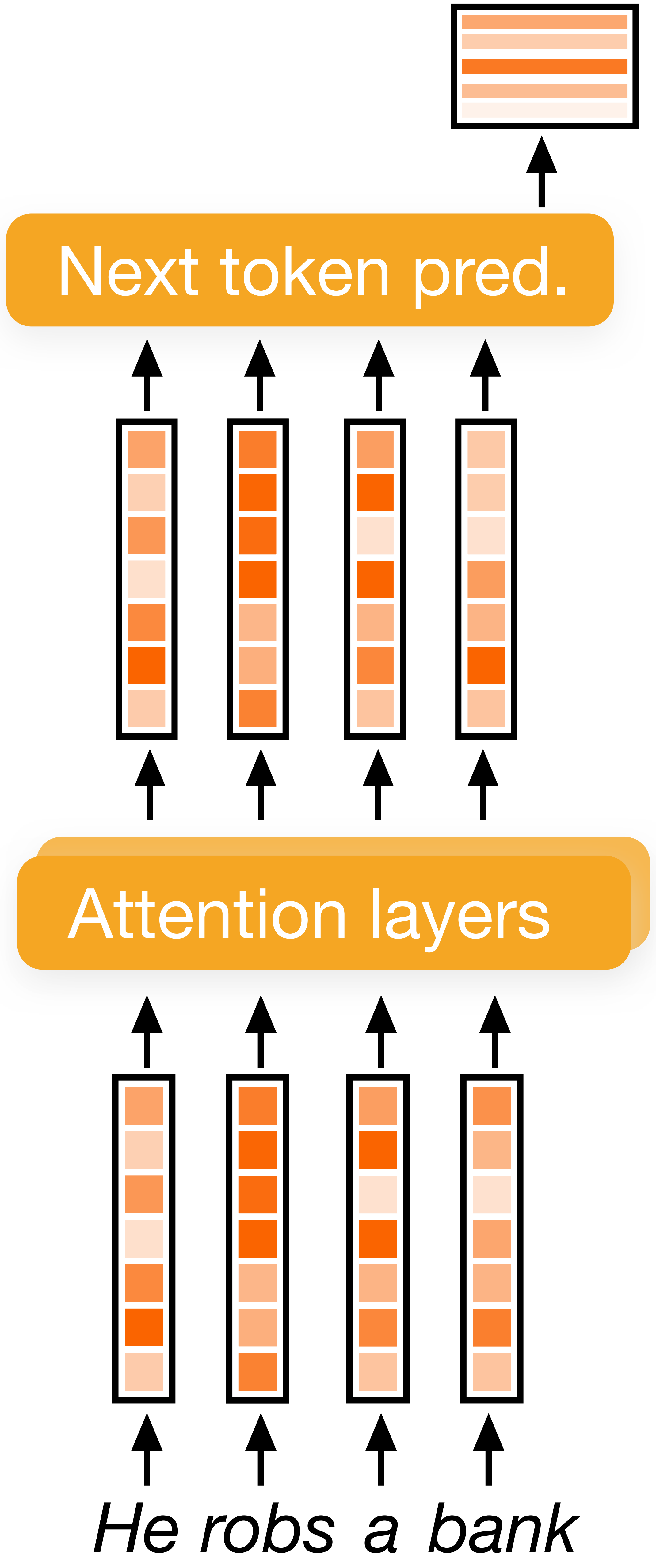
Attention layers

He robs a bank

LLMs use context to learn embeddings to address polysemy







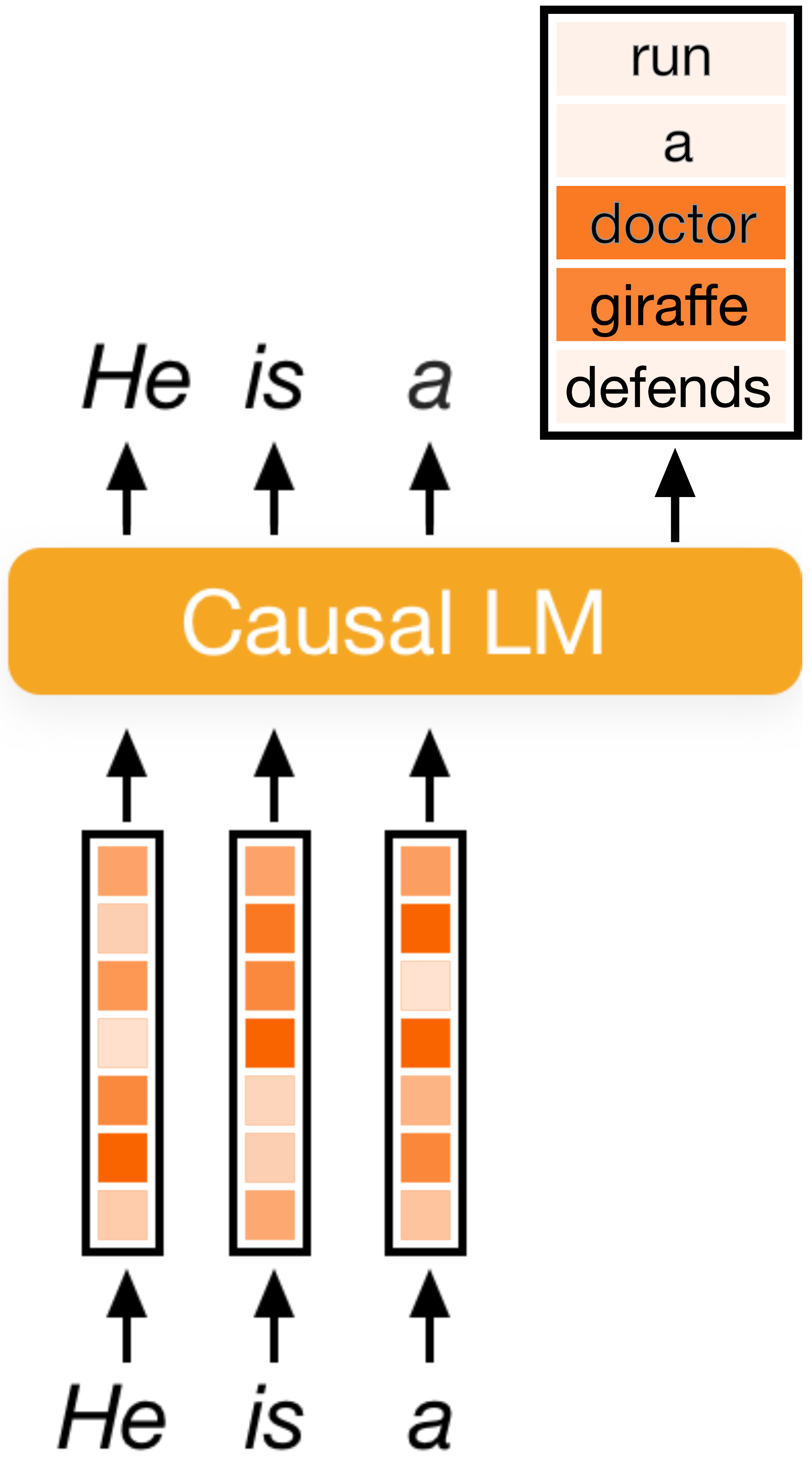
