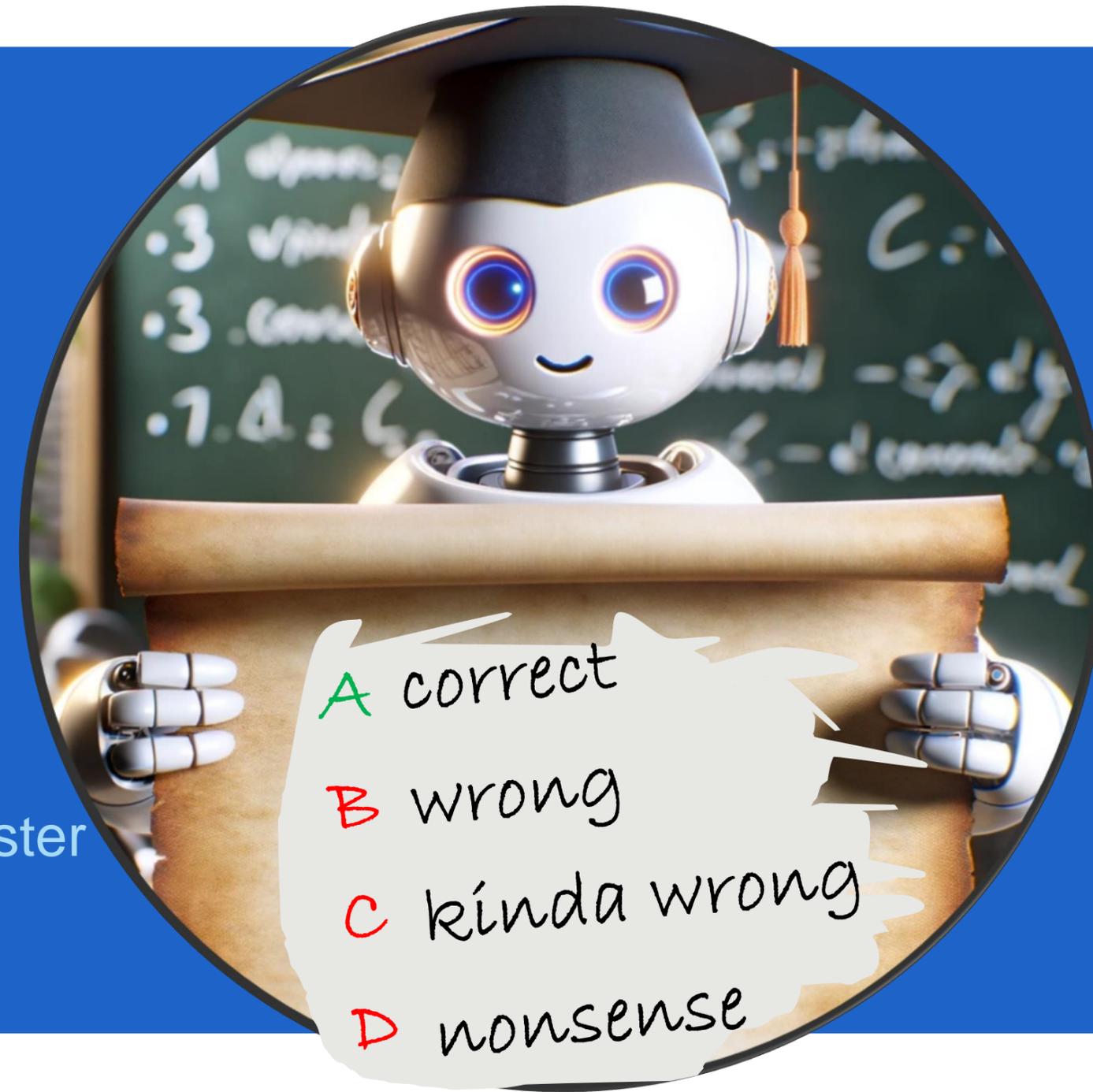


Language Model Adaptation with Applications in AI for Education

Ph.D. candidate: Semere Kiros Bitew
Promotors: prof. Chris Develder, prof. Thomas Demeester
March 08th, 2024



Outline

- AI in Education
- Research Motivation
- Language Modeling
- Research Contributions
 - Distractor generation task
 - Gap-filling grammar exercise generation
- Conclusion

Let's do a quick poll: Personal Trivia

Let's do a quick poll: Personal Trivia

What is my last name?

- A Kiros
- B Bitew
- C Kiros Bitew
- D Semere

Let's do a quick poll: Personal Trivia

What is my last name?

- A Kiros
- B Bitew
- C Kiros Bitew
- D Semere

What is my last name?

- A Smith
- B Bitew
- C Apple
- D BMW

Let's do a quick poll: Personal Trivia

What is my last name?

- A Kiros
- B Bitew**
- C Kiros Bitew
- D Semere

What is my last name?

- A Smith
- B Bitew**
- C Apple
- D BMW

Let's do a quick poll: Personal Trivia

Where am I originally from?

- A Mekelle
- B The moon
- C Pirate ship
- D Mars

Where am I originally from?

- A Mekelle
- B Axum
- C Addis Ababa
- D Nairobi

Let's do a quick poll: Personal Trivia

Where am I originally from?

- A Mekelle
- B The moon
- C Pirate ship
- D Mars

Where am I originally from?

- A Mekelle
- B Axum
- C Addis Ababa
- D Nairobi

Let's do a quick poll: Personal Trivia

What year was I born?

- A 1776
- B 1892
- C 2050
- D 1992

What year was I born?

- A 1984
- B 1989
- C 1995
- D 1992

Let's do a quick poll: Personal Trivia

What year was I born?

- A 1776
- B 1892
- C 2050
- D 1992**

What year was I born?

- A 1984
- B 1989
- C 1995
- D 1992**

Let's do a quick poll: Personal Trivia

My last name is **Bitew**, born in **1992**, and I am from **Mekelle!**

Introduction to AI in Education

AI in Education

Support for **teaching**, **learning**, and **school administration**.

AI in Education

AI in Education

AI in Education

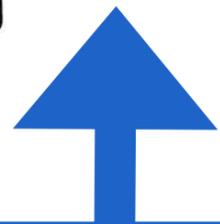


AI in Education

AI in Education

Examples

- FAQ Chatbot
- Essay Scoring

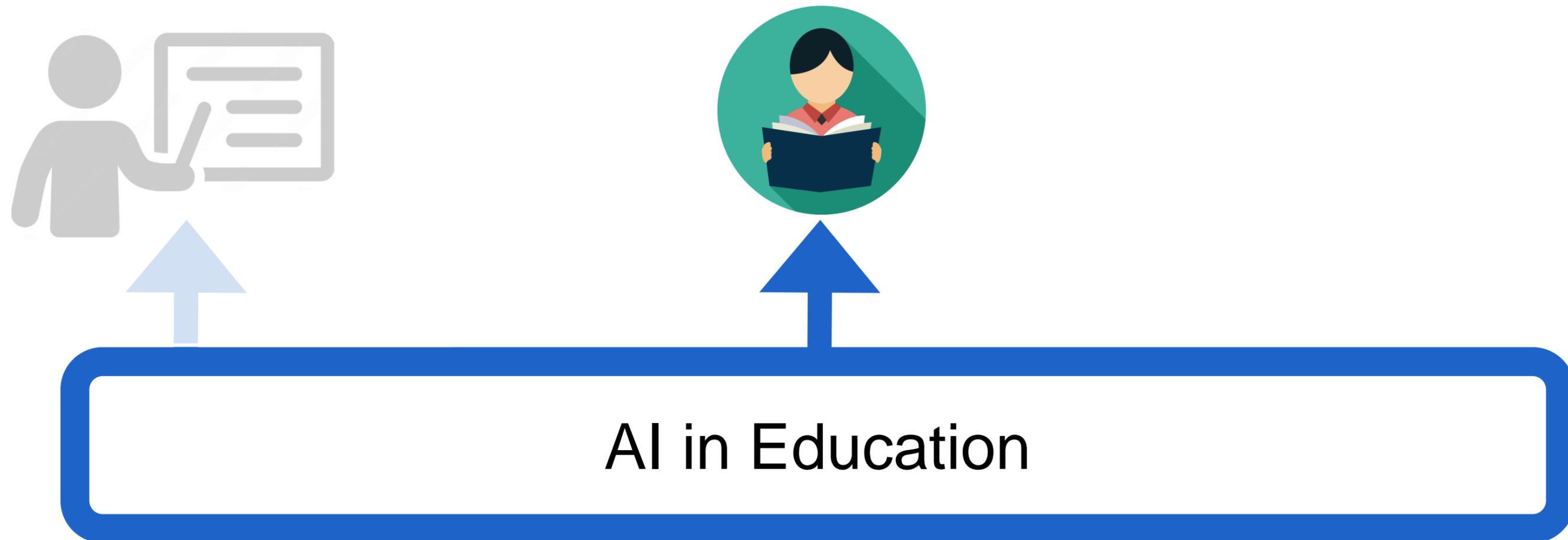


AI in Education

AI in Education

Examples

- FAQ Chatbot
- Essay Scoring



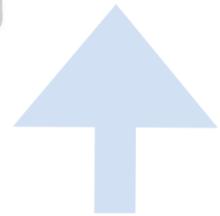
AI in Education

Examples

- FAQ Chatbot
- Essay Scoring



- **duolingo**
- Formative assessment



AI in Education

AI in Education

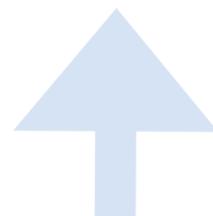
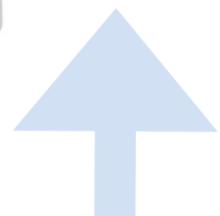
Examples

