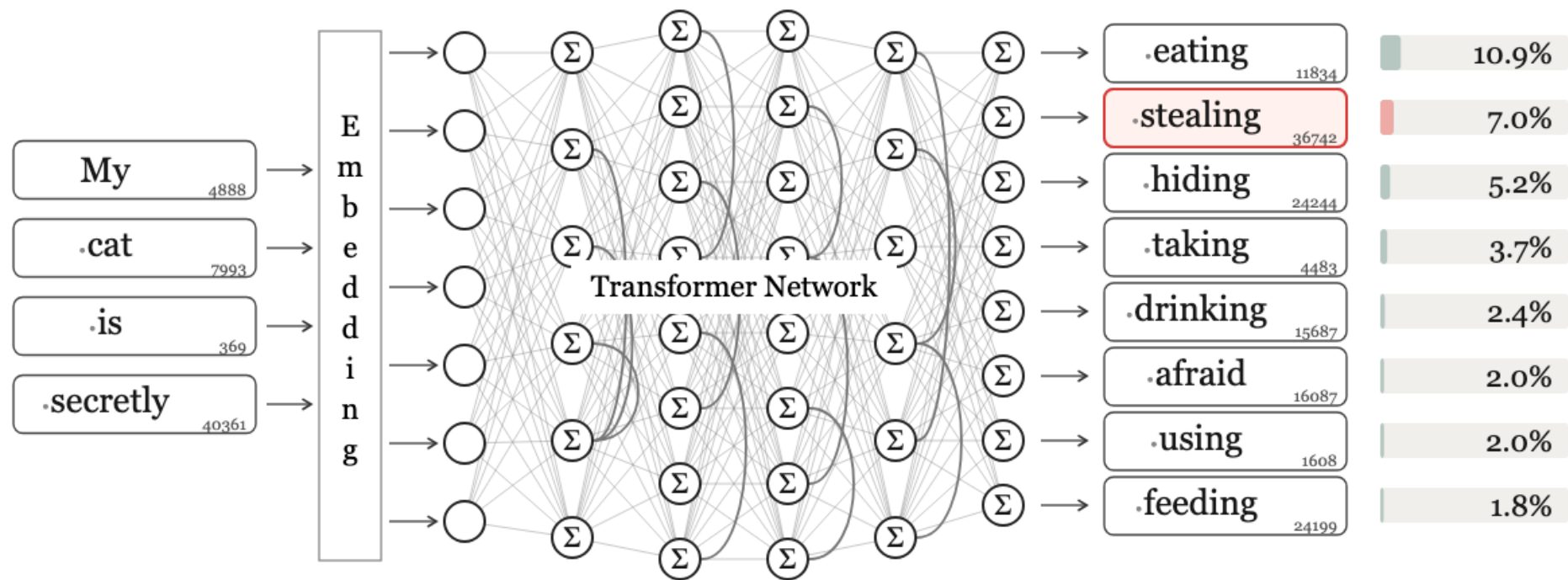




L-3

Large Language Models



DEMO 1 · L2 Exit Ticket Results

Language



Language

Our Most Powerful Invention

Language: The Meta-Tool

Our Most Powerful Tool

Foundation for human progress and innovation

Language: The Meta-Tool

Our Most Powerful Tool

Foundation for human progress and innovation

Meta-Tool: A Tool for Creating Other Tools

Enables **creation of new tools**

Allows **sharing of knowledge**

Supports refinement and **improvement over time**

Language: The Meta-Tool

Our Most Powerful Tool

Foundation for human progress and innovation

Meta-Tool: A Tool for Creating Other Tools

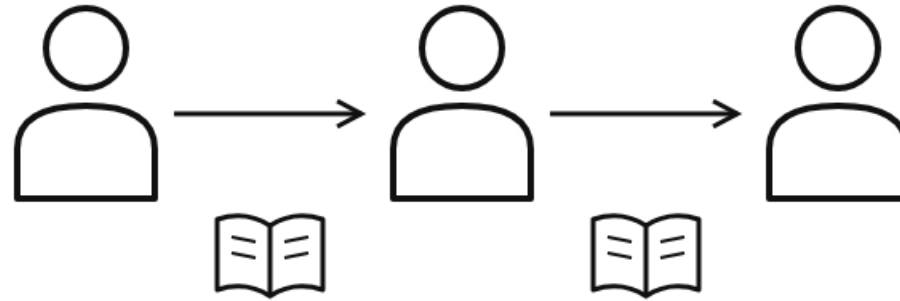
Enables **creation of new tools**

Allows **sharing of knowledge**

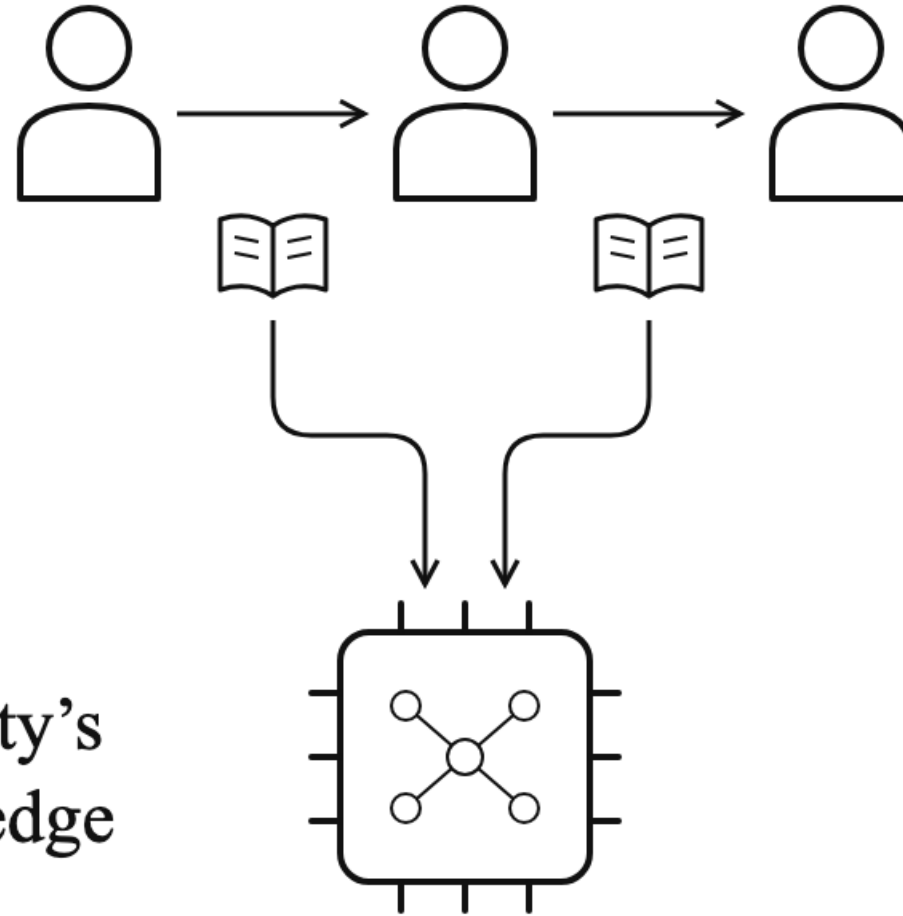
Supports refinement and **improvement over time**

From mastering fire to landing on the Moon,
language lets us build on each other's ideas

Language lets us pass knowledge across generations.

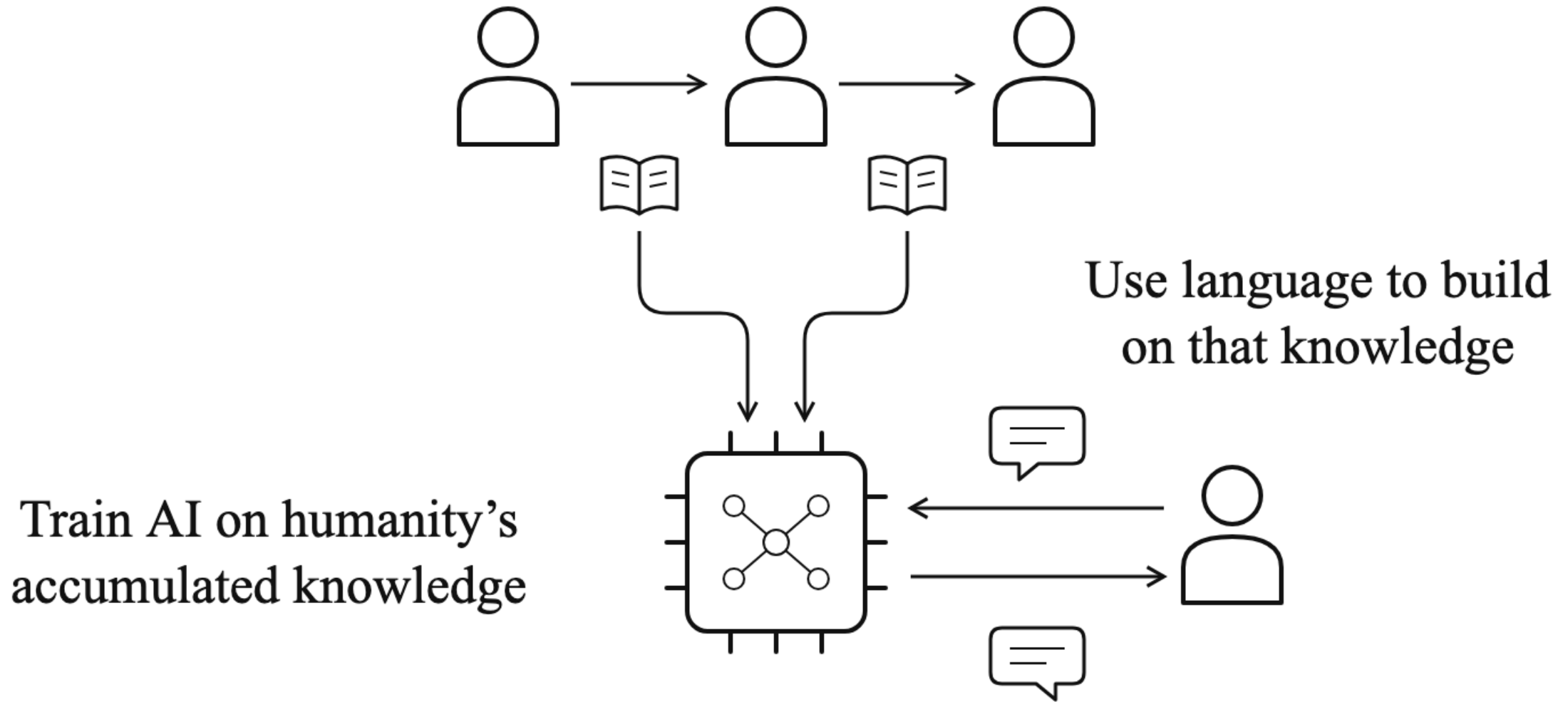


Language lets us pass knowledge across generations.



Train AI on humanity's
accumulated knowledge

Language lets us pass knowledge across generations.



Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Ambiguous

Context, tone, and culture shape meaning

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Ambiguous

Context, tone, and culture shape meaning

“I saw her duck.”

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Ambiguous

Context, tone, and culture shape meaning

“I saw her duck.”

Creative

New words, slang, evolving structures

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Ambiguous

Context, tone, and culture shape meaning

“I saw her duck.”

Creative

New words, slang, evolving structures

“My brain has 47 tabs open, and I can’t find which one is playing music.”

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Ambiguous

Context, tone, and culture shape meaning

“I saw her duck.”

Creative

New words, slang, evolving structures

“My brain has 47 tabs open, and I can’t find which one is playing music.”

Layered

Emotional, cultural, and social dimensions

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Ambiguous

Context, tone, and culture shape meaning

“I saw her duck.”

Creative

New words, slang, evolving structures

“My brain has 47 tabs open, and I can’t find which one is playing music.”

Layered

Emotional, cultural, and social dimensions

“**YOU** ate my cake?”

→ You, of all people?

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Ambiguous

Context, tone, and culture shape meaning

“I saw her duck.”

Creative

New words, slang, evolving structures

“My brain has 47 tabs open, and I can’t find which one is playing music.”