Predicting the next token

It is the tallest living terrestrial animal.

Giraffes live in herds.

He is a giraffe.

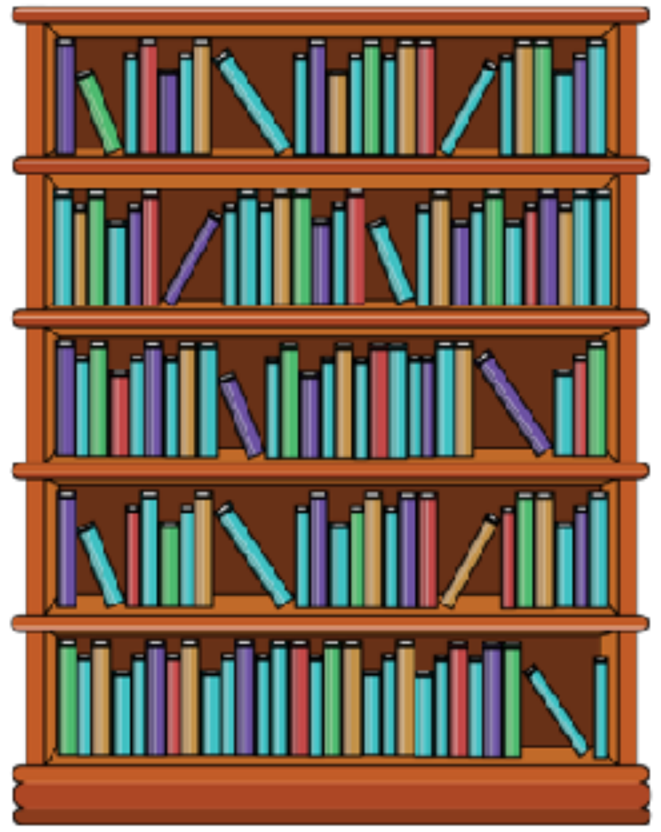
IUCN recognises one species of giraffe.