- FAQ Chatbot
- Essay Scoring



duolingo

- Formative assessment



AI in Education

AI in Education

Examples

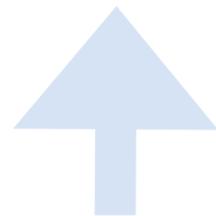
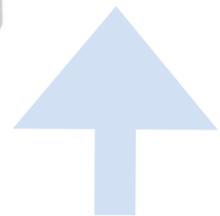
- FAQ Chatbot
- Essay Scoring



duolingo

- Formative assessment

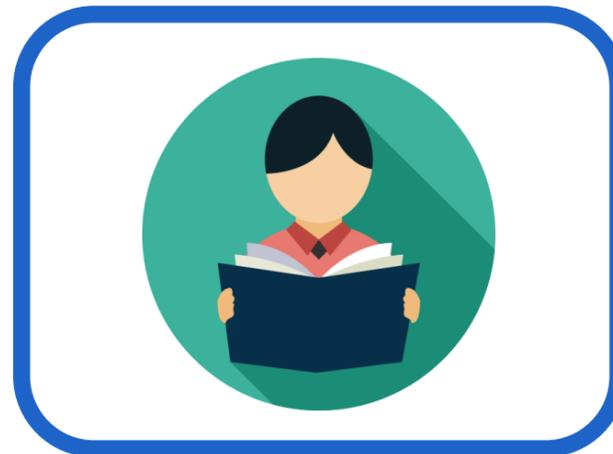
- Resource allocation



AI in Education

AI in Education

Focus of my research



AI in Education

Research Motivation

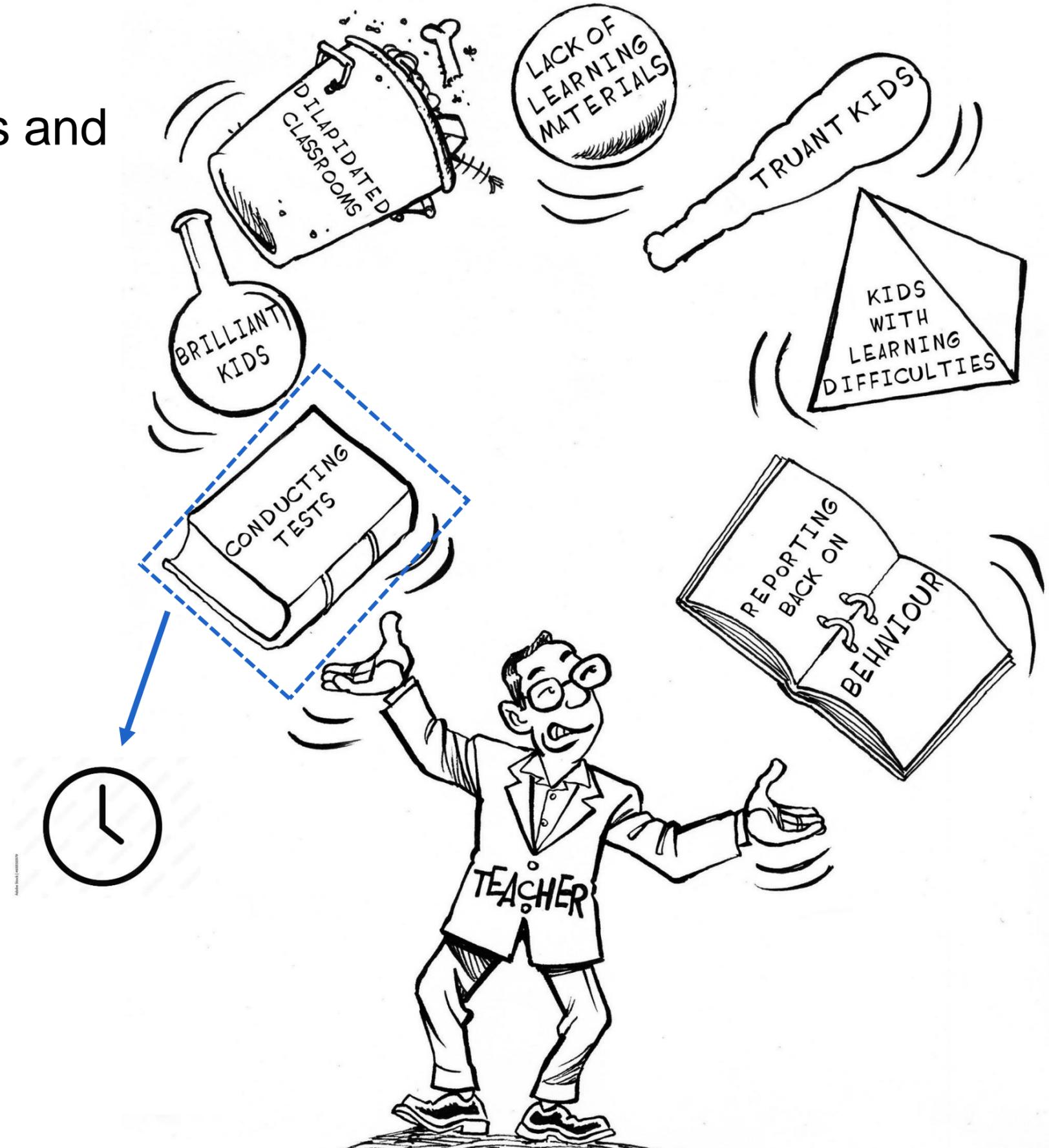
Burden on educators

- Heavy **workload** from creating exercises and tests



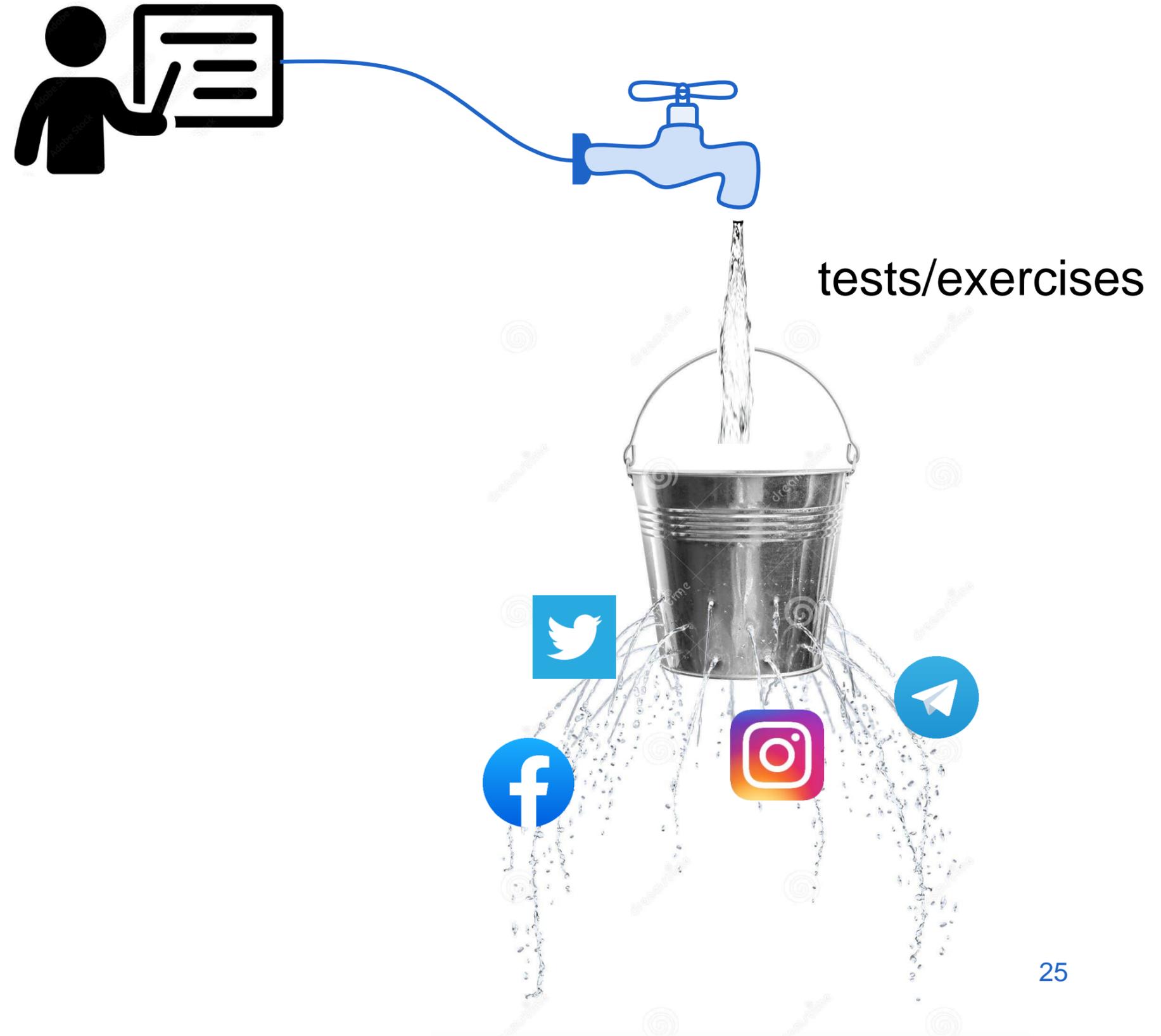
Burden on educators

- Heavy **workload** from creating exercises and tests
- Formulating **question** requires
 - Skill
 - Experience
 - Time