Layered

Emotional, cultural, and social dimensions

“**YOU** ate my cake?”

→ You, of all people?

“You **ATE** my cake?”

→ You ate it rather than saving it?

Why Is Language Difficult?

Complex

Irregularities, exceptions, idioms

“I broke the ice.”

Ambiguous

Context, tone, and culture shape meaning

“I saw her duck.”

Creative

New words, slang, evolving structures

“My brain has 47 tabs open, and I can’t find which one is playing music.”

Layered

Emotional, cultural, and social dimensions

“**YOU** ate my cake?”

→ You, of all people?

“You **ATE** my cake?”

→ You ate it rather than saving it?

“You ate **MY** cake?”

→ Mine, rather than yours?

How Do Machines Learn Language?

How Do Machines Learn Language?

Fundamental Idea: **Predict** What Comes Next

How Do Machines Learn Language?

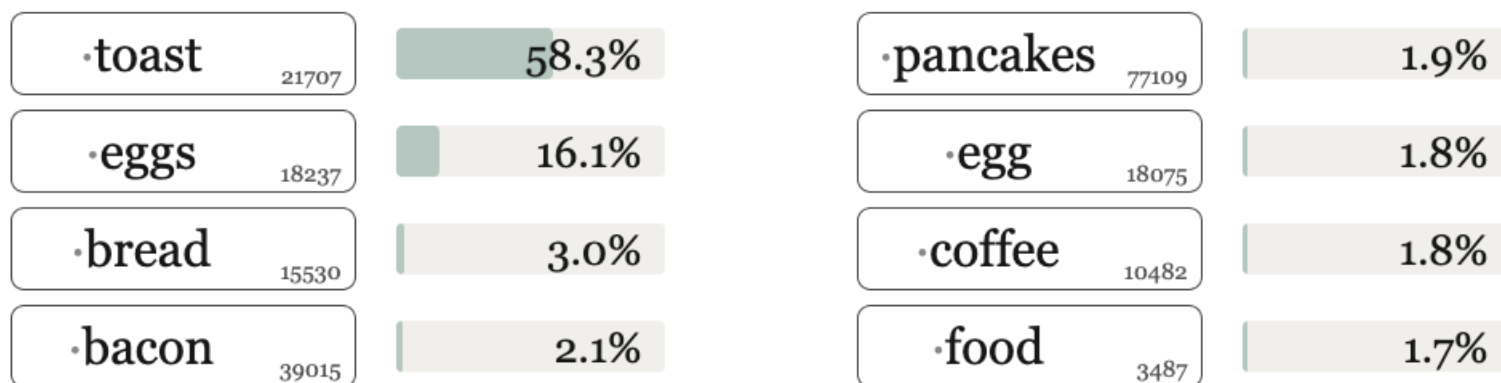
Fundamental Idea: **Predict** What Comes Next

The robot made breakfast, but burned the ...

How Do Machines Learn Language?

Fundamental Idea: **Predict** What Comes Next

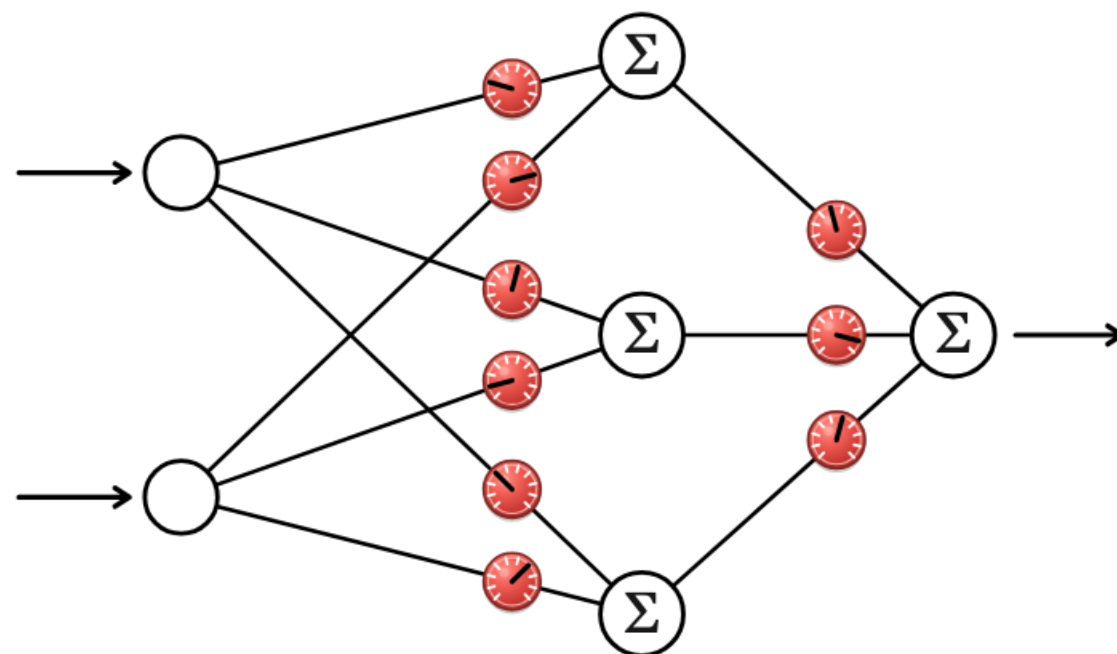
The robot made breakfast, but burned the ...



Next-token probabilities

Different Inputs, Different Outputs

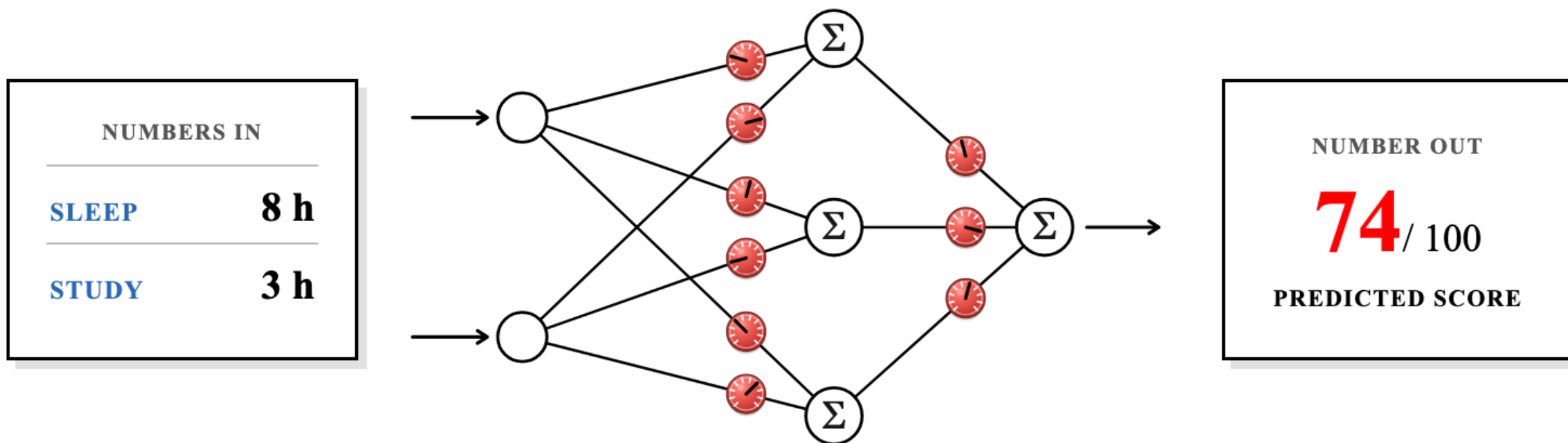
NUMBERS IN	
SLEEP	8 h
STUDY	3 h



NUMBER OUT	
74	/ 100
PREDICTED SCORE	

OUR EXAMPLE numbers $\rightarrow$ predicted score

Different Inputs, Different Outputs

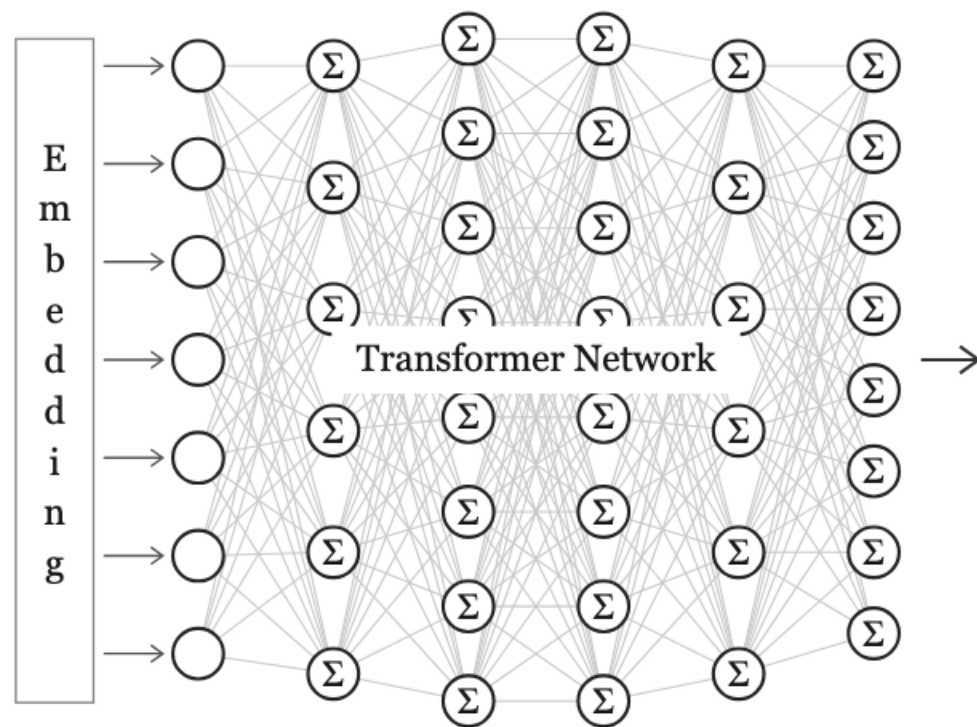


Last week: **predict** a score

This week: **predict** what comes next in text

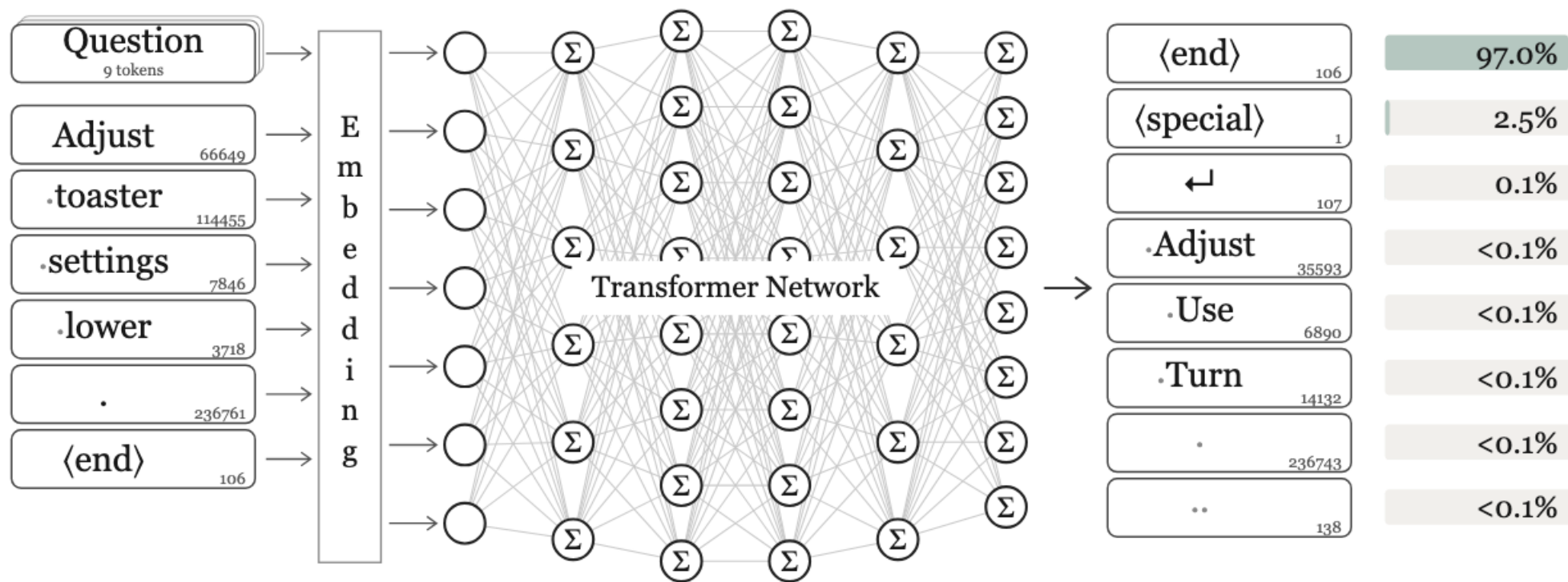
From One Prediction to a Reply

How 3910	·can 740	·I 564	·stop 4721	·my 1041	·toast 40233	·from 699	·burning 18830	? 236881
-------------	-------------	-----------	---------------	-------------	-----------------	--------------	-------------------	-------------



From One Prediction to a Reply

How₃₉₁₀ ·can₇₄₀ ·I₅₆₄ ·stop₄₇₂₁ ·my₁₀₄₁ ·toast₄₀₂₃₃ ·from₆₉₉ ·burning₁₈₈₃₀ ?₂₃₆₈₈₁



Qwen3.5-9B Base

Vocabulary: 248,320 tokens

DEMO 2 · Tokenization

What Is a Token?

unbelievable unbelievably unbelievable!