Large training corpuses are used

One book

40-50k tokens



One bookshelf

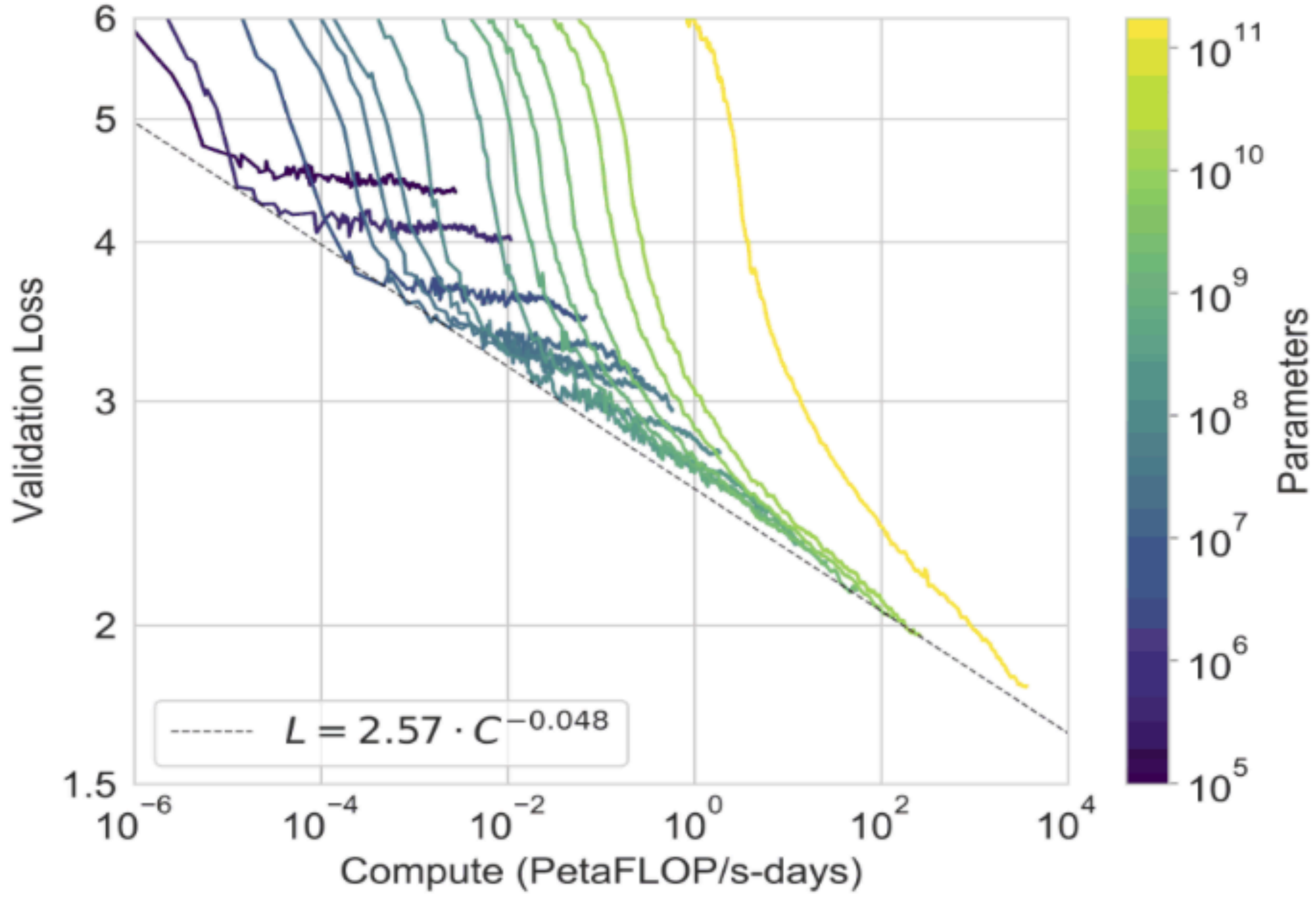
1.6M - 2.5M tokens

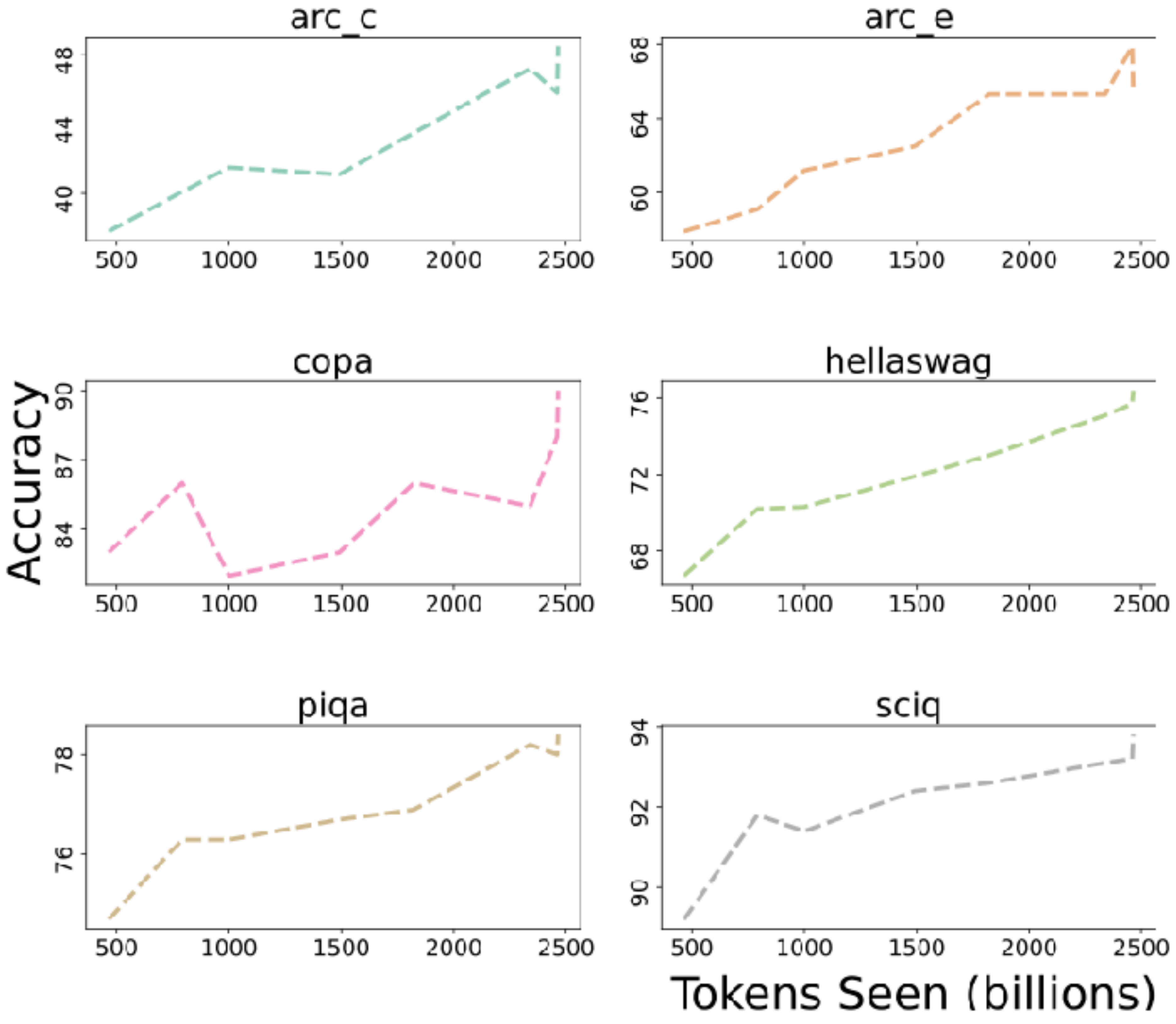
One LLM training set

2.5T - 6T tokens

~2 500 000 bookshelves