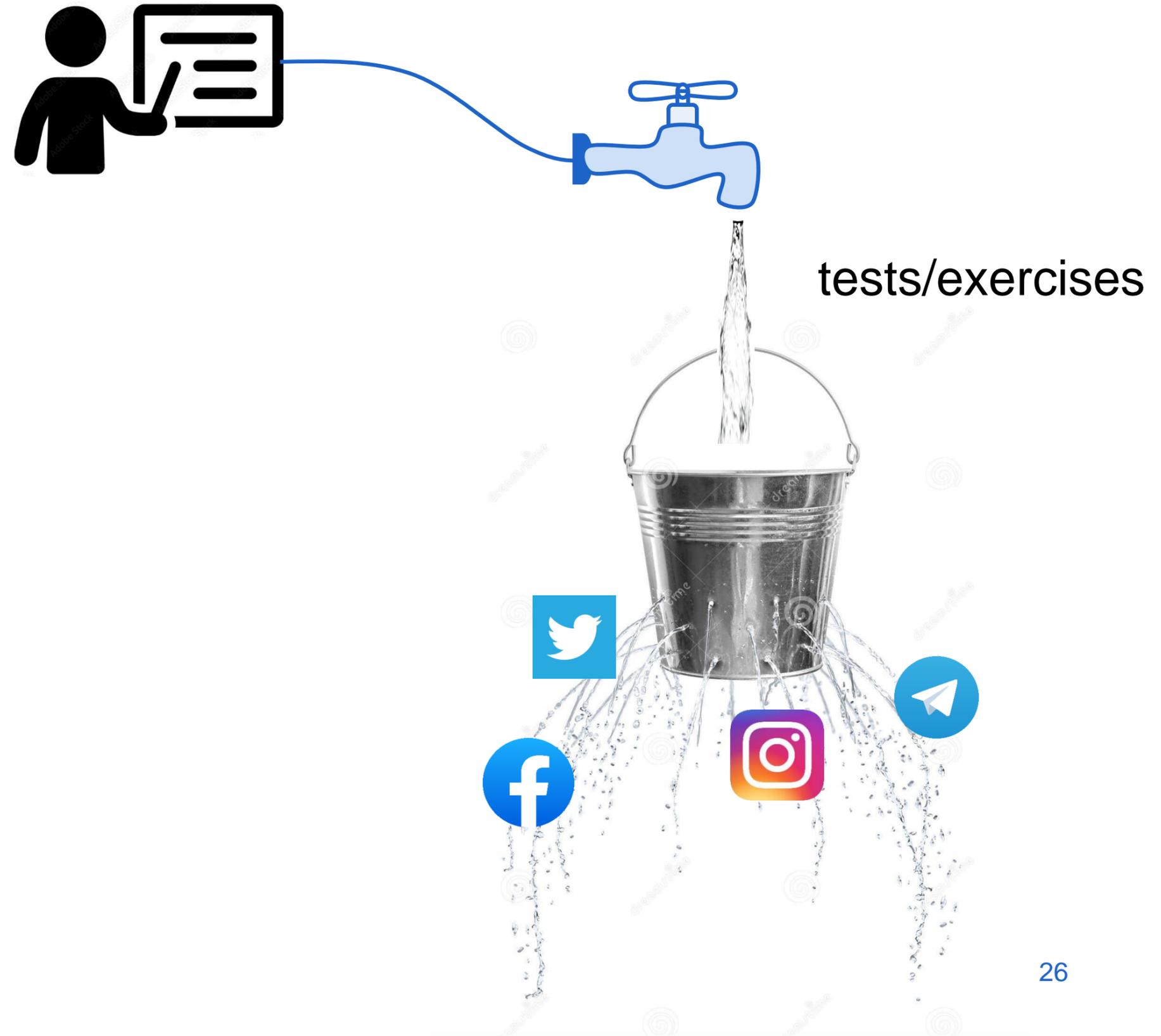
Single-use dilemma

- Question sharing and leakage



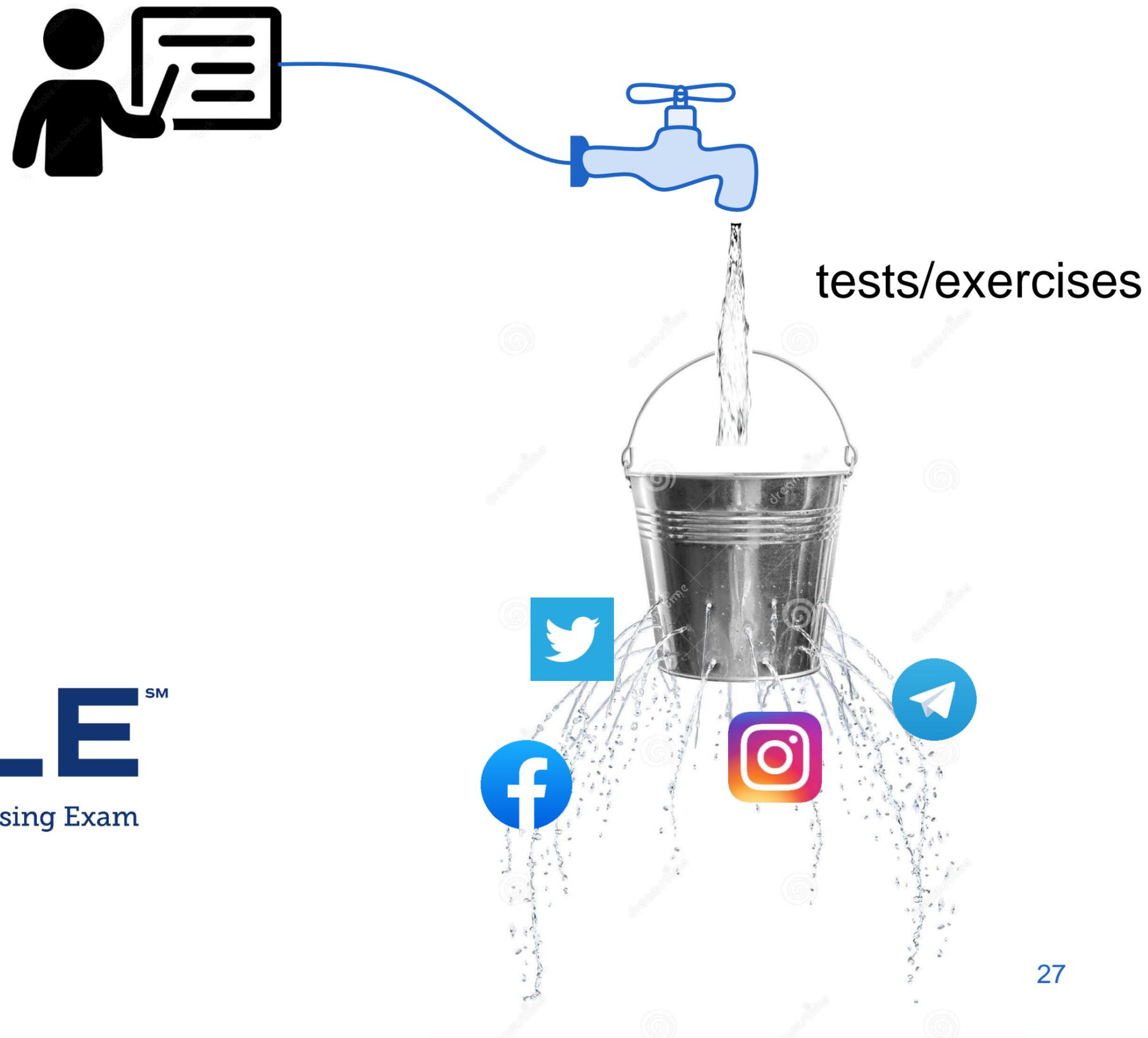
Single-use dilemma

- Question sharing and leakage
- **Problem** in high-stakes:
 - Certification, Job tests



Single-use dilemma

- Question sharing and leakage
- **Problem** in high-stakes:
 - Certification, Job tests



 ETS TOEFL


USMLESM
United States Medical Licensing Exam

Shift to digital education

Rapid digitization of educational materials, especially after COVID19

Shift to digital education

Rapid digitization of educational materials, especially after COVID19



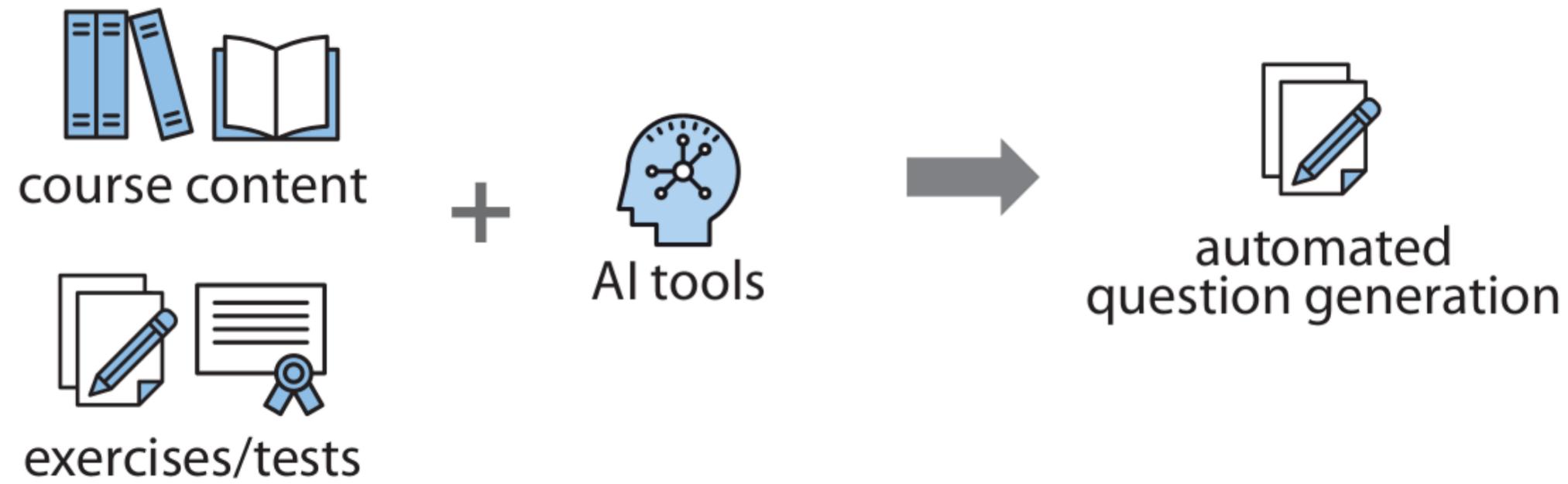
course content



exercises/tests

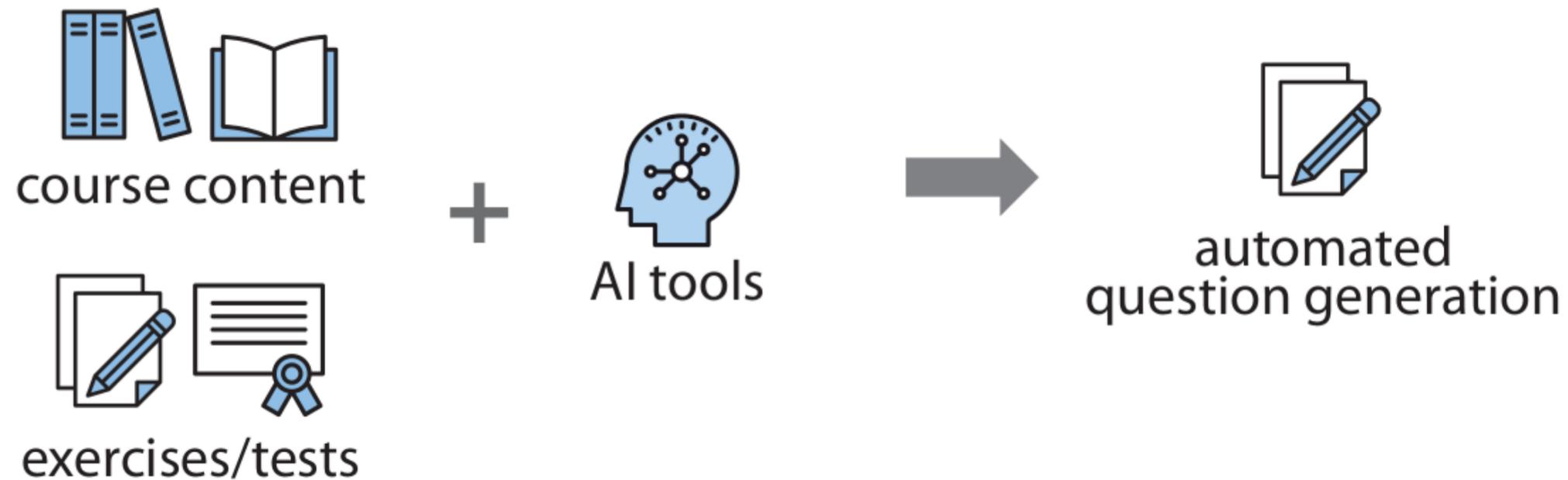
Shift to digital education

Rapid digitization of educational materials, especially after COVID19



Shift to digital education

Rapid digitization of educational materials, especially after COVID19



Goal: automatic educational question generation to support teachers & students

Concrete Example 1 (Question Answering)

Paragraph about Belgium

Belgium is a small country in Europe distinguished by its remarkable combination of historical significance and contemporary governance. Brussels is the capital city of Belgium. Other major cities include Antwerp, which is known for its diamond trade; Ghent, which has intact medieval buildings; and Liège, a city with a rich industrial past and culinary fame for its delicious waffles.