How many tokens?

What Is a Token?

unbelievable unbelievably unbelievable!

How many tokens?

un belie vable .unbelie vably .unbelievable !

7 tokens

Tokenization varies by model · Qwen3.5-9B Base

What Is a Token?

unbelievable unbelievably unbelievable!

How many tokens?

un belie vable .unbelie vably .unbelievable !

7 tokens

Tokenization varies by model · Qwen3.5-9B Base

A token is a piece of text:
a whole word, part of a word, or punctuation

What Is a Token?

unbelievable unbelievably unbelievable!

How many tokens?



7 tokens

Tokenization varies by model · Qwen3.5-9B Base

A token is a piece of text:
a whole word, part of a word, or punctuation

Each token has an ID that the model uses

What Is a Token?

unbelievable unbelievably unbelievable!

How many tokens?



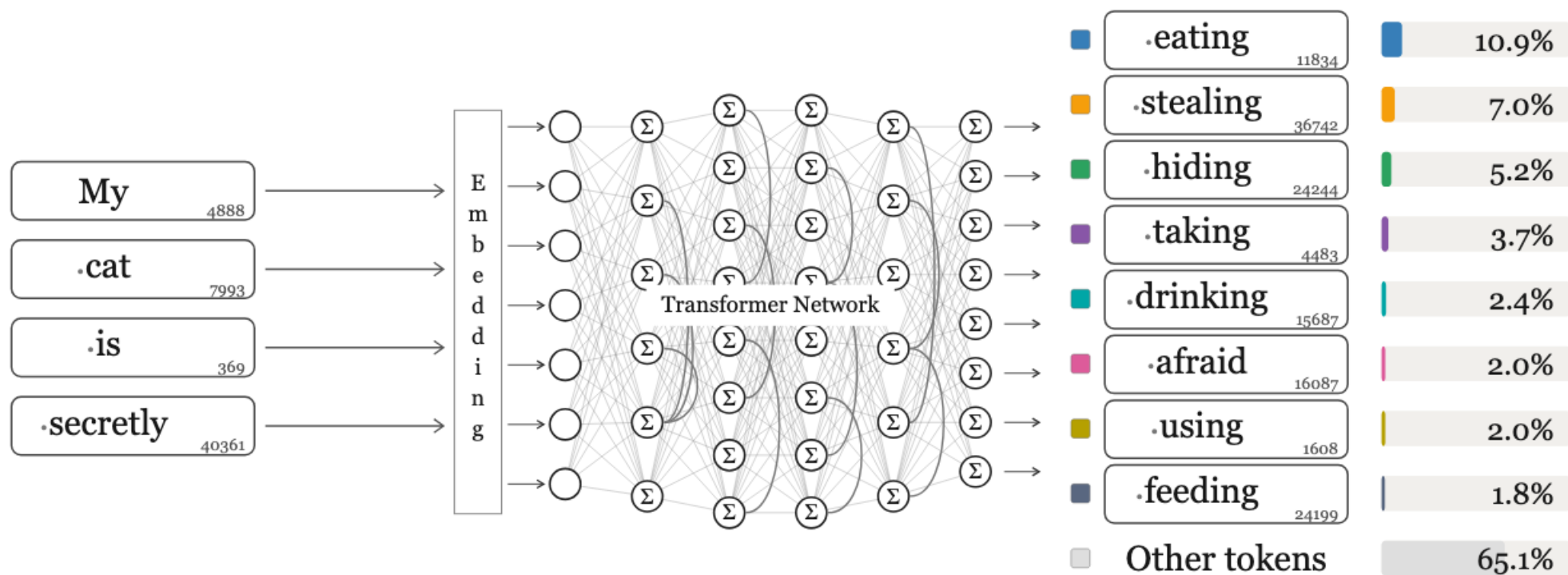
7 tokens

Tokenization varies by model · Qwen3.5-9B Base

A token is a piece of text:
a whole word, part of a word, or punctuation

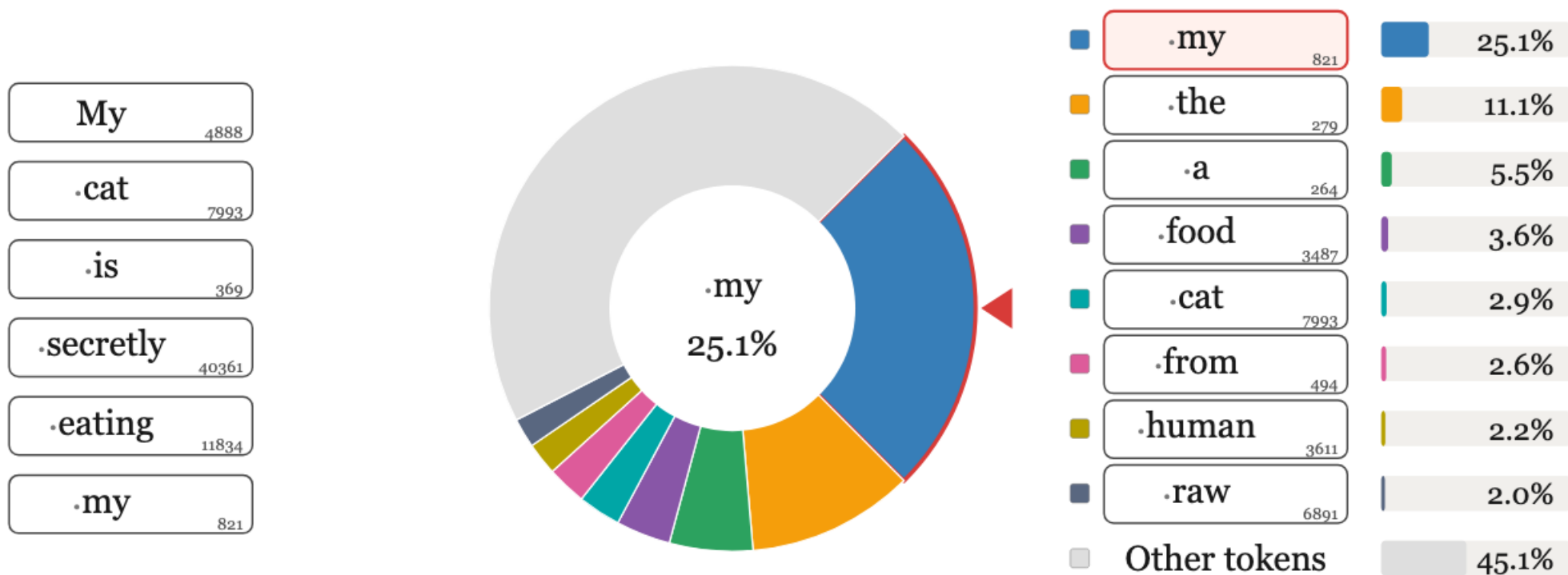
Token ID → Embedding → List of numbers

Same Input, Different Outputs



The network predicts the chances

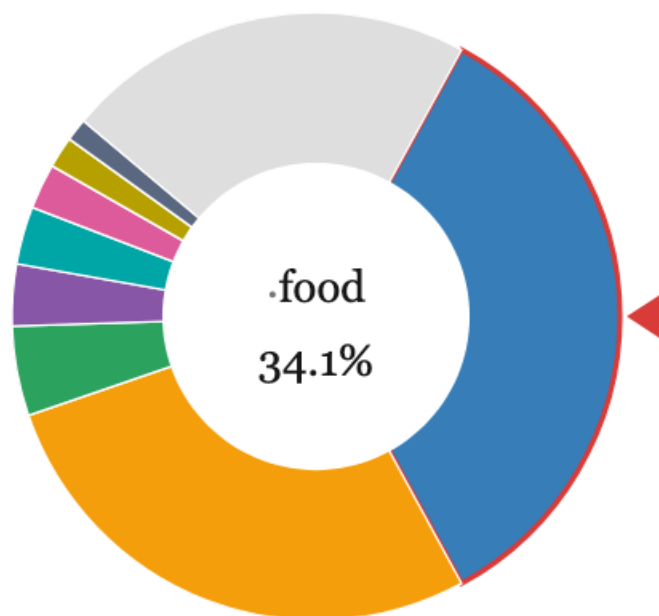
Same Input, Different Outputs



Larger section = greater chance

Same Input, Different Outputs

My	4888
.cat	7993
.is	369
.secretly	40361
.stealing	36742
.food	3487



.food	3487	34.1%
.my	821	27.7%
.and	321	4.8%
.the	279	3.3%
.treats	30697	3.1%
.small	2526	2.4%
.pieces	9378	1.7%
.fish	7397	1.2%
Other tokens		21.8%

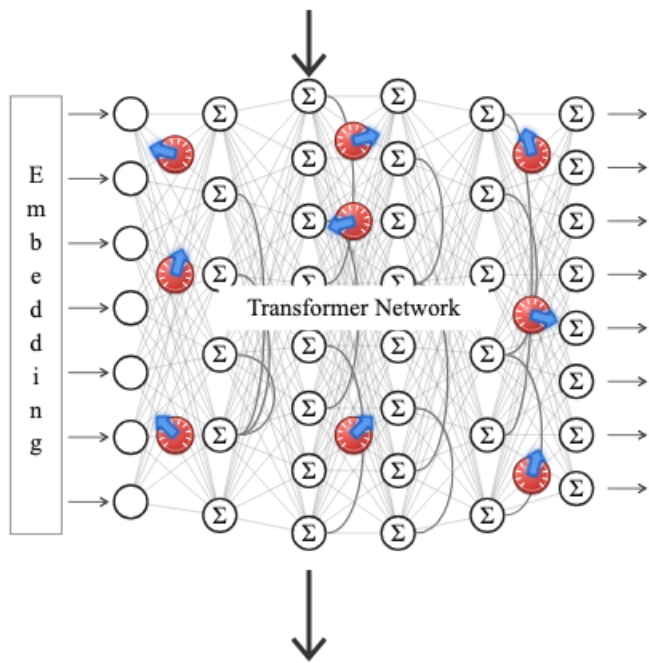
Qwen3.5-9B Base

DEMO 3 · Choosing the Next Token

How Is the Model Trained?