Pretraining is expensive, but worth it

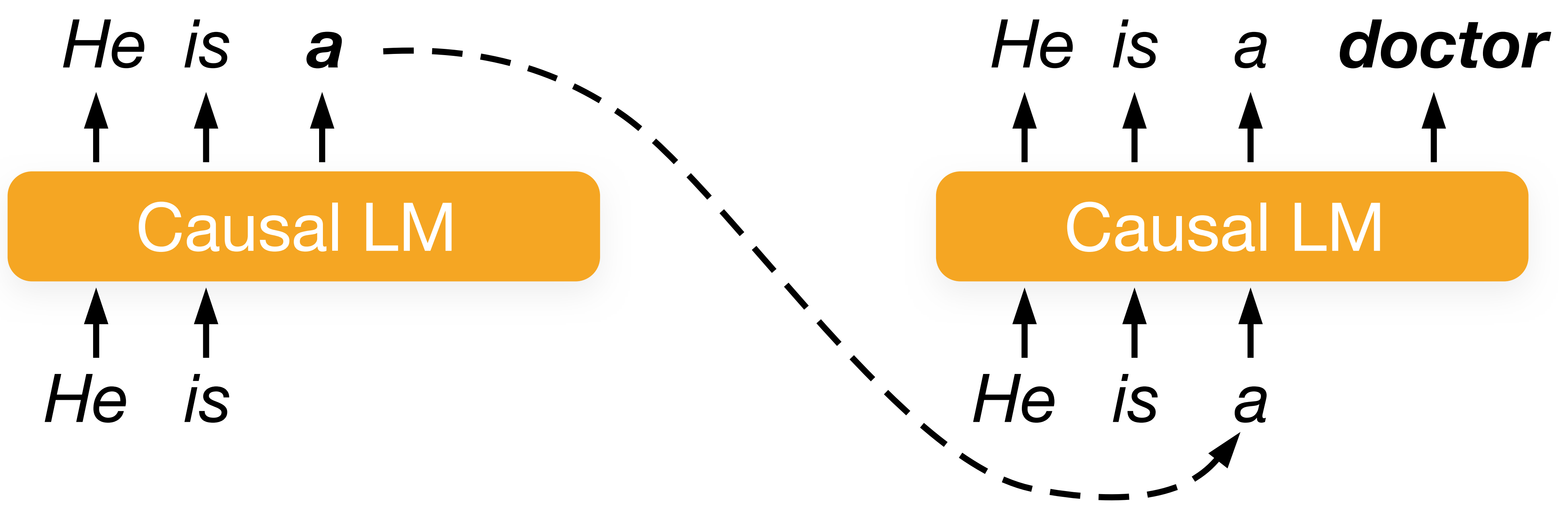




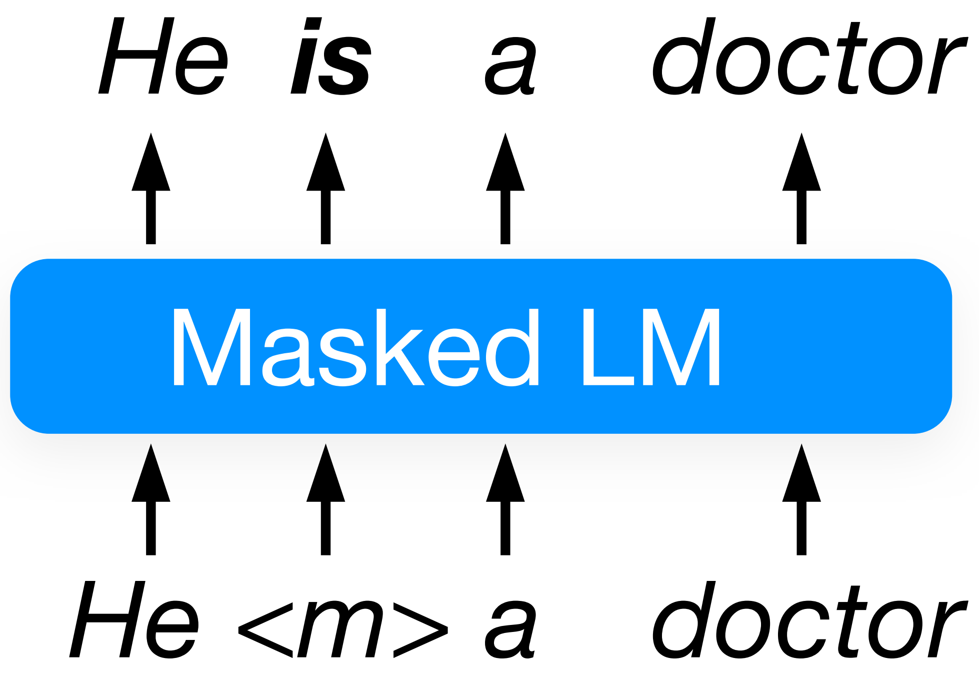
https://openai.com/research/scaling-laws-for-neural-language-models
https://allenai.org/olmo