What is the Capital of Belgium?

AI tools

Automatically **Extracted**
Answer

Brussels

Concrete Example 2 (Question Generation)

Paragraph about Belgium

Belgium is a small country in Europe distinguished by its remarkable combination of historical significance and contemporary governance. Brussels is the capital city of Belgium. Other major cities include Antwerp, which is known for its diamond trade; Ghent, which has intact medieval buildings; and Liège, a city with a rich industrial past and culinary fame for its delicious waffles.

AI tools

Automatically Generated Multiple-Choice Question

What is the Capital of Belgium?

- A Brussels
- B Antwerp
- C Ghent
- D Liège

Input and output in human language

Paragraph about Belgium

Belgium is a small country in Europe distinguished by its remarkable combination of historical significance and contemporary governance. Brussels is the capital city of Belgium. Other major cities include Antwerp, which is known for its diamond trade; Ghent, which has intact medieval buildings; and Liège, a city with a rich industrial past and culinary fame for its delicious waffles.

Human language

Automatically Generated Multiple-Choice Question

What is the Capital of Belgium?

- A Brussels
- B Antwerp
- C Ghent
- D Liège

AI tools

How do computers understand language?

Paragraph about Belgium

Belgium is a small country in Western Europe that is distinguished by its remarkable combination of historical significance and contemporary governance. Brussels is the capital city of Belgium and serves as the seat of the EU. Other major cities include Antwerp, which is known for its diamond trade; Ghent, which has intact medieval buildings; and Liège, a city with a rich industrial past and culinary fame for its delicious waffles.

Human language

How?

AI tools

Automatically Generated Multiple-Choice Question

What is the Capital of Belgium?

- A Brussels
- B Antwerp
- C Ghent
- D Liège

Language models

Paragraph about Belgium

Belgium is a small country in Western Europe that is distinguished by its remarkable combination of historical significance and contemporary governance. Brussels is the capital city of Belgium and serves as the seat of the EU. Other major cities include Antwerp, which is known for its diamond trade; Ghent, which has intact medieval buildings; and Liège, a city with a rich industrial past and culinary fame for its delicious waffles.

Human language

How?

Language modeling

Automatically Generated Multiple-Choice Question

What is the Capital of Belgium?

- A Brussels
- B Antwerp
- C Ghent
- D Liège

Language Modeling

Language models (LMs)

Language model is a *probabilistic model* of the human language.

Language models (LMs)

Language model is a *probabilistic model* of the human language.

- 1 Predicting the next word

Language models (LMs)

Language model is a *probabilistic model* of the human language.

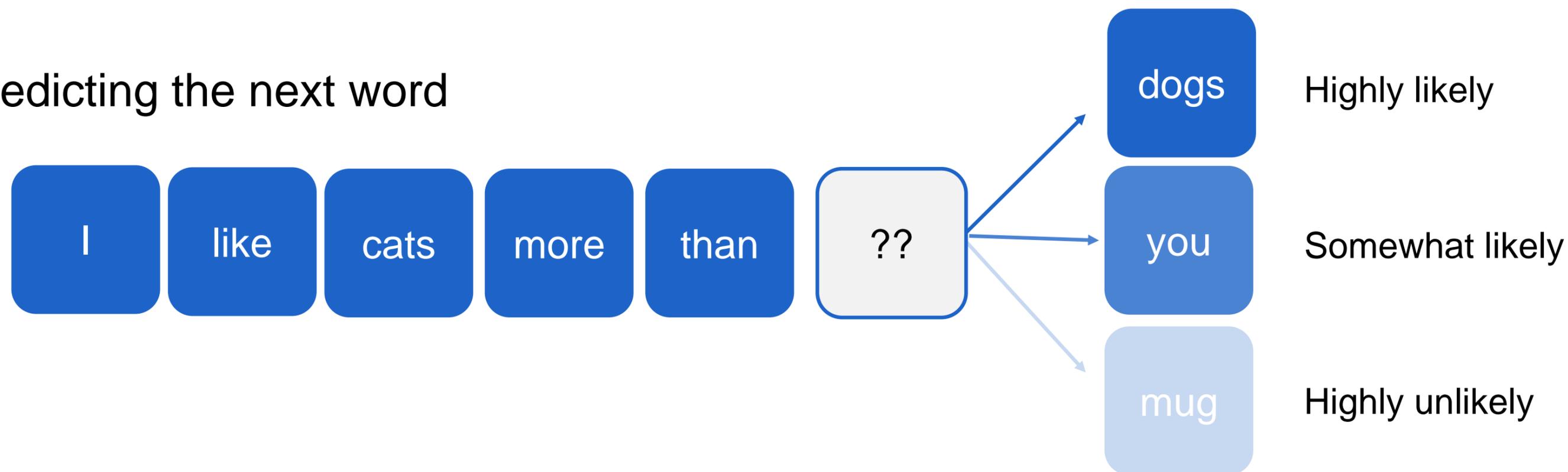
1 Predicting the next word



Language models (LMs)

Language model is a *probabilistic model* of the human language.

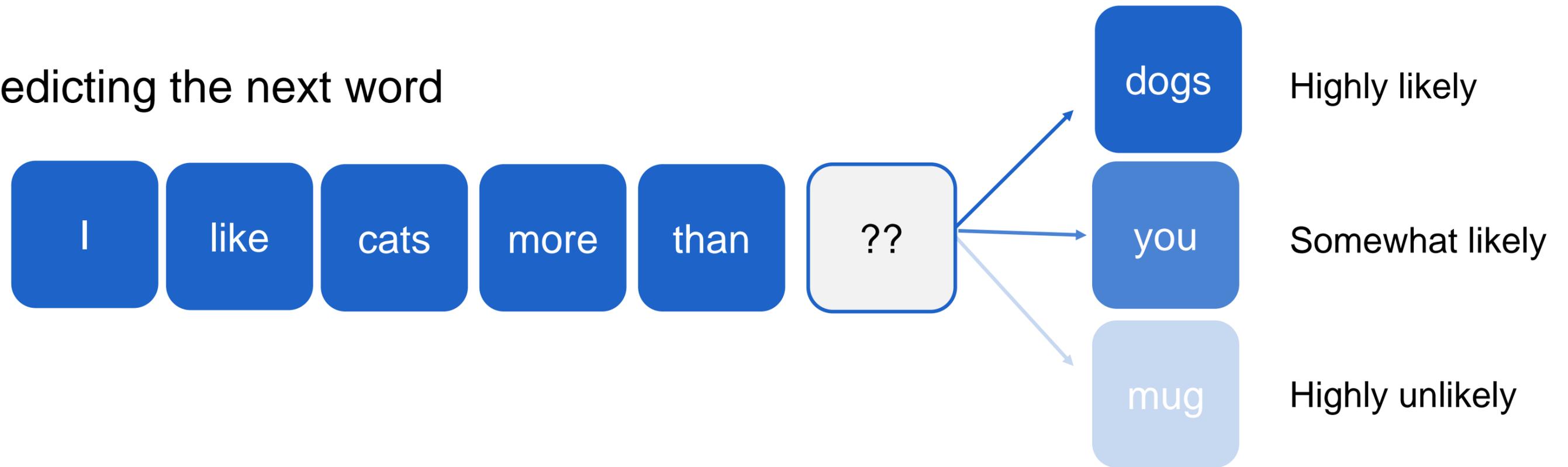
1 Predicting the next word



Language models (LMs)

Language model is a *probabilistic model* of the human language.

1 Predicting the next word

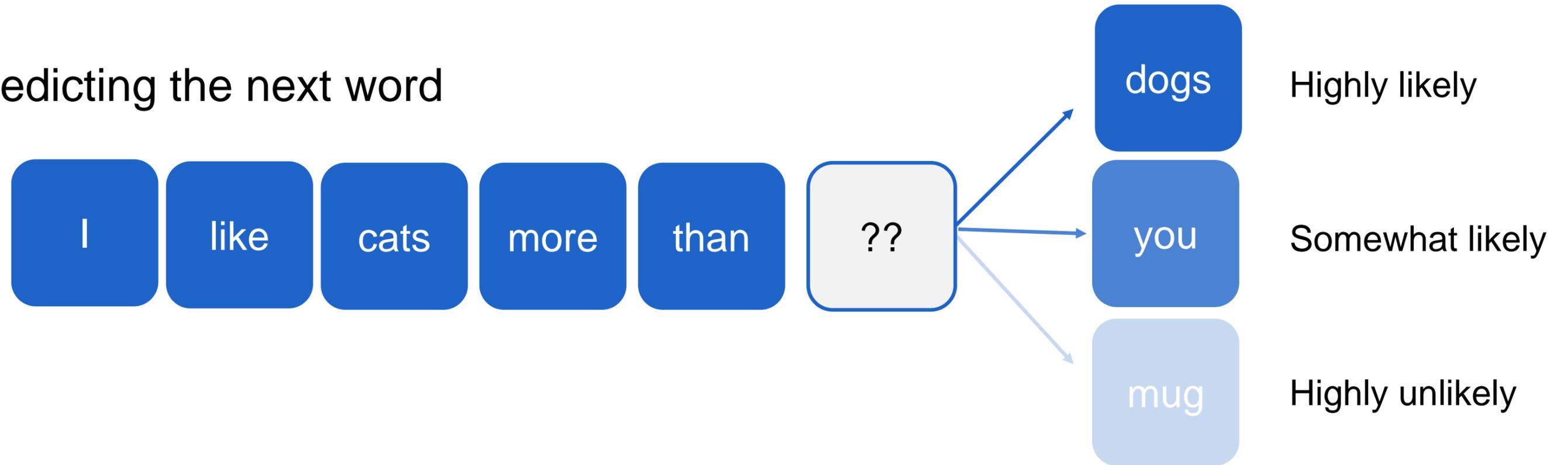


2 Likelihood of the sentence

Language models (LMs)

Language model is a *probabilistic model* of the human language.

1 Predicting the next word



2 Likelihood of the sentence

$$P(\text{The sky is the limit.}) \gg P(\text{limit is The the sky.})$$

Why assign probabilities?

Why assign probabilities?

Real-world applications

Machine Translation

Dutch ↔ English

hij heeft hoge normen. ×

he has high standards. ✓

English

he has tall standards. X

Why assign probabilities?

Real-world applications

Machine Translation

Dutch ↔ English

hij heeft hoge normen. ×

he has high standards. ✓

English

he has tall standards. X

Speech recognition



?? → I saw a van

?? → eyes awe of an

Why assign probabilities?

Real-world applications

Machine Translation

Dutch ↔ English

hij heeft hoge normen. ×

he has high standards. ✓

English

he has tall standards. X

Speech recognition



I saw a van ✓

eyes awe of an X

Why assign probabilities?