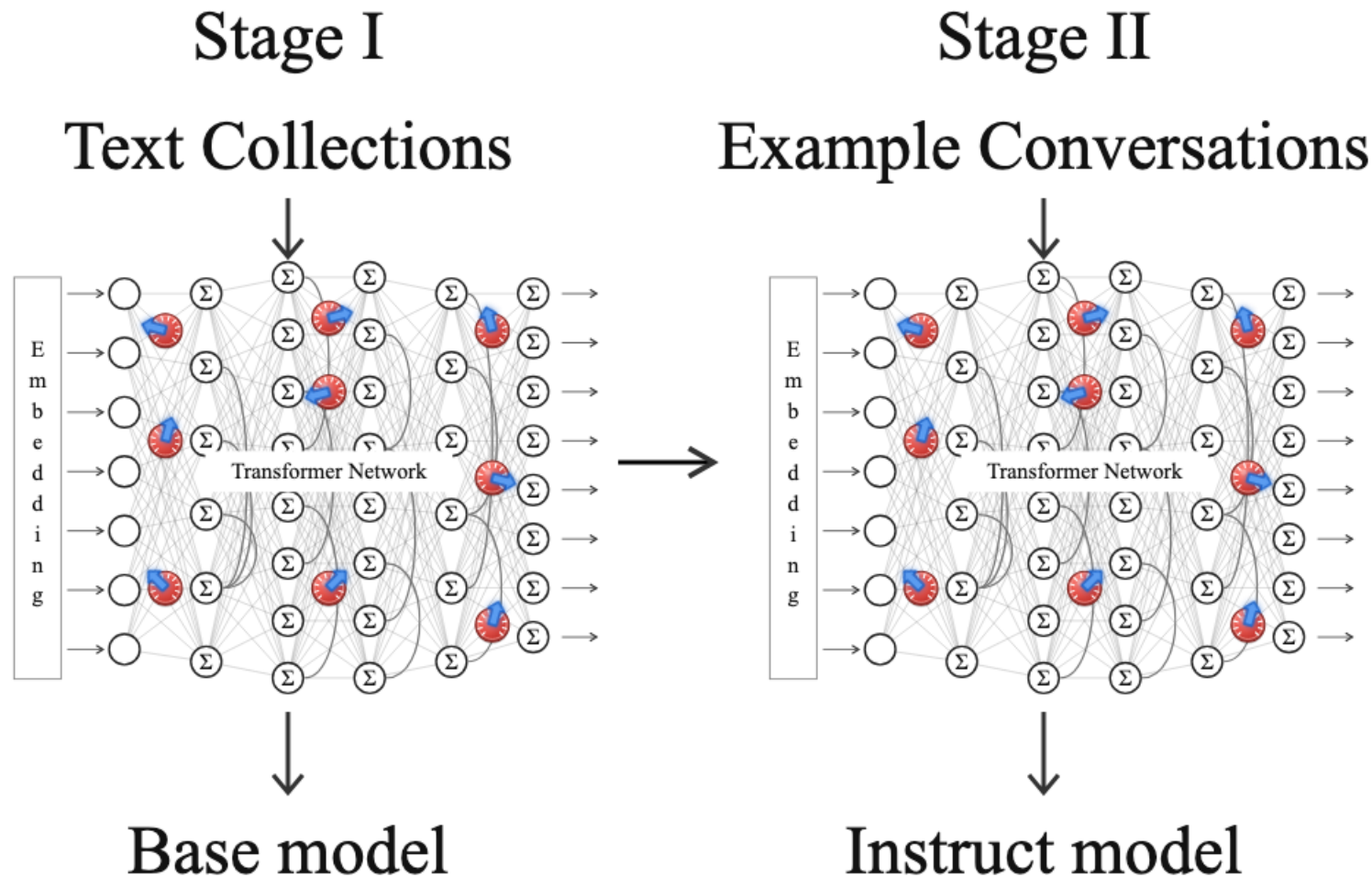
Stage I

Text Collections

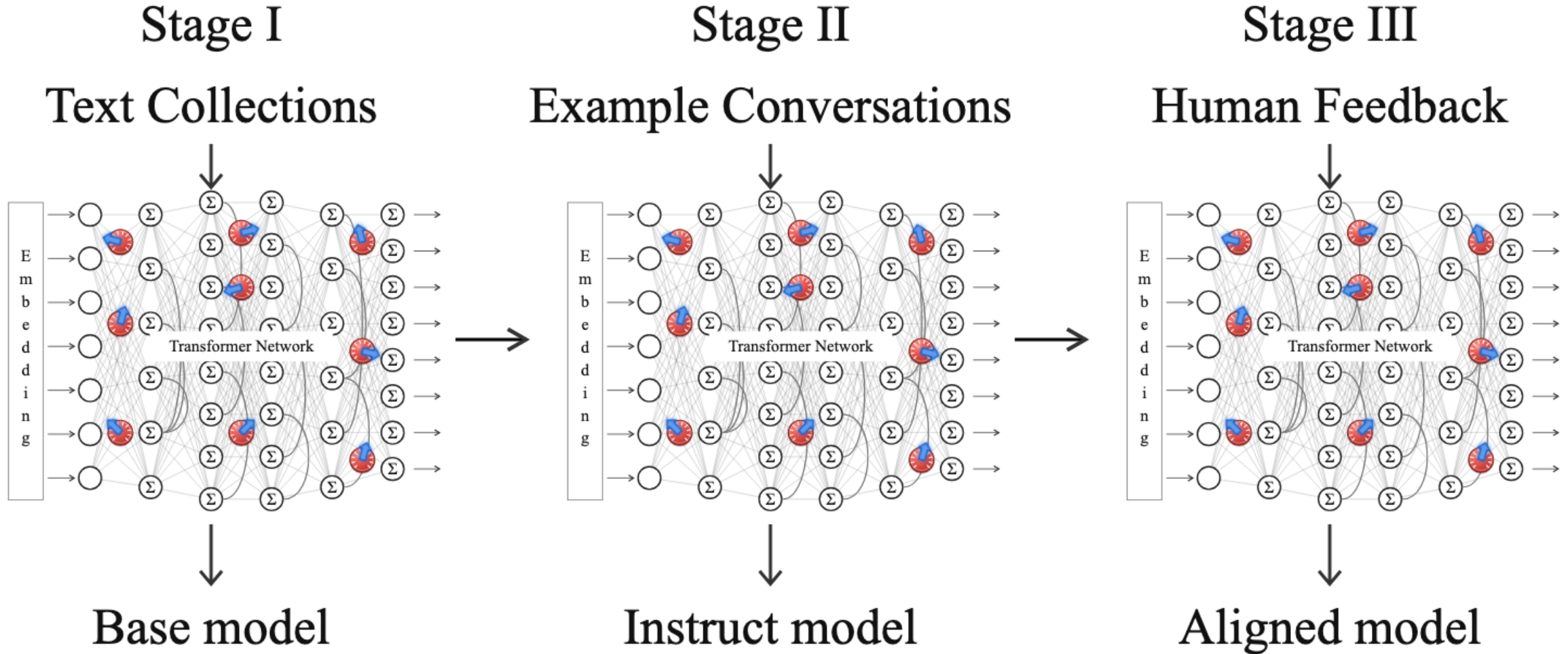


Base model

How Is the Model Trained?



How Is the Model Trained?



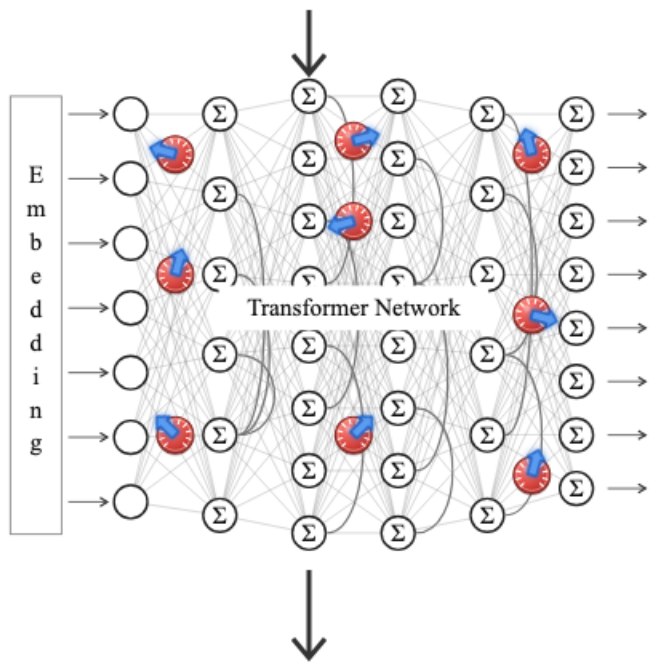
How Is the Model Trained?

Pre-training

Fine-tuning

Stage I

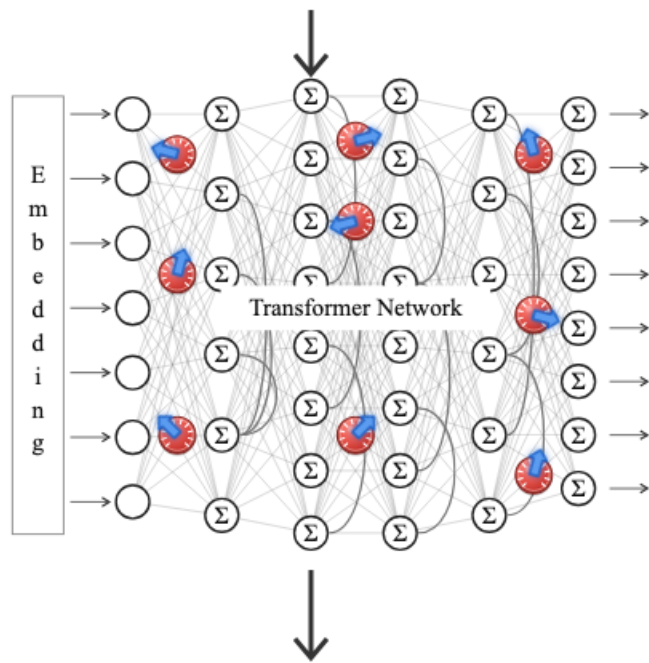
Text Collections



Base model

Stage II

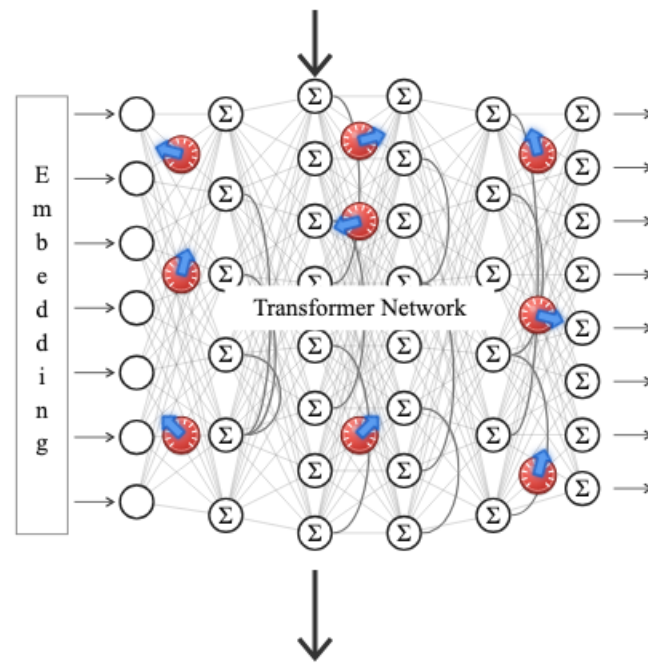
Example Conversations



Instruct model

Stage III

Human Feedback



Aligned model

Stage I



One crawl · August 2026
2.14 billion web pages
33.1 million registered domains
360 TiB uncompressed

The logo for FineWeb, where the word "FineWeb" is written in a bold, black, sans-serif font. The letter "i" is replaced by a bunch of purple grapes with a green leaf, and the letter "e" is replaced by a wine glass filled with red wine.

The finest collection of data the web has to offer



How Was FineWeb Created?

An example of pre-training data

Stage I



One crawl · August 2026

2.14 billion web pages

33.1 million registered domains

360 TiB uncompressed



Remove websites
on the blocklist

The logo for FineWeb, where the word "FineWeb" is written in a large, bold, black sans-serif font. The letter "i" in "Fine" is replaced by a small illustration of a purple bunch of grapes, and the letter "u" in "Web" is replaced by a small illustration of a wine glass filled with red liquid.

The finest collection of data the web has to offer



How Was FineWeb Created?