Language modeling

1. Causal language modeling (CLM)

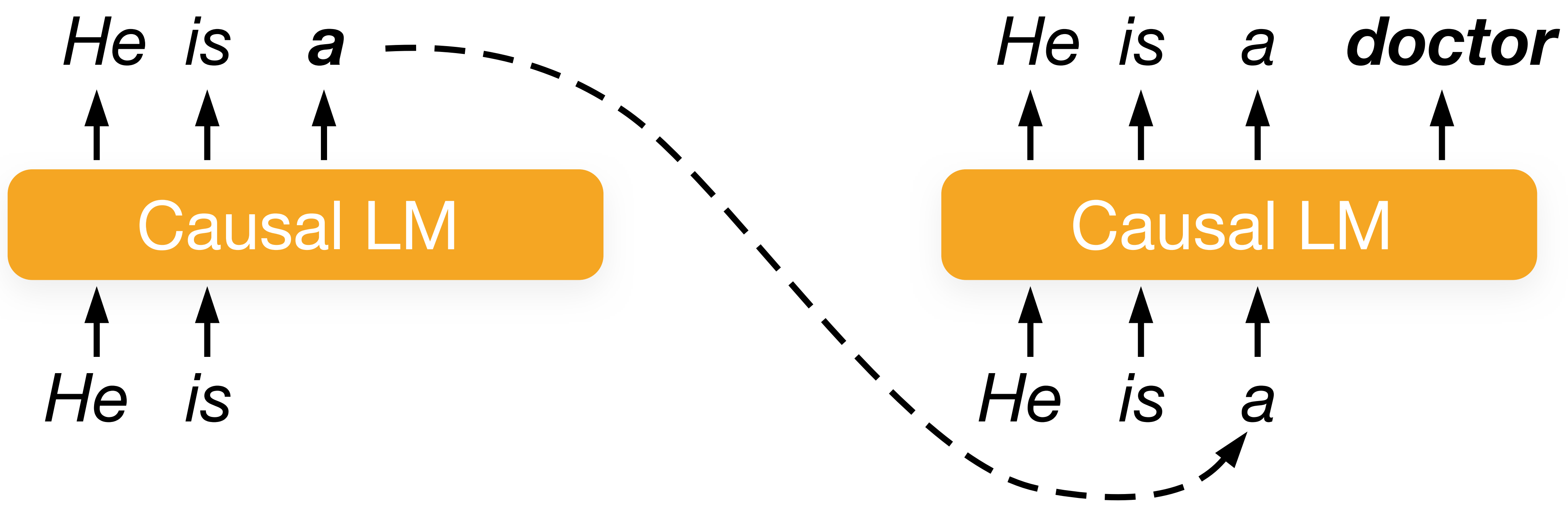


2. Masked language modeling (MLM)

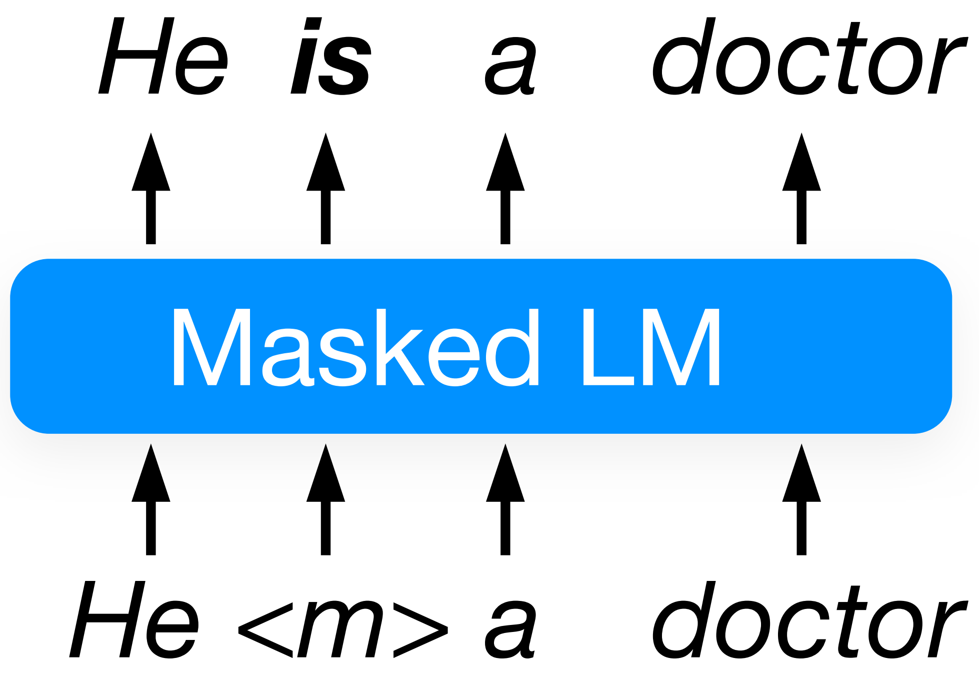


Language modeling

1. Causal language modeling (CLM)