Real-world applications

Machine Translation

Dutch ↔ English

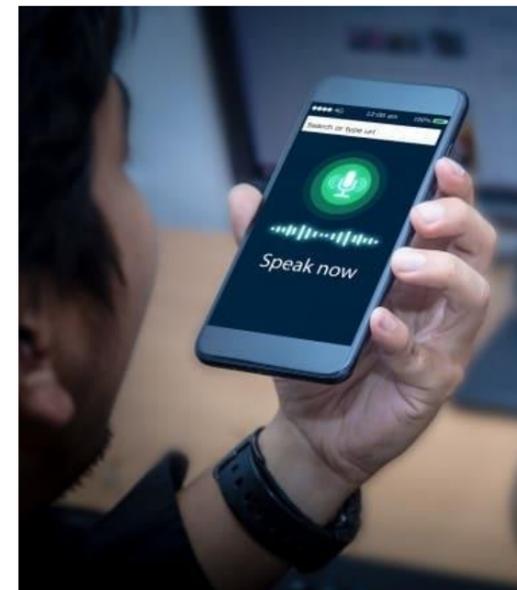
hij heeft hoge normen. ×

he has high standards. ✓

English

he has tall standards. X

Speech recognition



I saw a van ✓

eyes awe of an X

Sentence auto-completion

I had such a

great great time lovely

q w e r t y u i o p

@ # & * - + = ()

a s d f g h j k l

z x c v b n m

123 , SwiftKey .!? 48

How do LMs assign probabilities?

How do LMs ~~assign probabilities?~~ Learn?

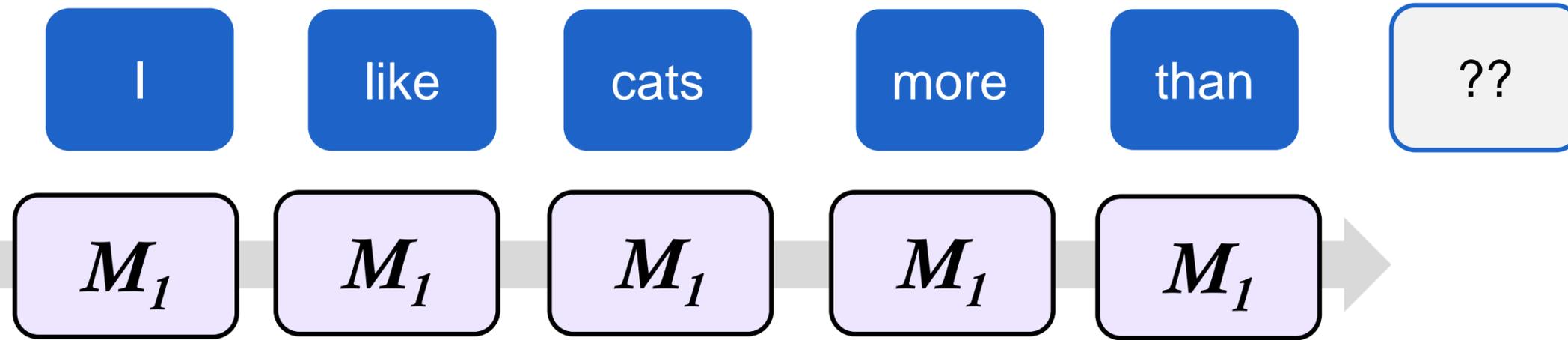
How do LMs Learn?

Two common tasks to teach LMs

- 1 *Causal Language modeling*
- 2 *Masked language modeling*

1 Causal Language Modeling

Predict the **next word** from a vocabulary using an untrained model M_1

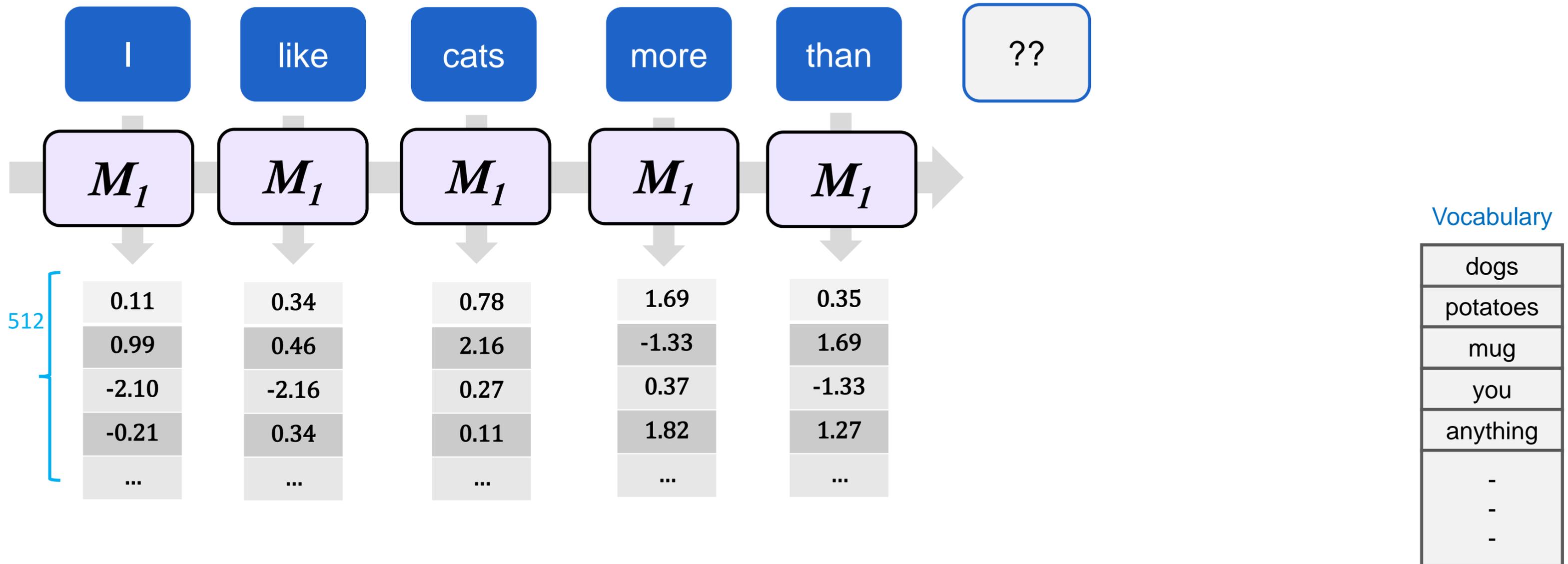


Vocabulary

dogs
potatoes
mug
you
anything
-
-
-

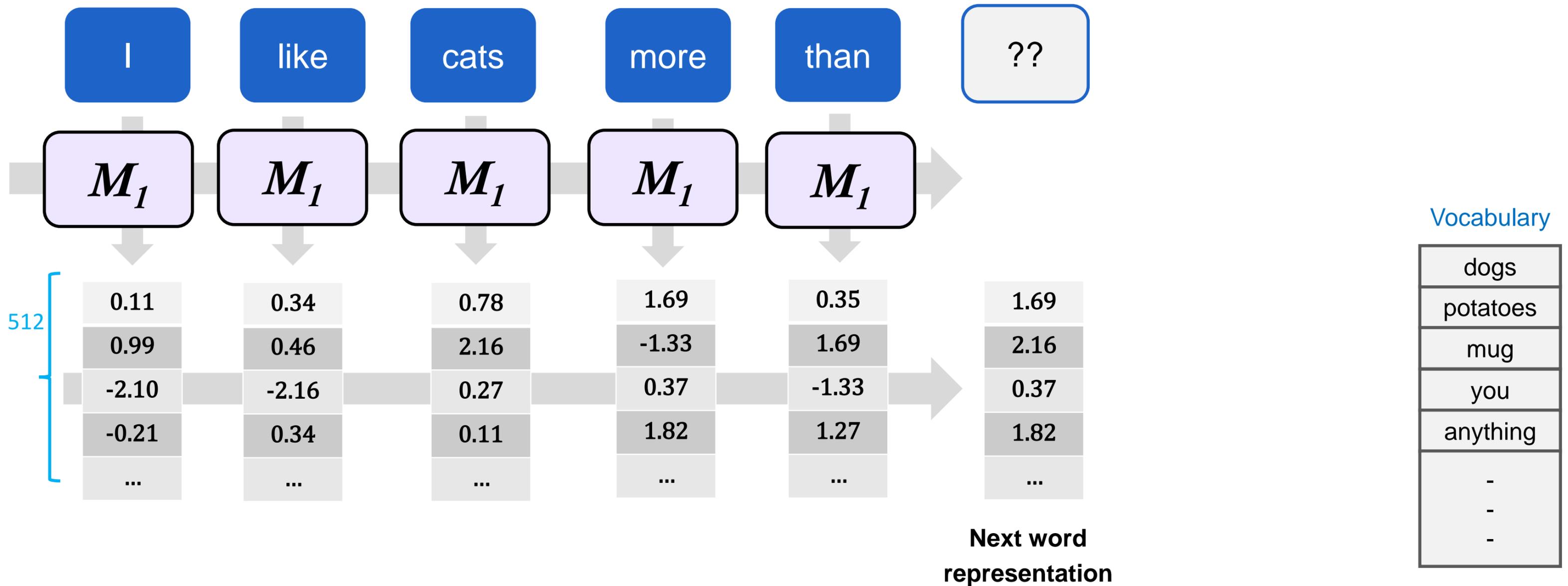
1 Causal Language Modeling

First the model M_0 converts each word to a numerical vector (e.g., 512)