An example of pre-training data

Stage I



One crawl · August 2026
2.14 billion web pages
33.1 million registered domains
360 TiB uncompressed



Remove websites
on the blocklist



Extract readable text
from HTML pages

FineWeb

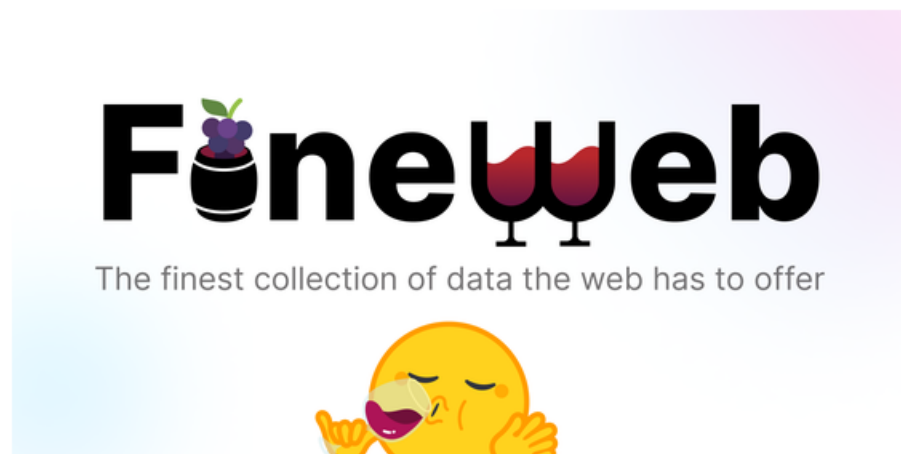
The finest collection of data the web has to offer



How Was FineWeb Created?

An example of pre-training data

Stage I



How Was FineWeb Created?

An example of pre-training data



URL
Filtering

Text
Extraction

Language
Filtering

One crawl · August 2026
2.14 billion web pages
33.1 million registered domains
360 TiB uncompressed



Remove websites
on the blocklist



Extract readable text
from HTML pages



Keep English text
(language score ≥ 0.65)

Stage I



URL
Filtering

Text
Extraction

Language
Filtering

MinHash
dedup

One crawl · August 2026

2.14 billion web pages

33.1 million registered domains

360 TiB uncompressed



Remove websites
on the blocklist



Extract readable text
from HTML pages



Keep English text
(language score ≥ 0.65)



Filter low-quality text
and remove duplicates

FineWeb

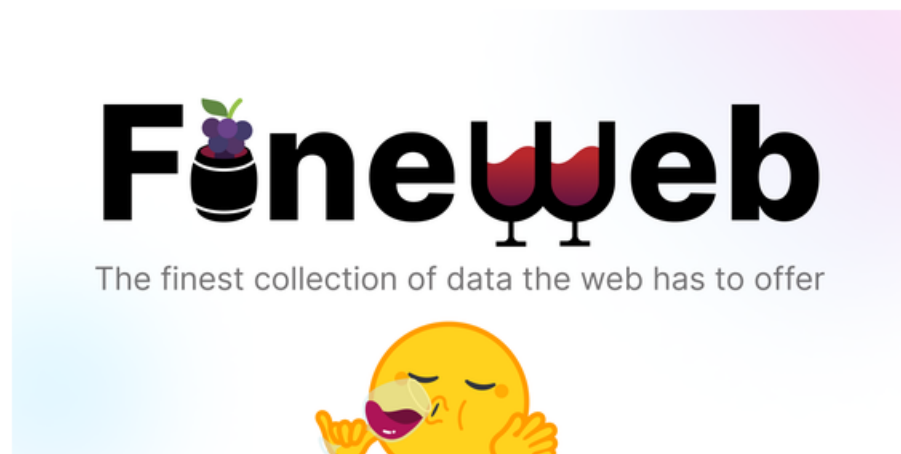
The finest collection of data the web has to offer



How Was FineWeb Created?

An example of pre-training data

Stage I



How Was FineWeb Created?

An example of pre-training data



URL
Filtering

Text
Extraction

Language
Filtering

MinHash
dedup

PII
Removal

One crawl · August 2026

2.14 billion web pages

33.1 million registered domains

360 TiB uncompressed



Remove websites
on the blocklist



Extract readable text
from HTML pages



Keep English text
(language score ≥ 0.65)



Filter low-quality text
and remove duplicates



Mask personal information
(email and IP addresses)

Stage I



URL
Filtering

Text
Extraction

Language
Filtering

MinHash
dedup

PII
Removal

One crawl · August 2026

2.14 billion web pages

33.1 million registered domains

360 TiB uncompressed



Remove websites
on the blocklist



Extract readable text
from HTML pages



Keep English text
(language score ≥ 0.65)



Filter low-quality text
and remove duplicates



Mask personal information
(email and IP addresses)

FineWeb

The finest collection of data the web has to offer



How Was FineWeb Created?

An example of pre-training data

FineWeb2 extends this approach
to 1,000+ languages

Stage I

My cat is secretly **stealing** my socks

Stage I

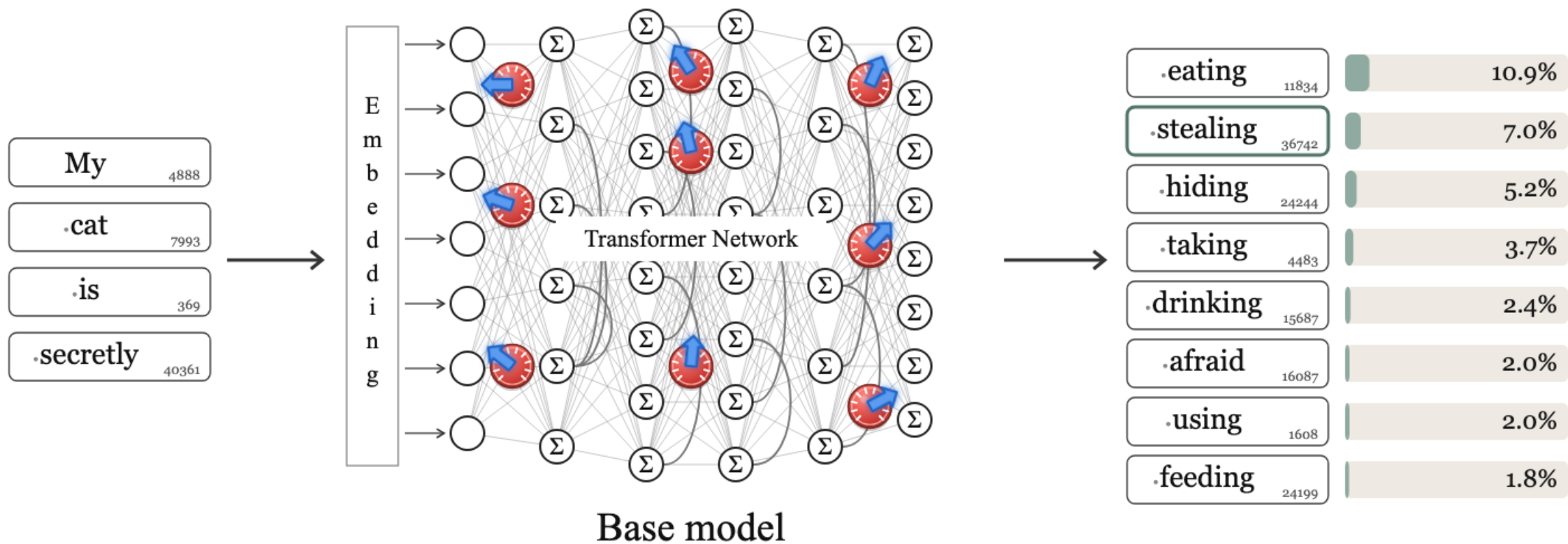
My cat is secretly **stealing** my socks

Target token → .stealing₃₆₇₄₂

Stage I

My cat is secretly **stealing** my socks

Target token → **stealing** 36742

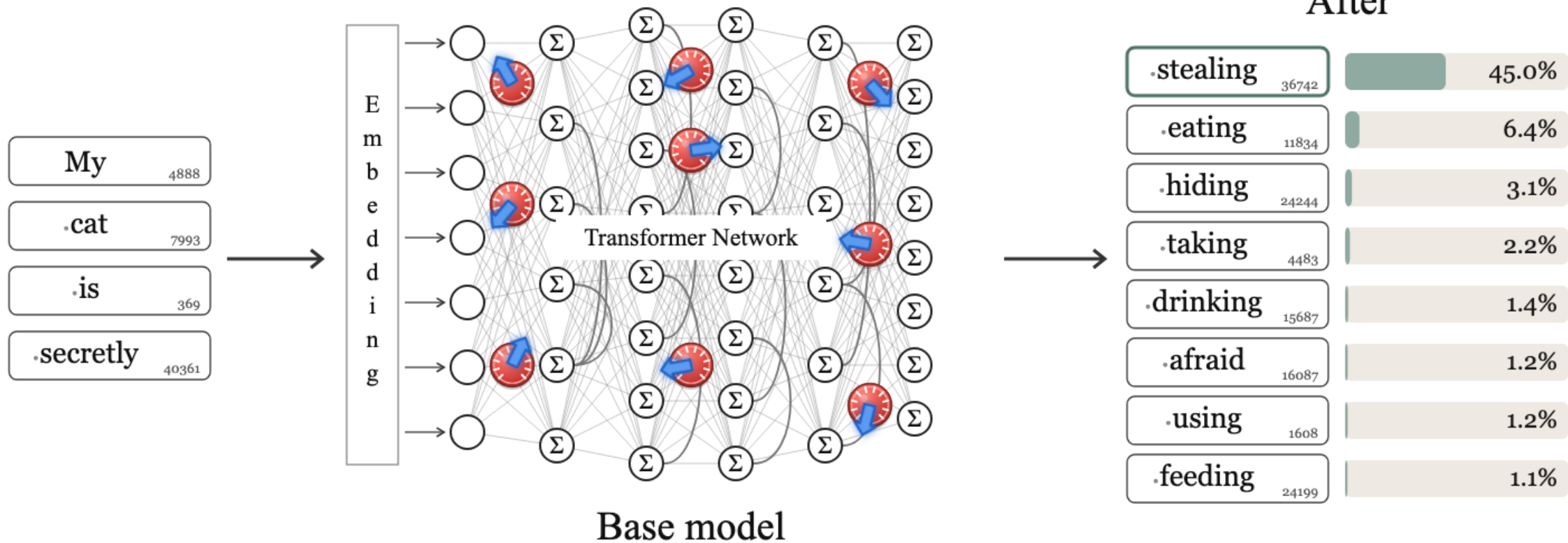


Stage I

My cat is secretly **stealing** my socks

Target token → **stealing** 36742

After



Qwen3.5-9B Base

DEMO 4 · Continuous Generation

Stage II

Same neural network · Different training data

User

Explain monopsony in simple terms

Stage II

Same neural network · Different training data

User

Explain monopsony in simple terms

Assistant

Monopsony means there is one main buyer.
For example, a town's only large employer
may have power to keep wages low

Stage II

Same neural network · Different training data



Human-written
or generated
and reviewed

User

Explain monopsony in simple terms

Assistant

Monopsony means there is one main buyer.
For example, a town's only large employer
may have power to keep wages low