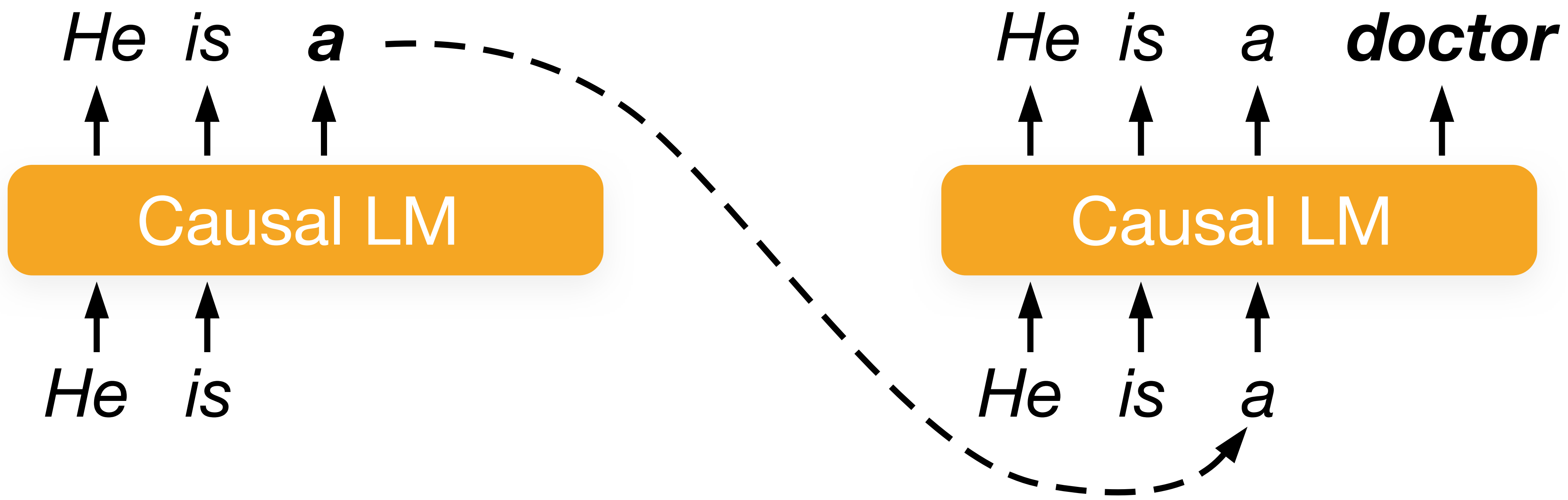


2. Masked language modeling (MLM)

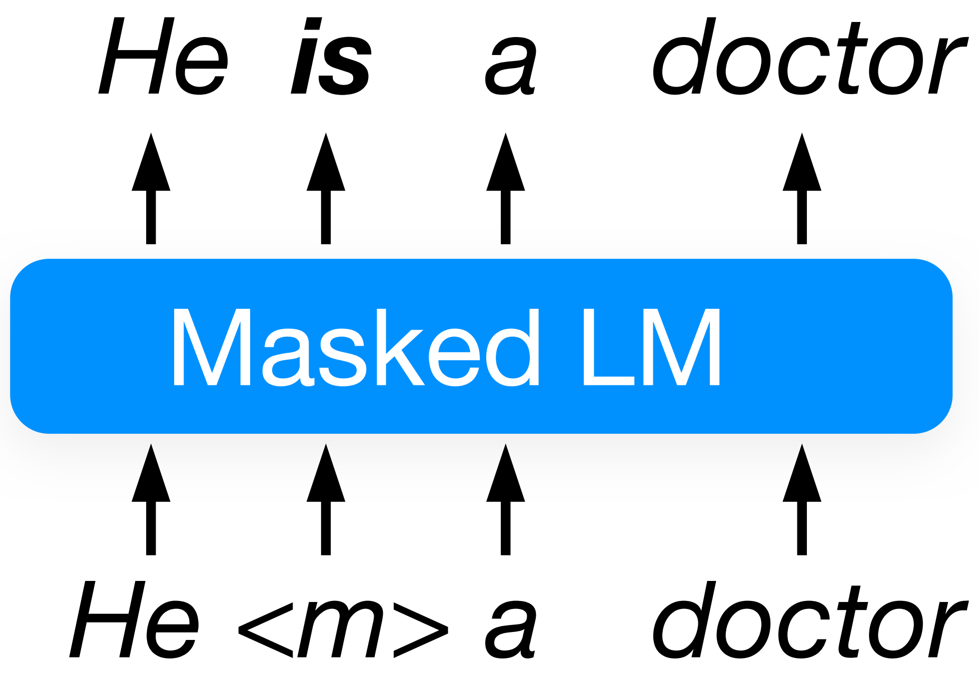


Language modeling

1. Causal language modeling (CLM)



2. Masked language modeling (MLM)



RobBERT

https://pieter.ai/robbert/

Hoe werkt dit?