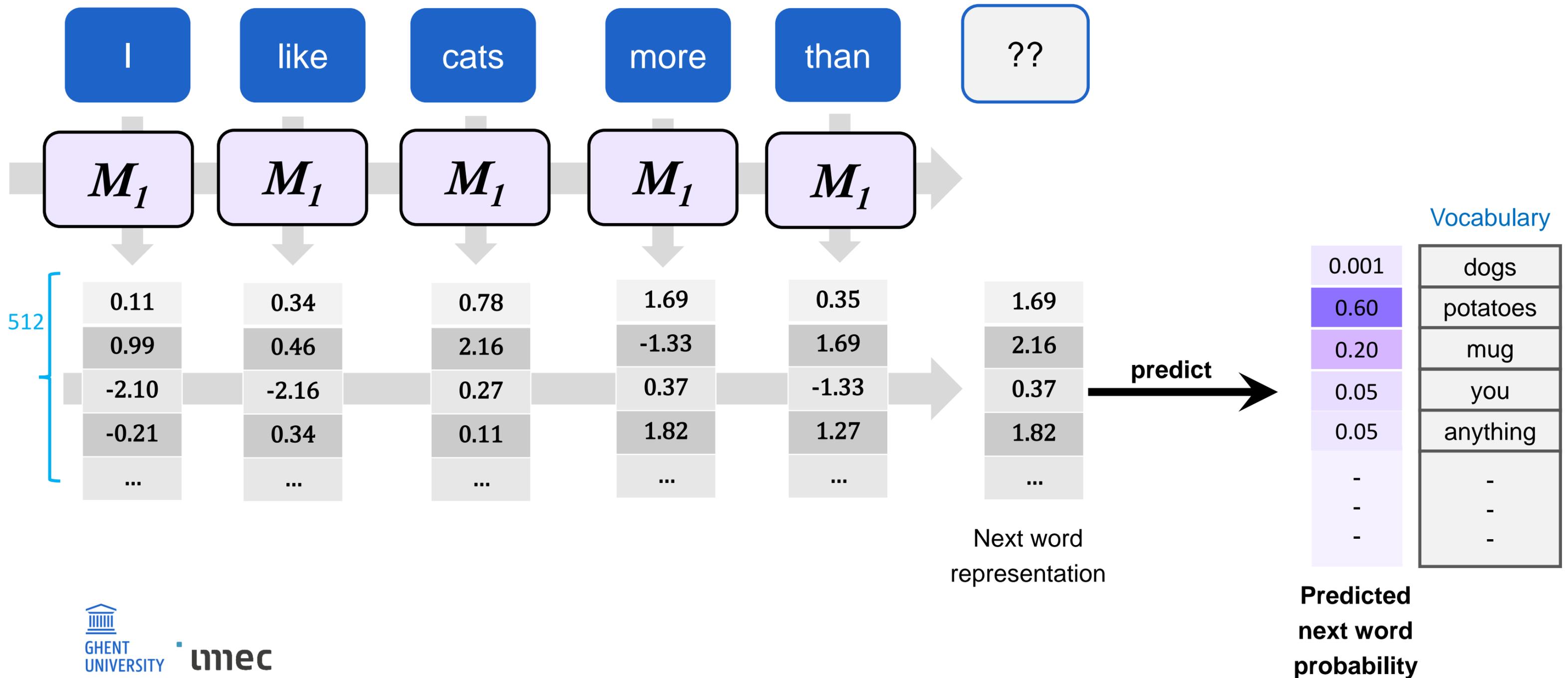
1 Causal Language Modeling

Combine all the word vectors into one **next word** representation vector



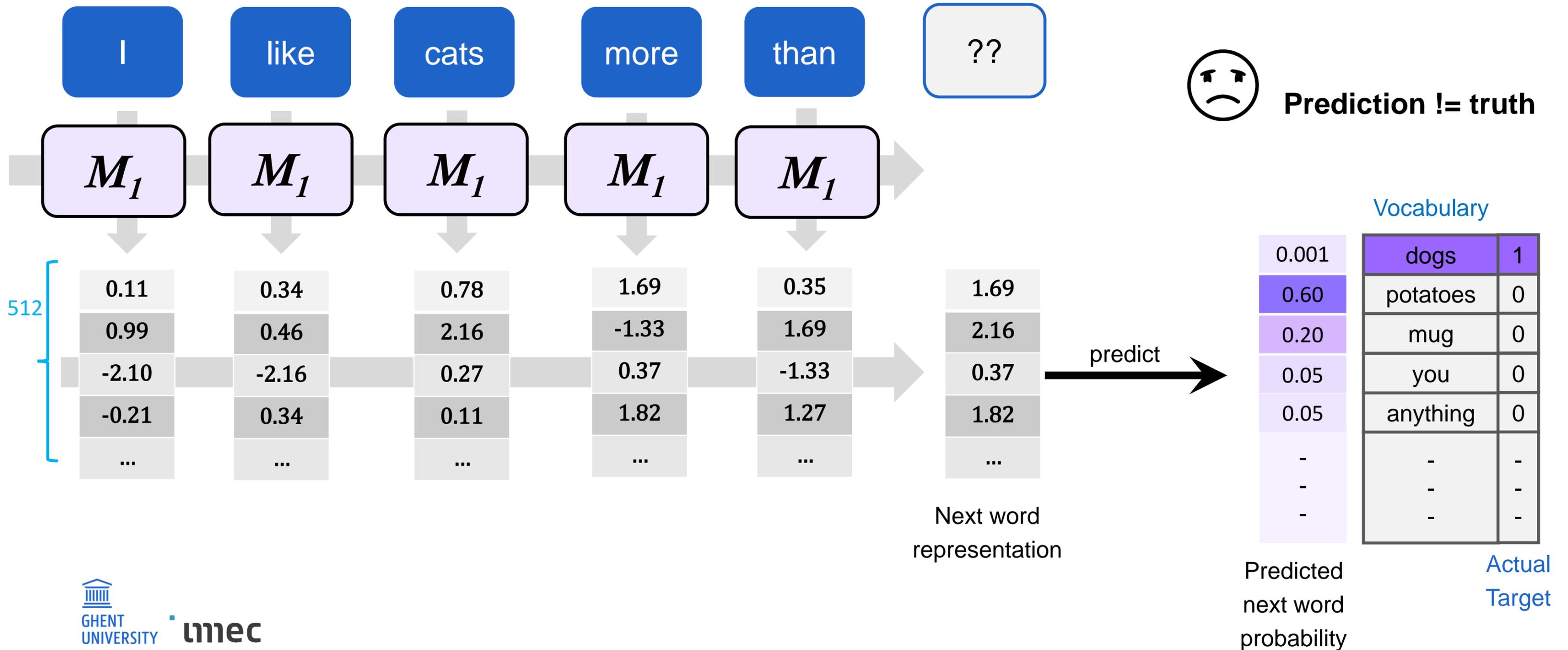
1 Causal Language Modeling

Using the next word representation, **predict** the probability of next word



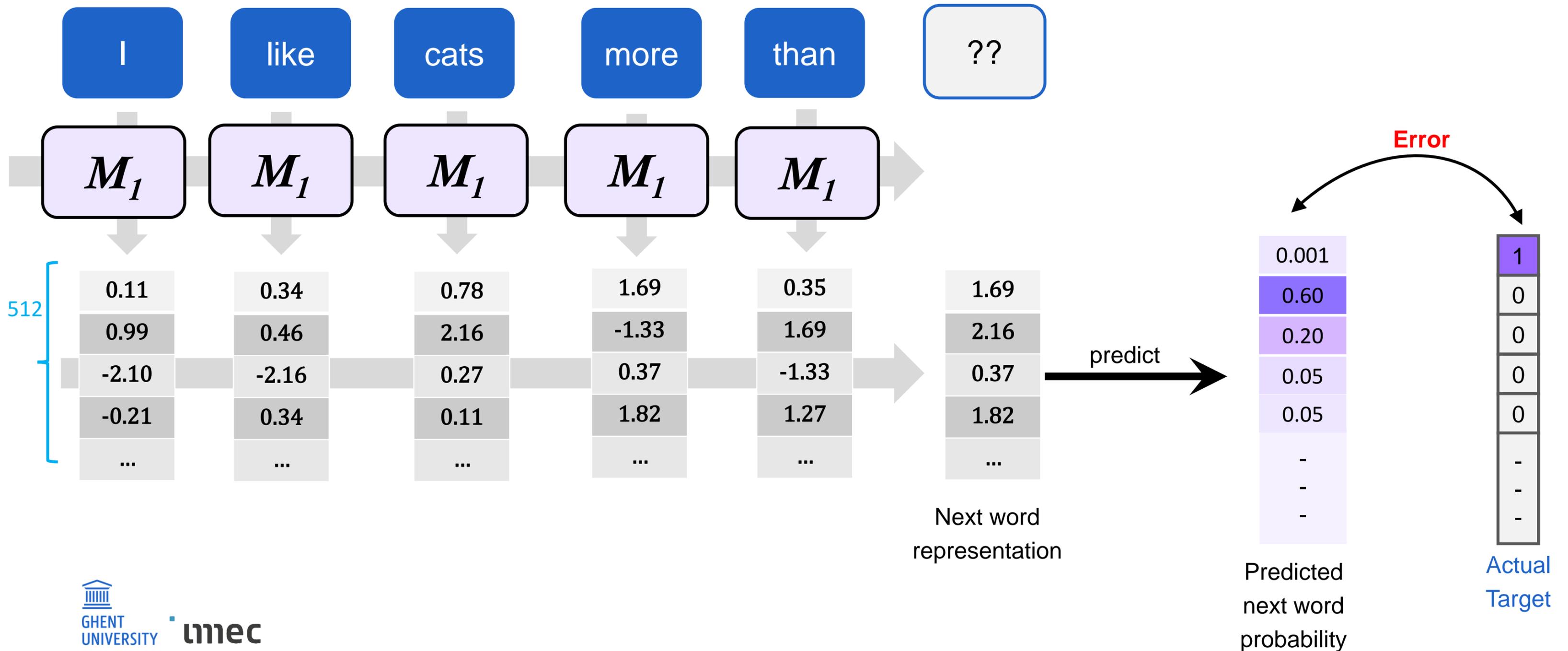
1 Causal Language Modeling

Mismatch between prediction and truth. How to **solve** this?