Collect carefully checked request–response examples

Stage II

Monopsony means there is **one** main buyer

Stage II

Monopsony means there is **one** main buyer

Target token → .one₇₉₉

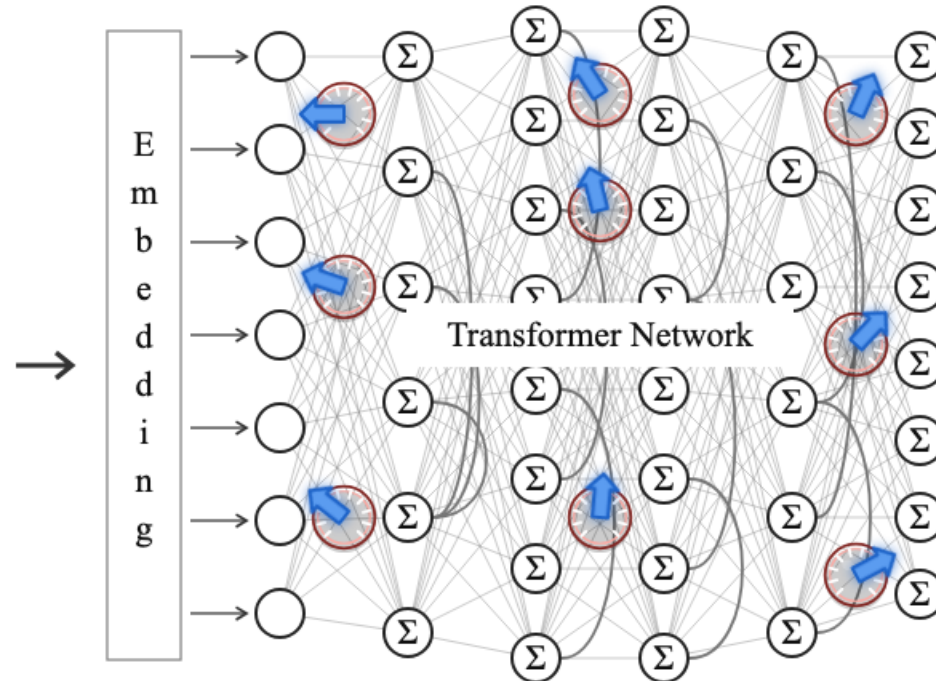
Stage II

Monopsony means there is **one** main buyer

Target token → .one₇₉₉



Reviewed example
Explain monopsony
in simple terms
Monopsony means
there is



Instruct model

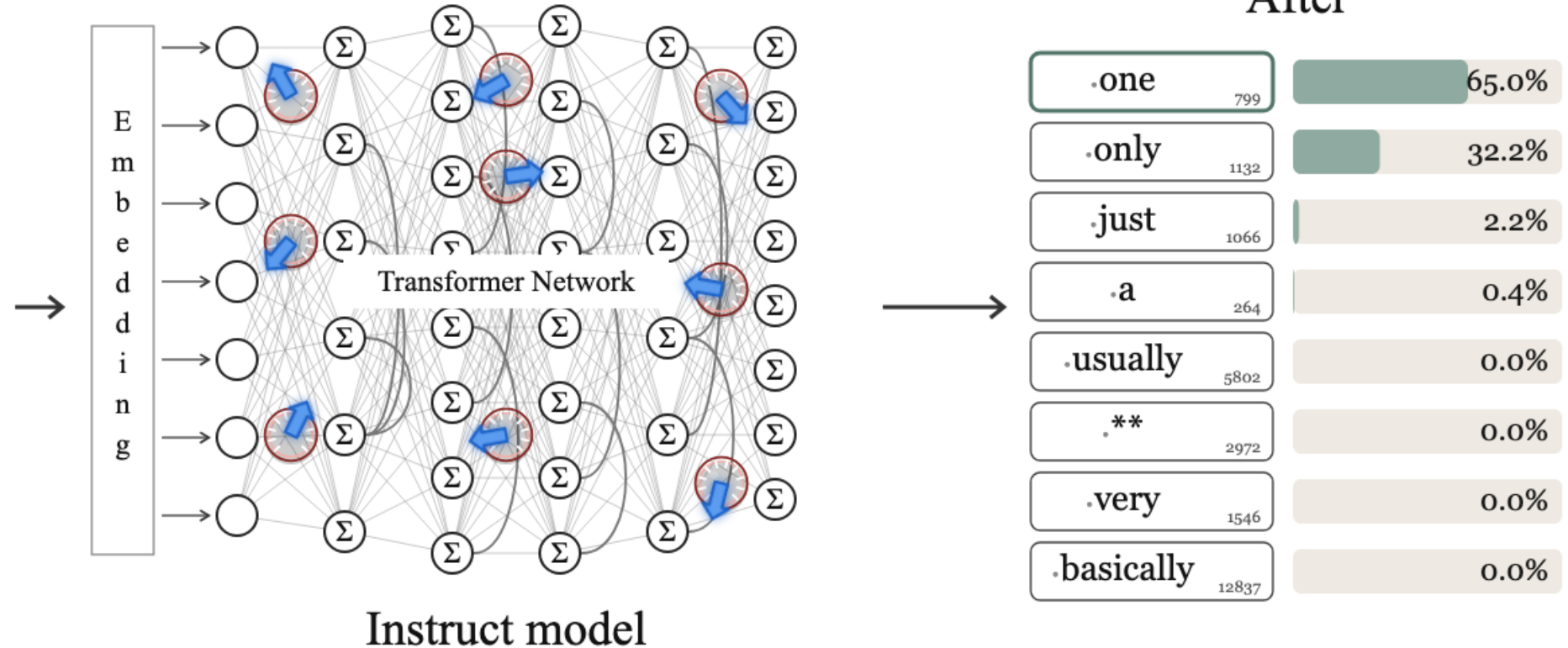
.only ₁₁₃₂	88.2%
.just ₁₀₆₆	6.0%
.one ₇₉₉	4.2%
.a ₂₆₄	1.1%
.usually ₅₈₀₂	0.1%
.** ₂₉₇₂	0.0%
.very ₁₅₄₆	0.0%
.basically ₁₂₈₃₇	0.0%

Stage II

Monopsony means there is **one** main buyer

Target token → .one
799

After



Reviewed example
Explain monopsony
in simple terms
Monopsony means
there is

Stage II

Qwen3.5-9B



DEMO 5 · Base vs. Instruct

Stage III

Explain monopsony to a first-year student

A

Monopsony is a market structure characterised by a single buyer with market power over suppliers

B

When one employer is the main buyer of labour in a town, workers have fewer alternatives and less power to negotiate pay

Which is clearer for this reader, and why?

Stage III



People compare answers
Across many examples

Stage III



People compare answers
Across many examples

Train



Reward model
Predicts preference scores

Stage III



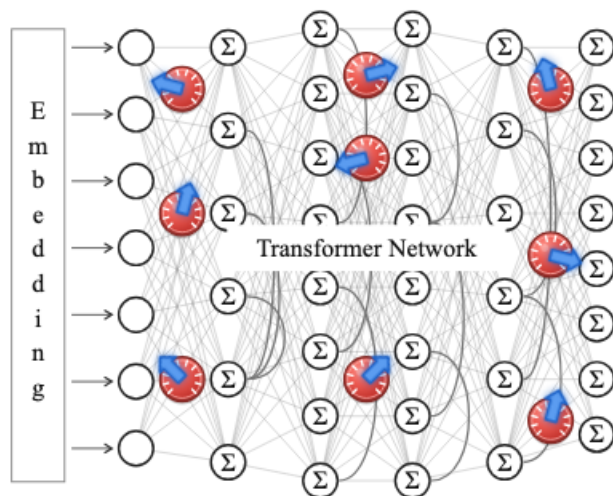
People compare answers
Across many examples

Train

Reward model
Predicts preference scores

Predicts next tokens

Text



LLM

Prompt + answer

Formal definition

Concrete example

Stage III



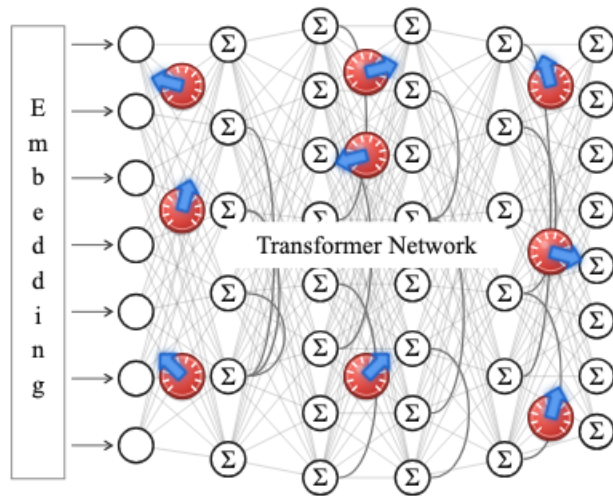
People compare answers
Across many examples