Hoe kan ik dit gebruiken?

Different access modes

Closed source

No access to training data or model weights

GPT

Open model weights

No access to training data

LLAMA 2

Mistral

Open

Access to training data and model weights

olmo

Different access modes

Closed source

No access to training data or model weights

GPT

Open model weights

No access to training data

LLAMA 2

Mistral

Geitje-7b

Tweety-7b-dutch

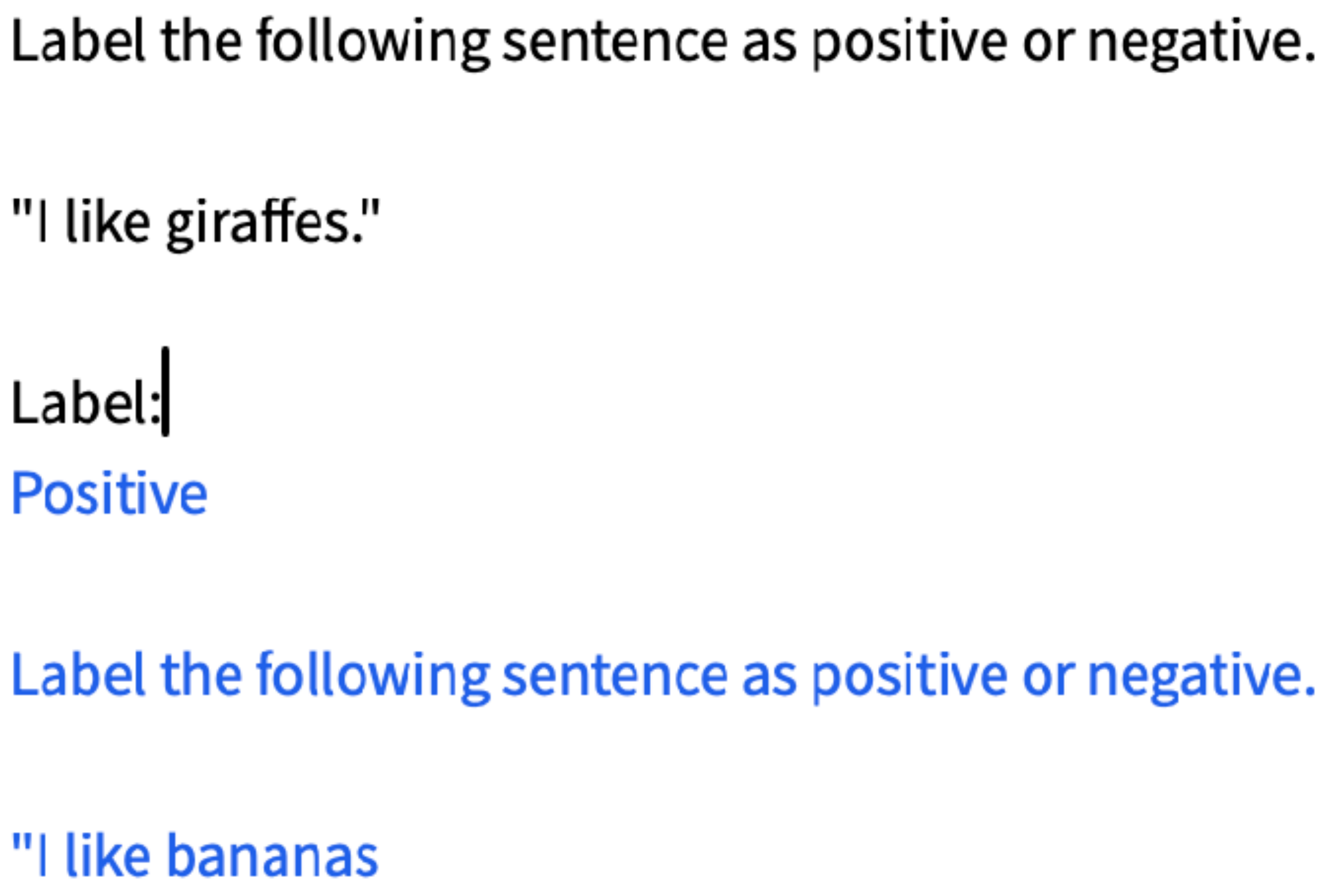
Open

Access to training data and model weights

pieter.ai/tweety-7b-dutch

Instruction tuning

Base model



Label the following sentence as positive or negative.

"I like giraffes."

Label:

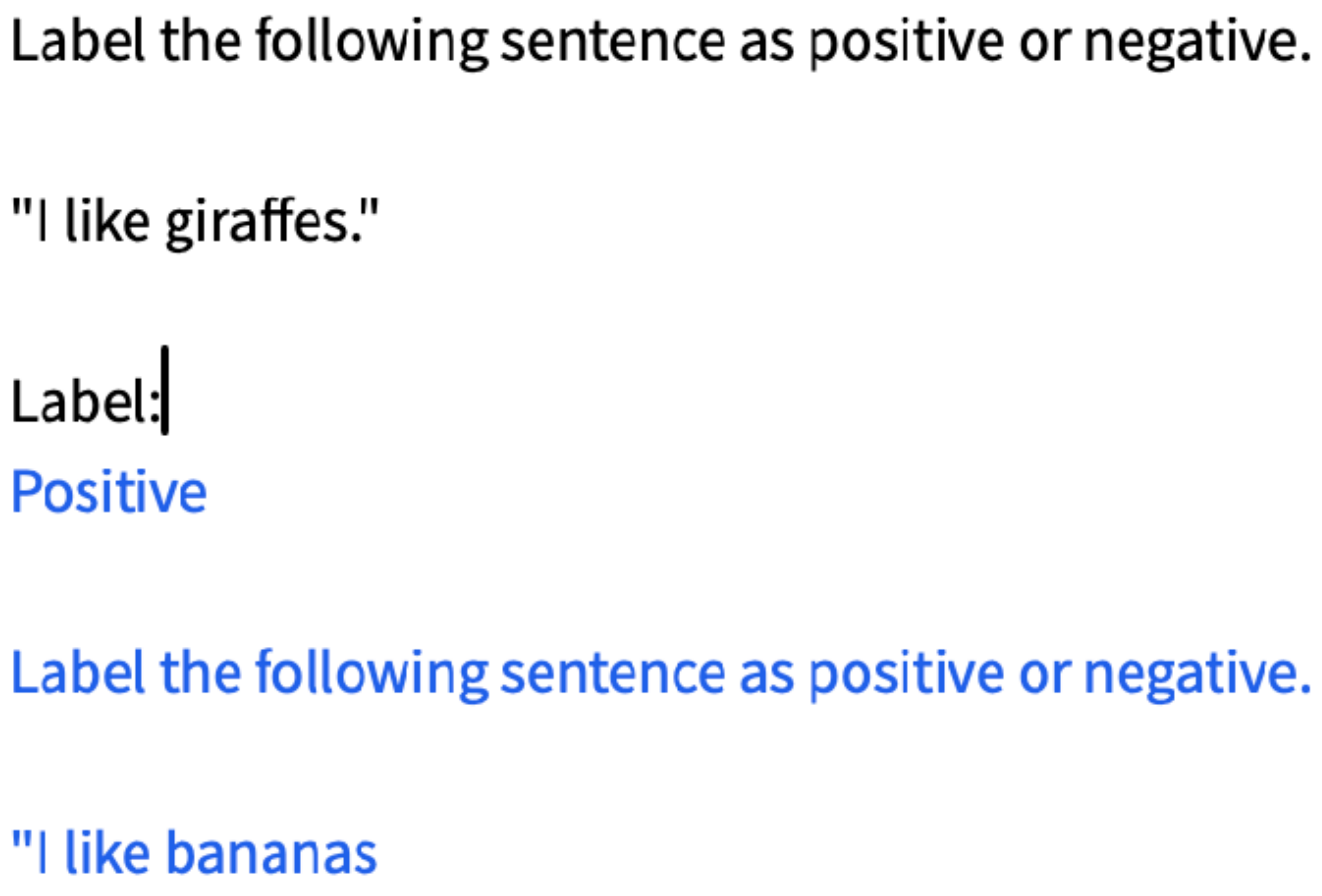
Positive

Label the following sentence as positive or negative.

"I like bananas

Instruction tuning

Base model



Label the following sentence as positive or negative.

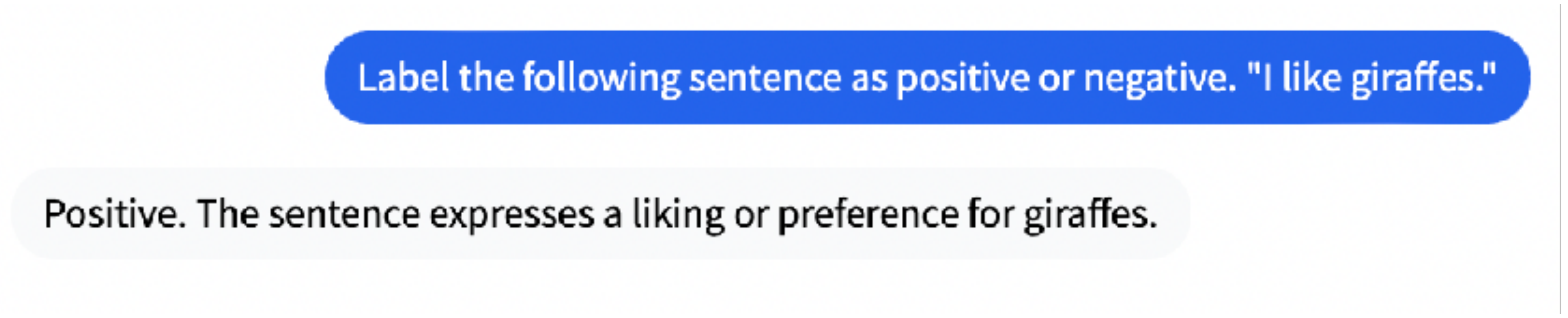
"I like giraffes."

Label:
Positive

Label the following sentence as positive or negative.

"I like bananas

Instruction-tuned model
with chat-templates



Label the following sentence as positive or negative. "I like giraffes."

Positive. The sentence expresses a liking or preference for giraffes.

```
<s>[INST] Label the following sentence as positive
or negative... [/INST]"
"Well, Positive. The sentence expresses a liking
for …</s> "
"[INST] And this sentence: “…” [/INST]
```

Van klanken tot woorden

Intro tot taalmodellen

Pieter Delobelle

June 11, 2024

slides: **pieter.ai/appearances.html**