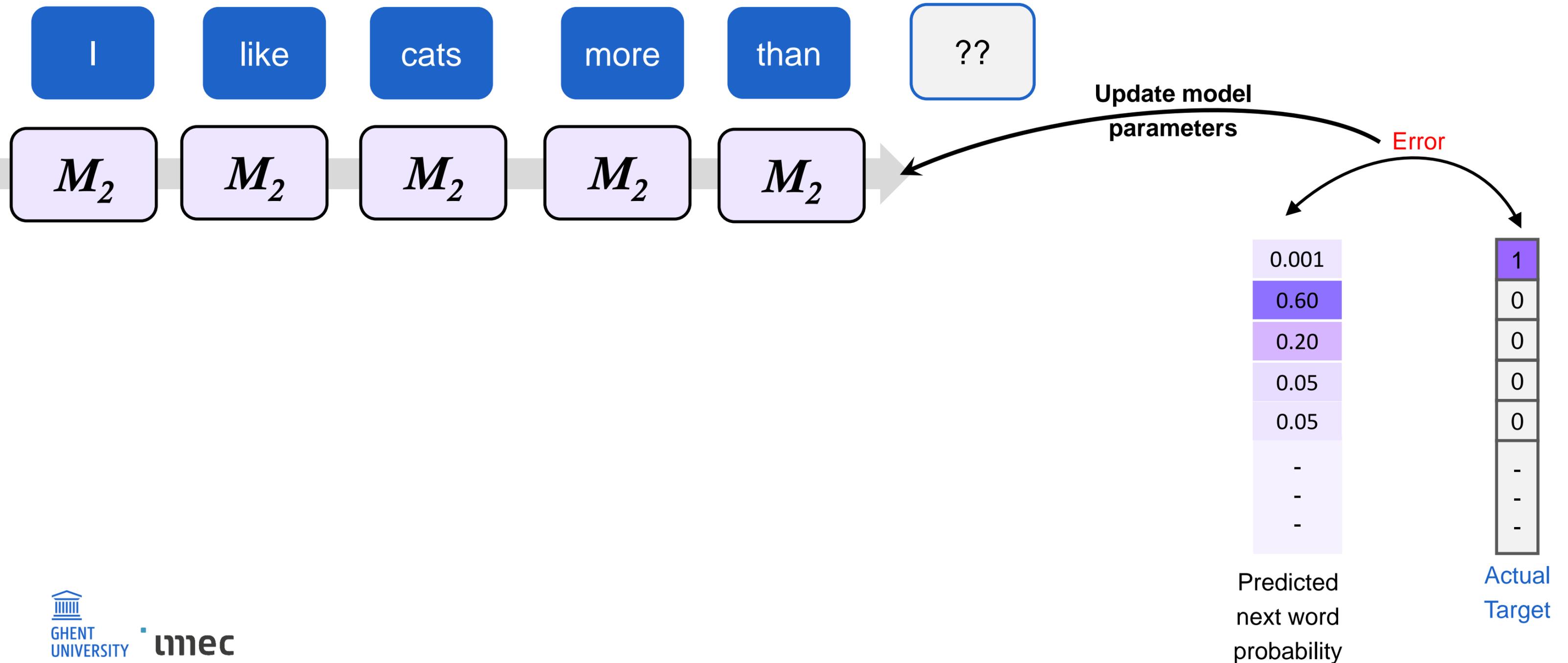
1 Causal Language Modeling

Calculate the **error (loss)** between the prediction and the actual target



1 Causal Language Modeling

Update the model's parameters $M: M_1 \rightarrow M_2$



1 Causal Language Modeling

Repeat this process to minimize the **error**

Training goal:
Minimize the error

I like cats more than ??



Update model parameters

Error

0.01
0.40
0.10
0.05
0.05
-
-
-

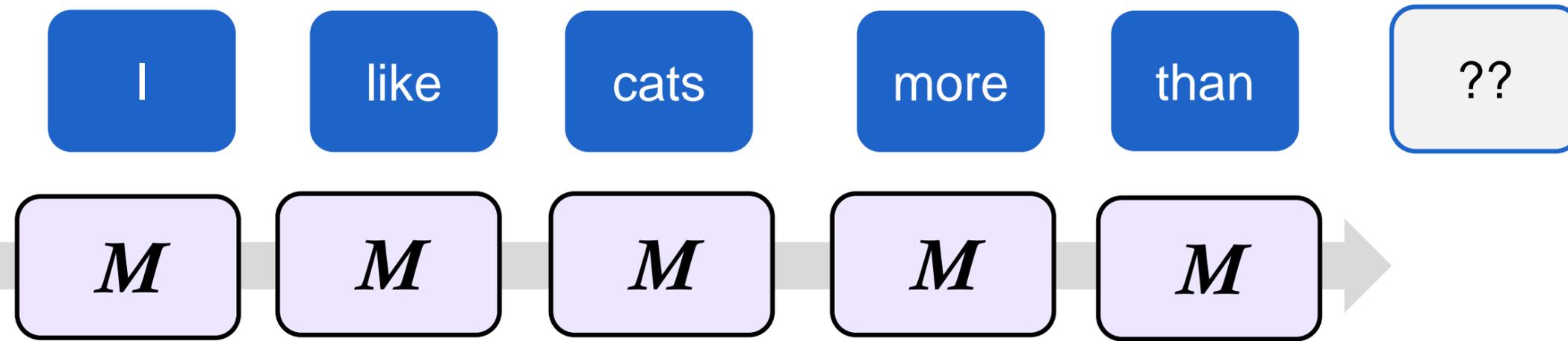
Predicted next word probability

1
0
0
0
0
-
-
-

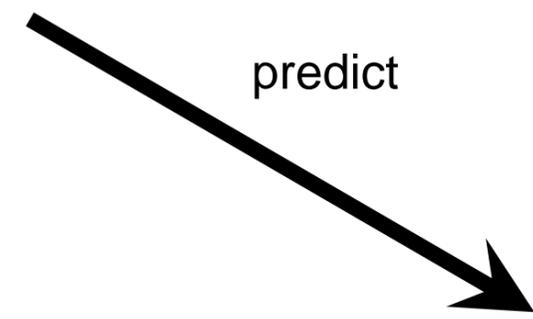
Actual Target

1 Causal Language Modeling

The model is trained after several iterations (e.g., 100 iterations)



LM prediction similar to truth



Vocabulary

0.93	dogs	1
0.0001	potatoes	0
0.0002	mug	0
0.03	you	0
0.01	anything	0
-	-	-
-	-	-
-	-	-

Predicted next word probability

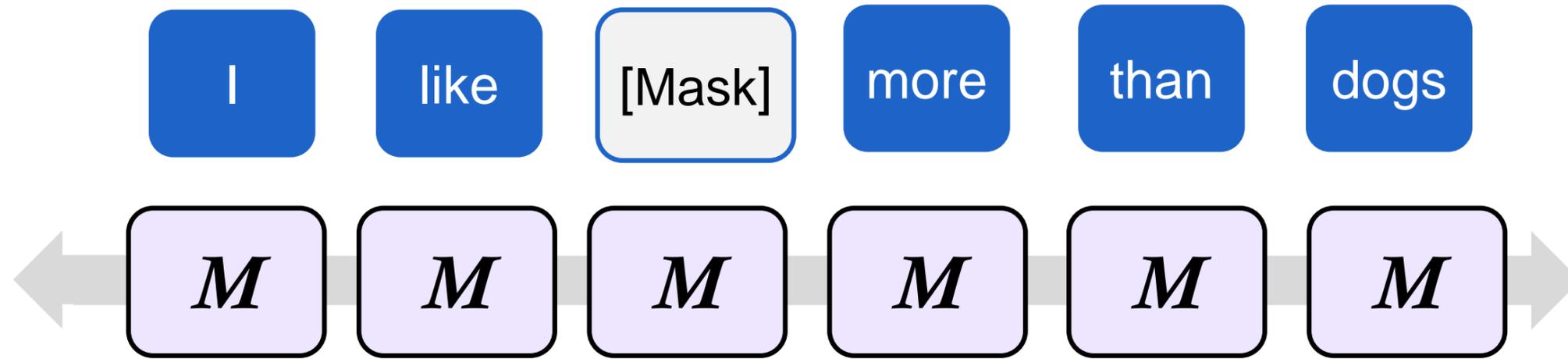
Actual Target

1 Causal Language Modeling

- Processes text from left to right
- Natural for generative tasks (e.g., [Question Generation](#))
- Examples: [GPT-3/4](#), [ChatGPT](#)