Train

Reward model
Predicts preference scores

Predicts next tokens

Text



LLM

Prompt + answer

Formal definition

Concrete example

Score

Scores guide LLM weight updates

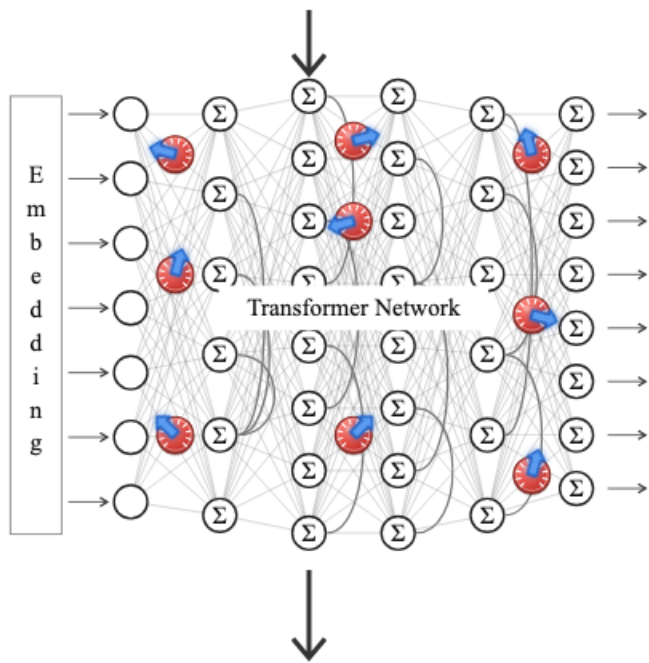
Training Complete

Pre-training

Fine-tuning

Stage I

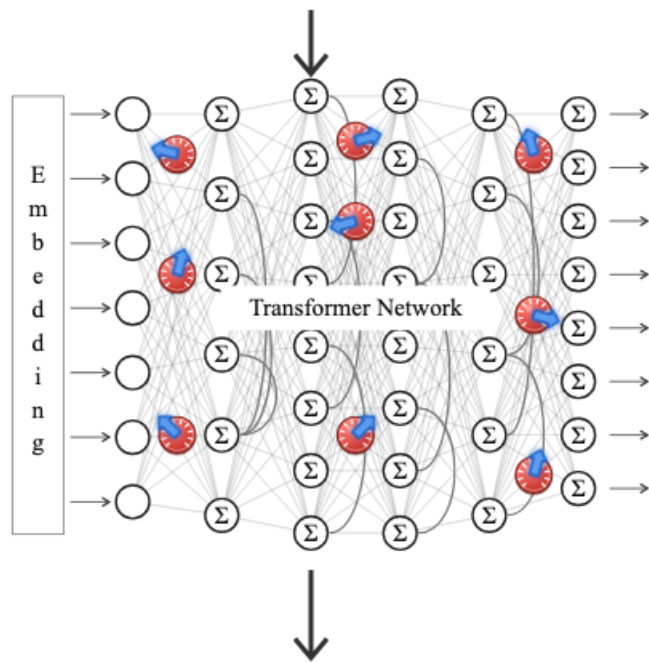
Text Collections



Base model

Stage II

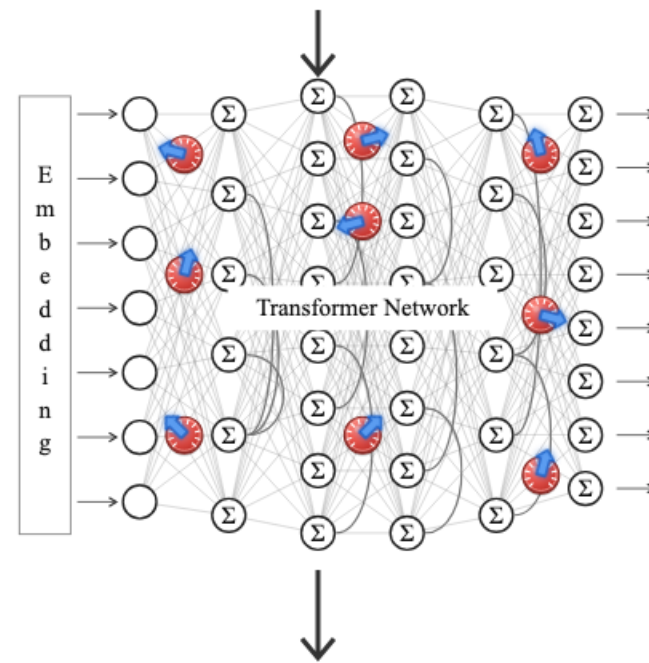
Example Conversations



Instruct model

Stage III

Human Feedback

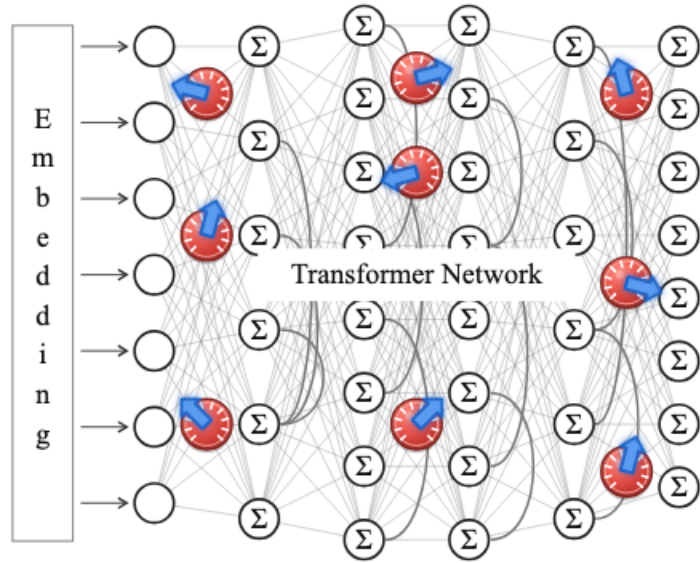


Aligned model

Training and Inference

Training

Learn from examples and feedback

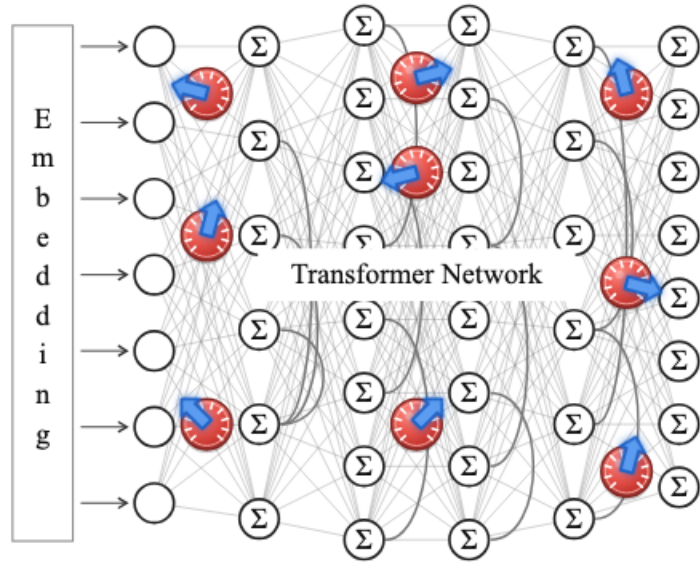


Weights change

Training and Inference

Training

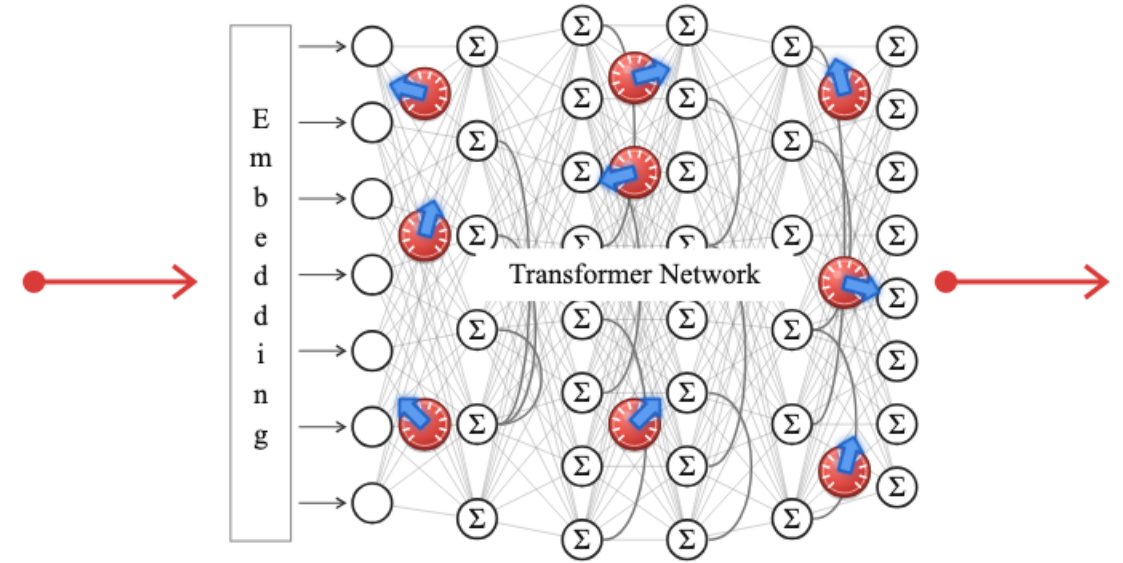
Learn from examples and feedback



Weights change

Inference

Use context to generate the next token

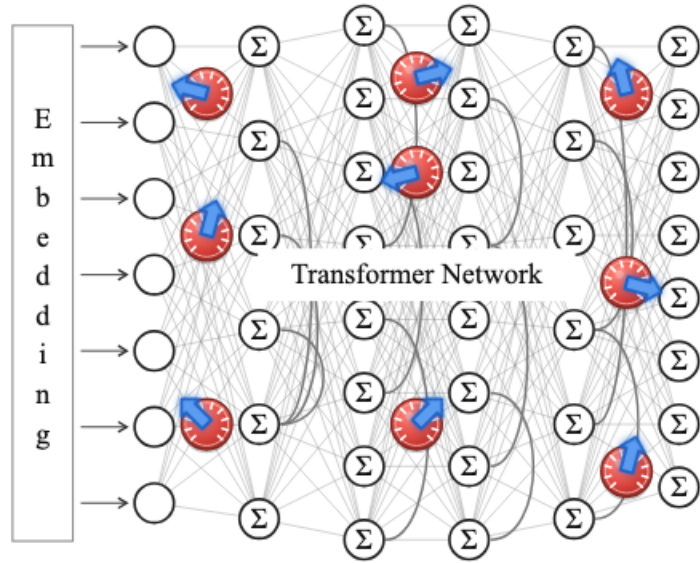


Weights stay fixed

Training and Inference

Training

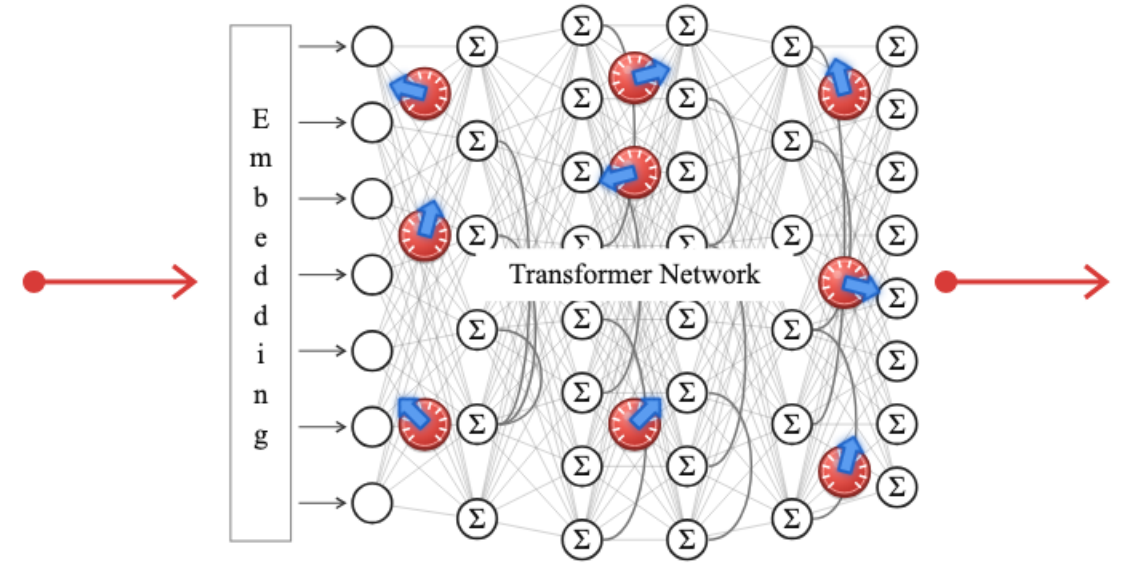
Learn from examples and feedback



Weights change

Inference

Use context to generate the next token



Weights stay fixed

A longer conversation changes the context

Attention Uses Context

Attention helps predict the next token by using relevant earlier tokens

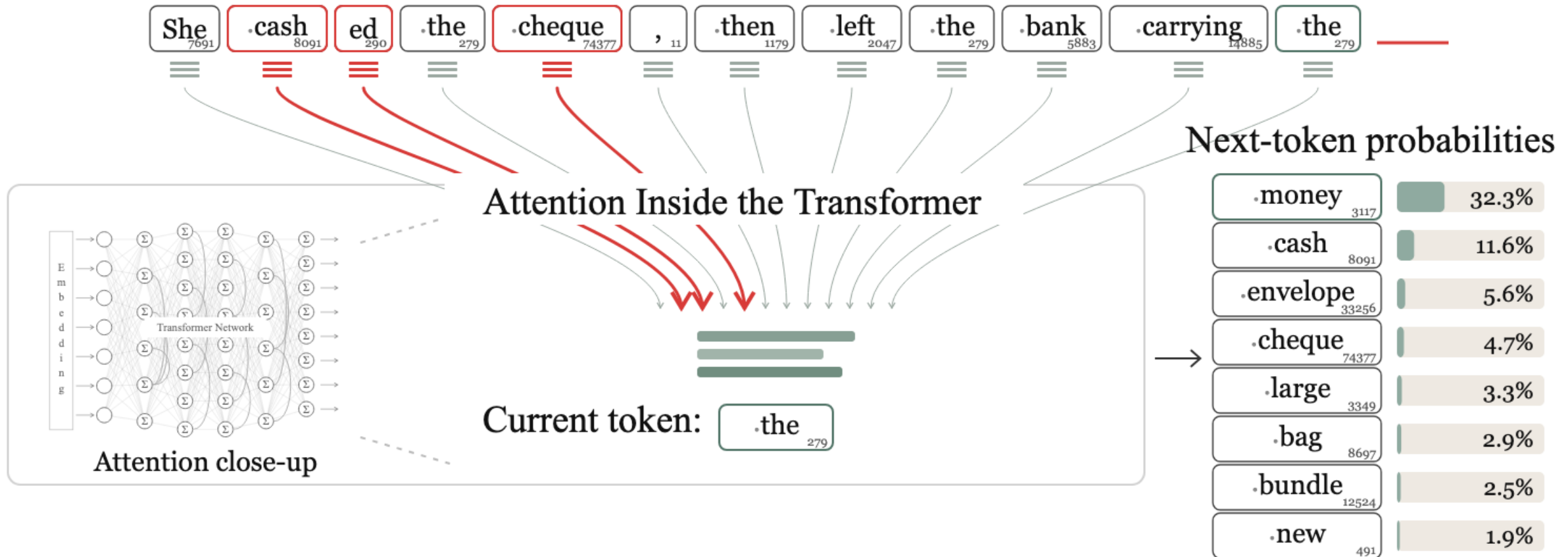
She₇₆₉₁ ·cash₈₀₉₁ ed₂₉₀ ·the₂₇₉ ·cheque₇₄₃₇₇ ,₁₁ ·then₁₁₇₉ ·left₂₀₄₇ ·the₂₇₉ ·bank₅₈₈₃ ·carrying₁₄₈₈₅ ·the₂₇₉ —

What comes **next**?

Which earlier tokens matter?

Attention Uses Context

Attention helps predict the next token by using relevant earlier tokens



Attention Uses Context

Attention helps predict the next token by using relevant earlier tokens

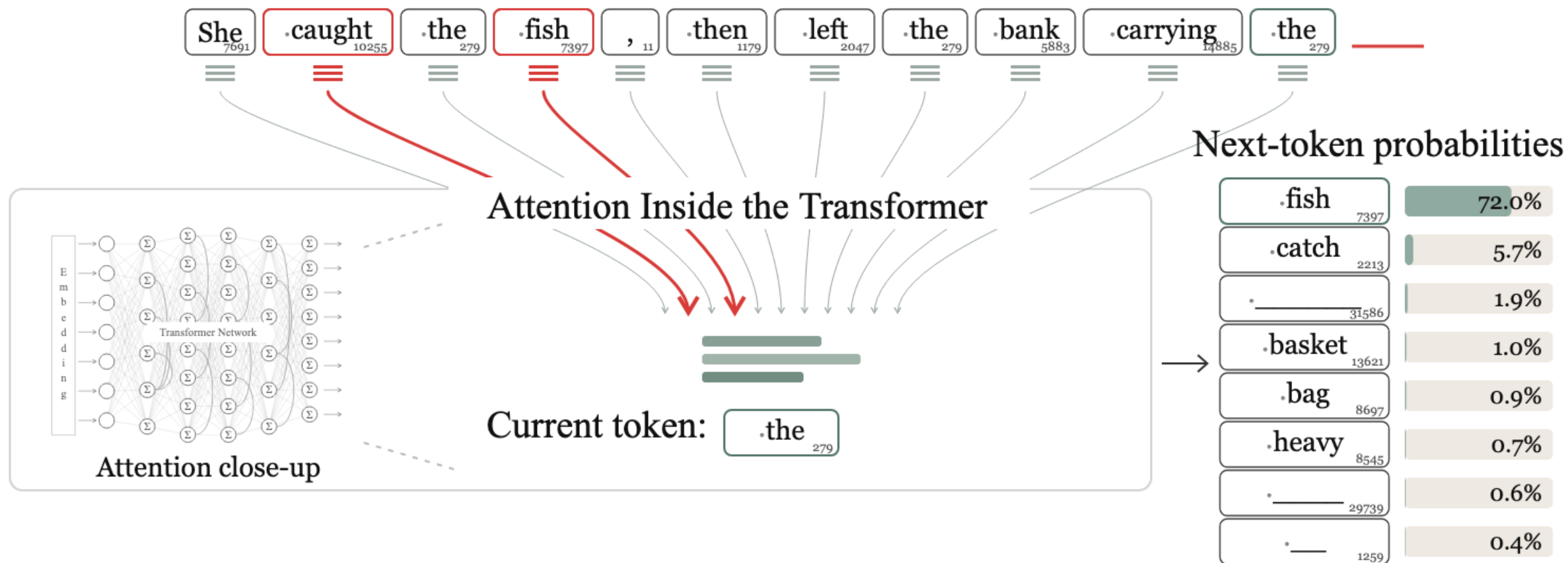


What comes **next**?

Which earlier tokens matter?

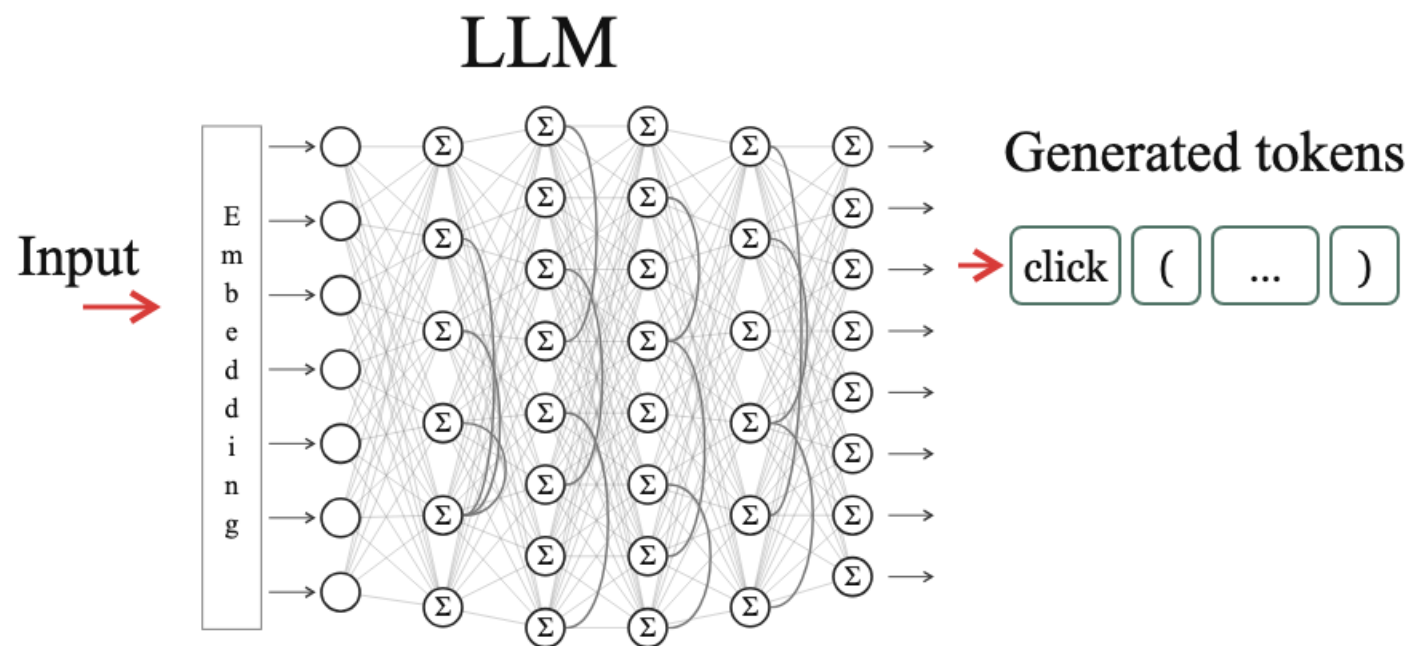
Attention Uses Context

Attention helps predict the next token by using relevant earlier tokens



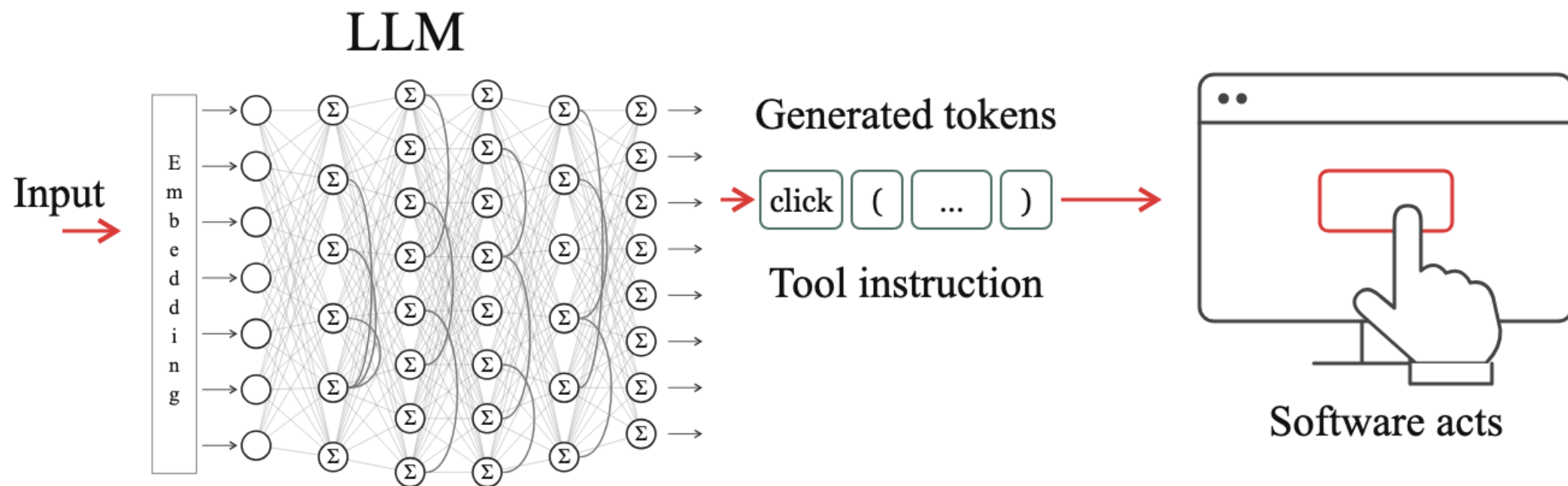
From LLM to Agent

An LLM generates tokens. An agent uses them to act.



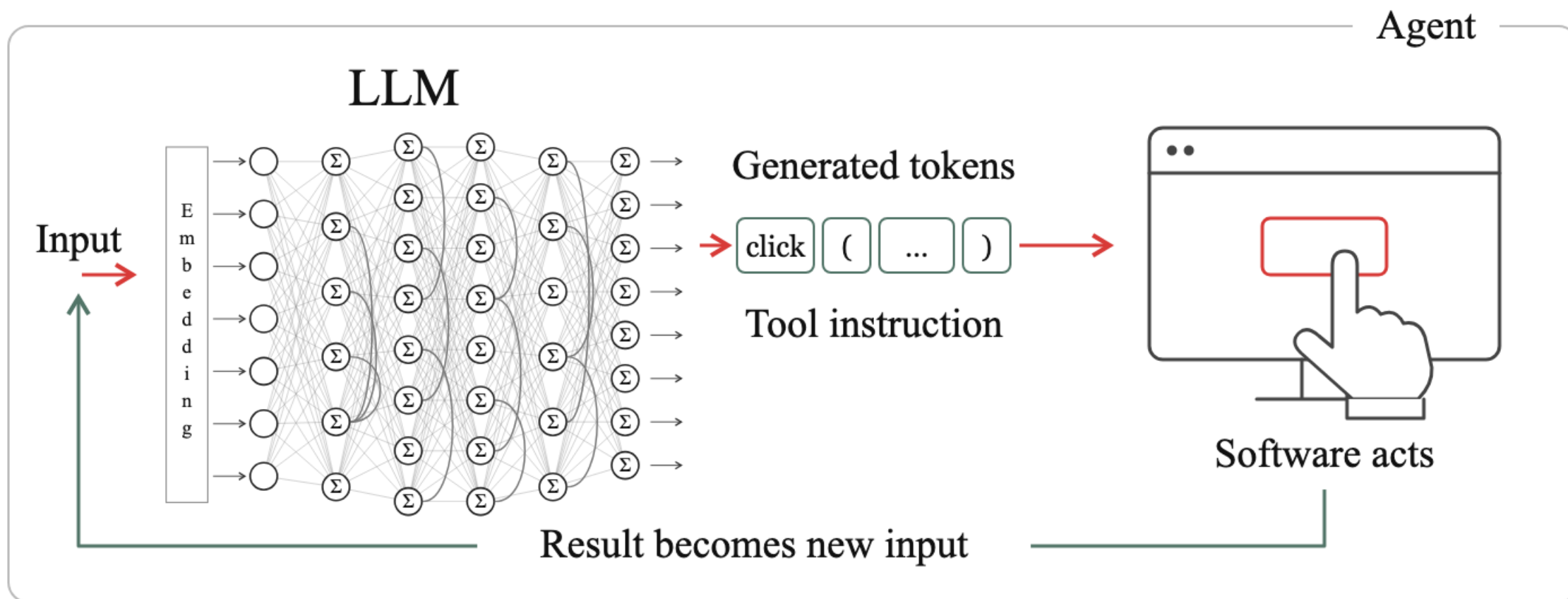
From LLM to Agent

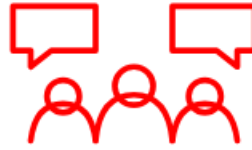
An LLM generates tokens. An agent uses them to act.



From LLM to Agent

An LLM generates tokens. An agent uses them to act.





1 Ed Login
on your phone



2 Camera App



3 Scan QR



4 Submit

ACTIVITY · EXIT TICKET