2 Masked Language Modeling

Predict the **masked** word

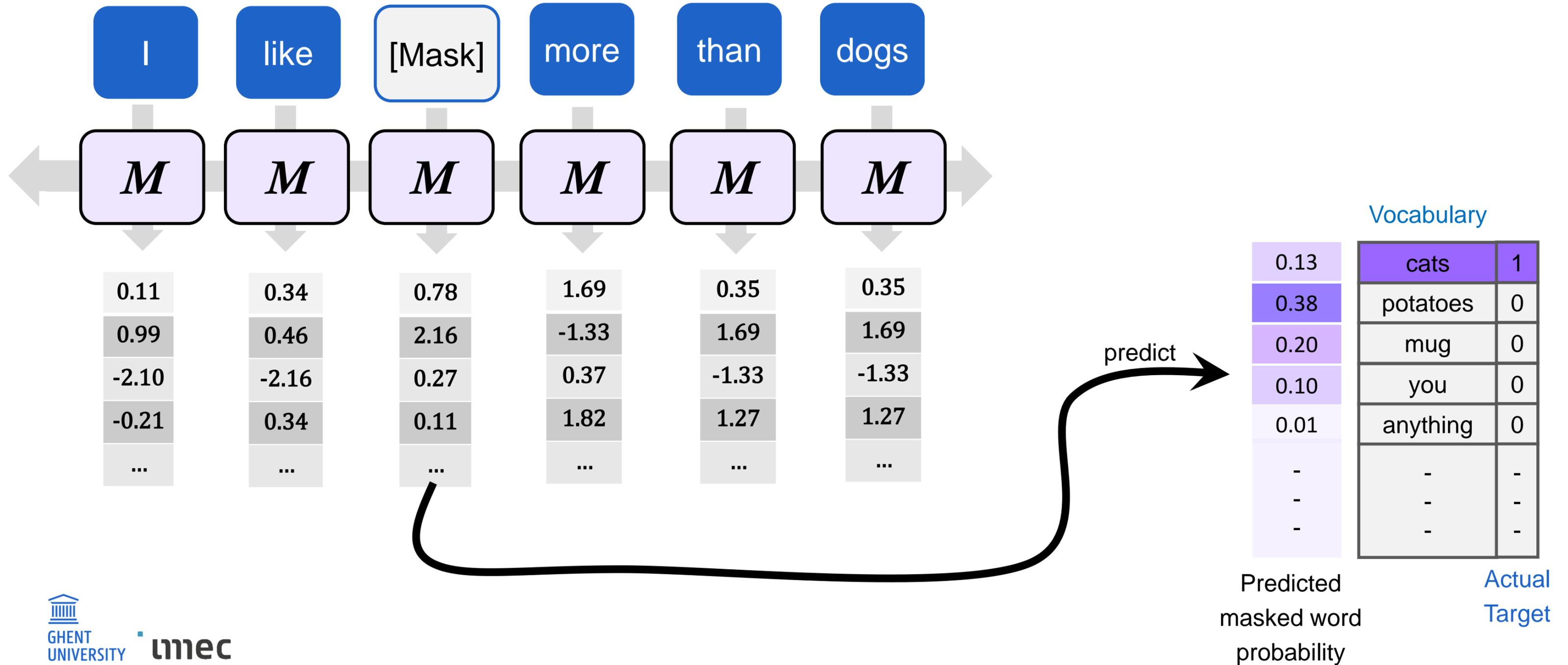


Vocabulary

cats
potatoes
mug
you
anything
-
-
-

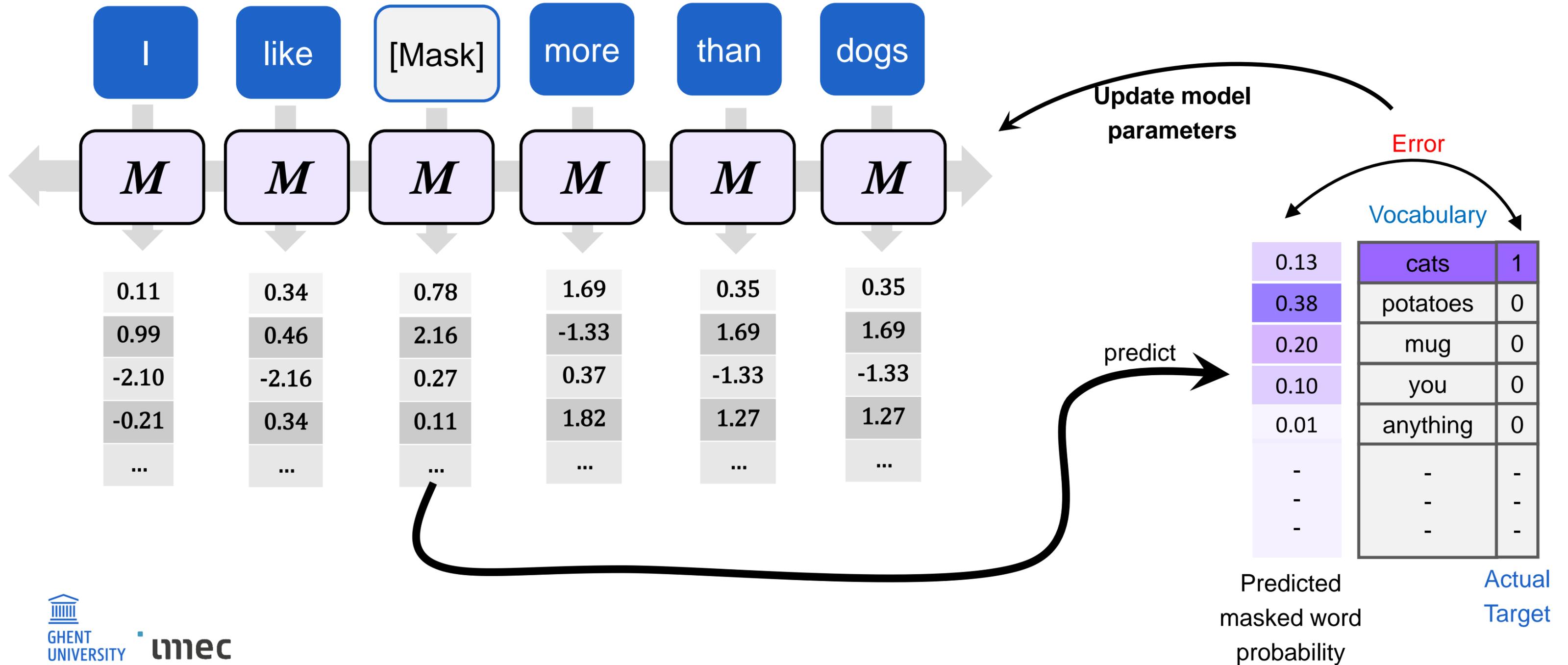
2 Masked Language Modeling

Predict the **masked** word



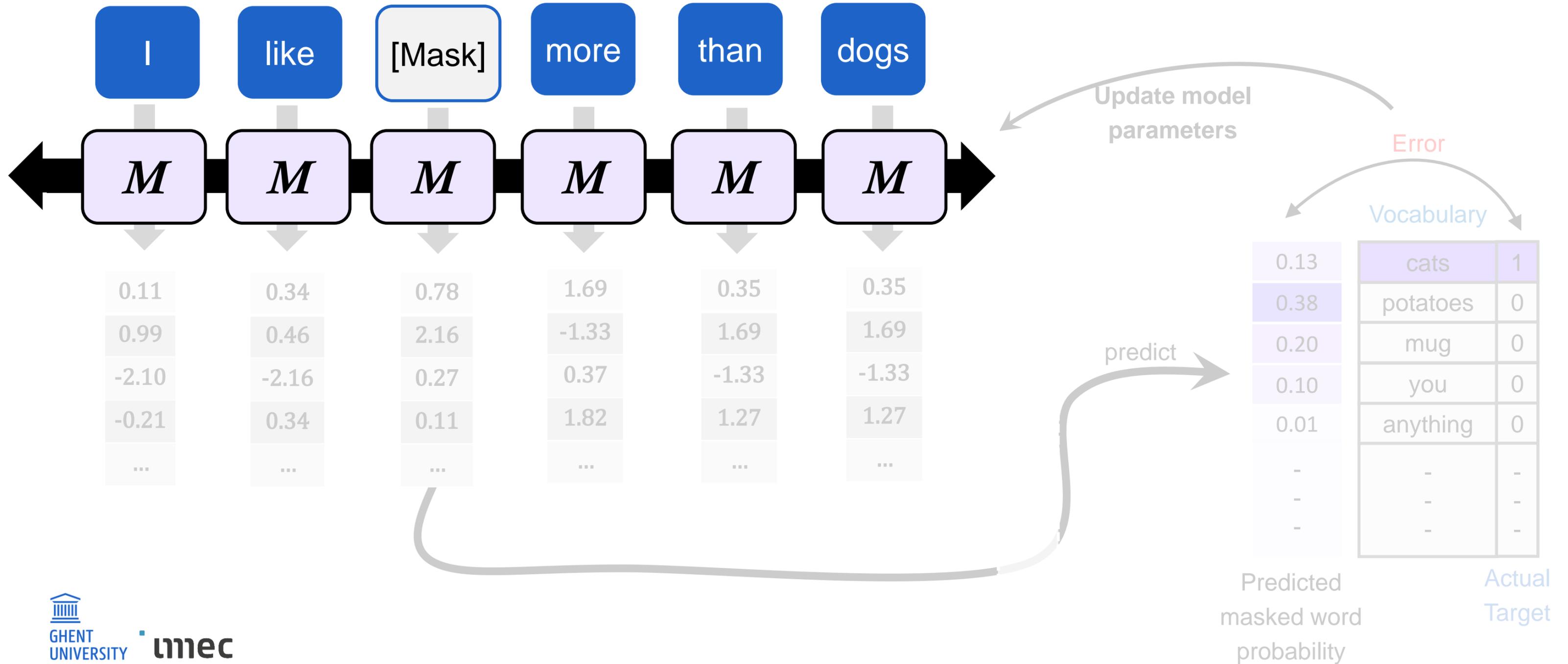
2 Masked Language Modeling

Predict the **masked** word



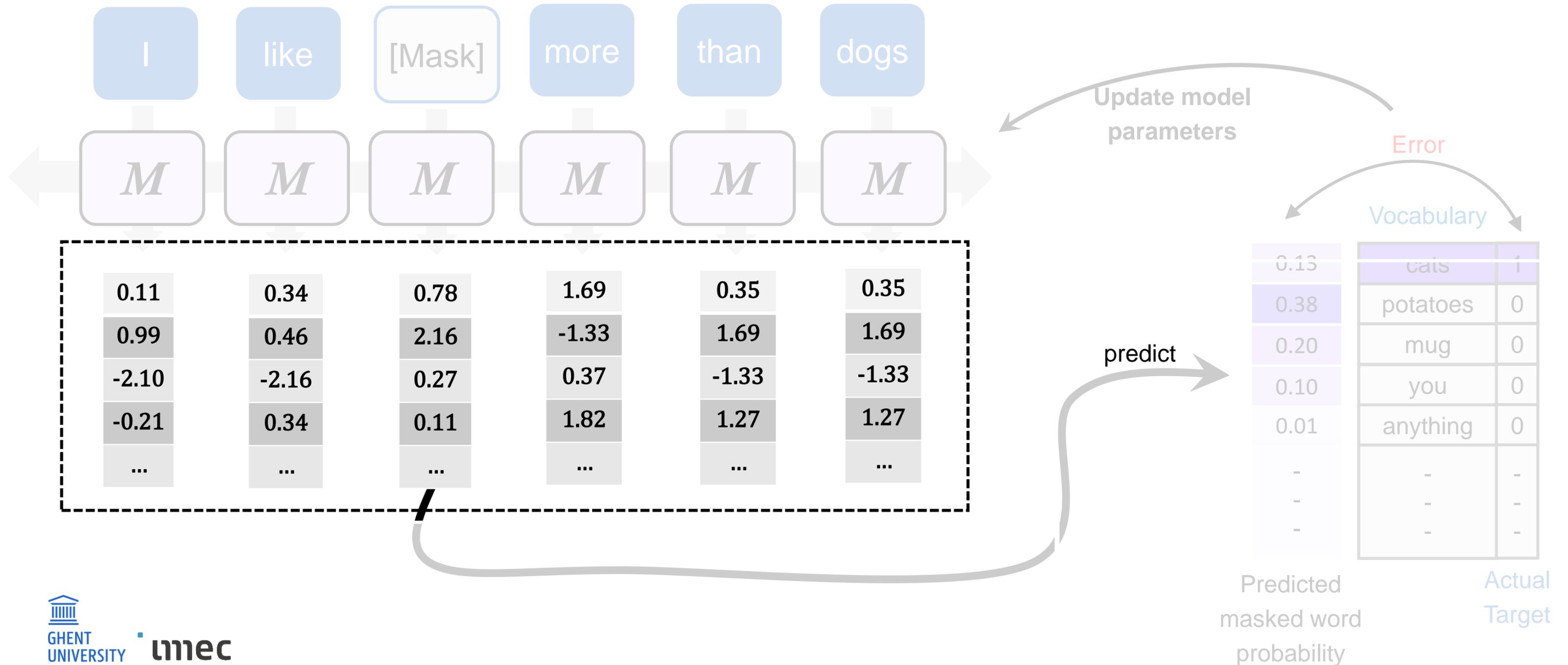
2 Masked Language Modeling

Bidirectional processing of text



2 Masked Language Modeling

- Suitable for text understanding tasks (e.g., question answering, gap-fill prediction)



2 Masked Language Modeling

- Suitable for text understanding tasks (e.g., question answering, gap-fill prediction)
- Examples: BERT, RoBERTa

Large Language Models (LLMs)

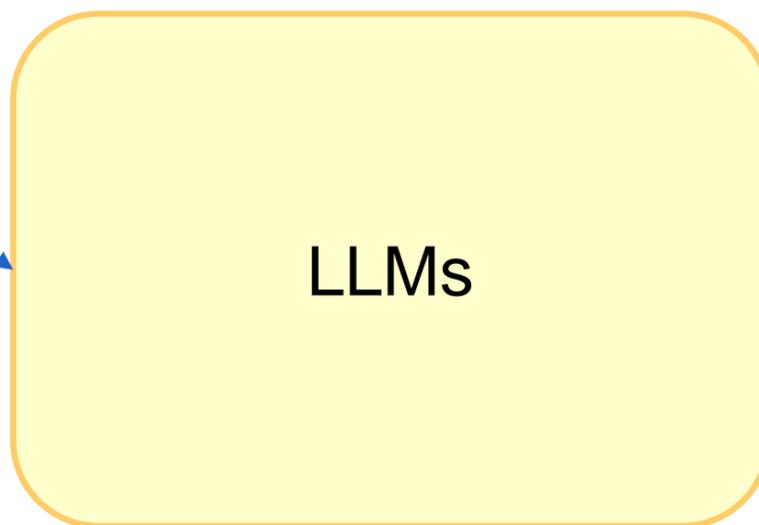
LLMs are pretrained on tons of textual data



WIKIPEDIA

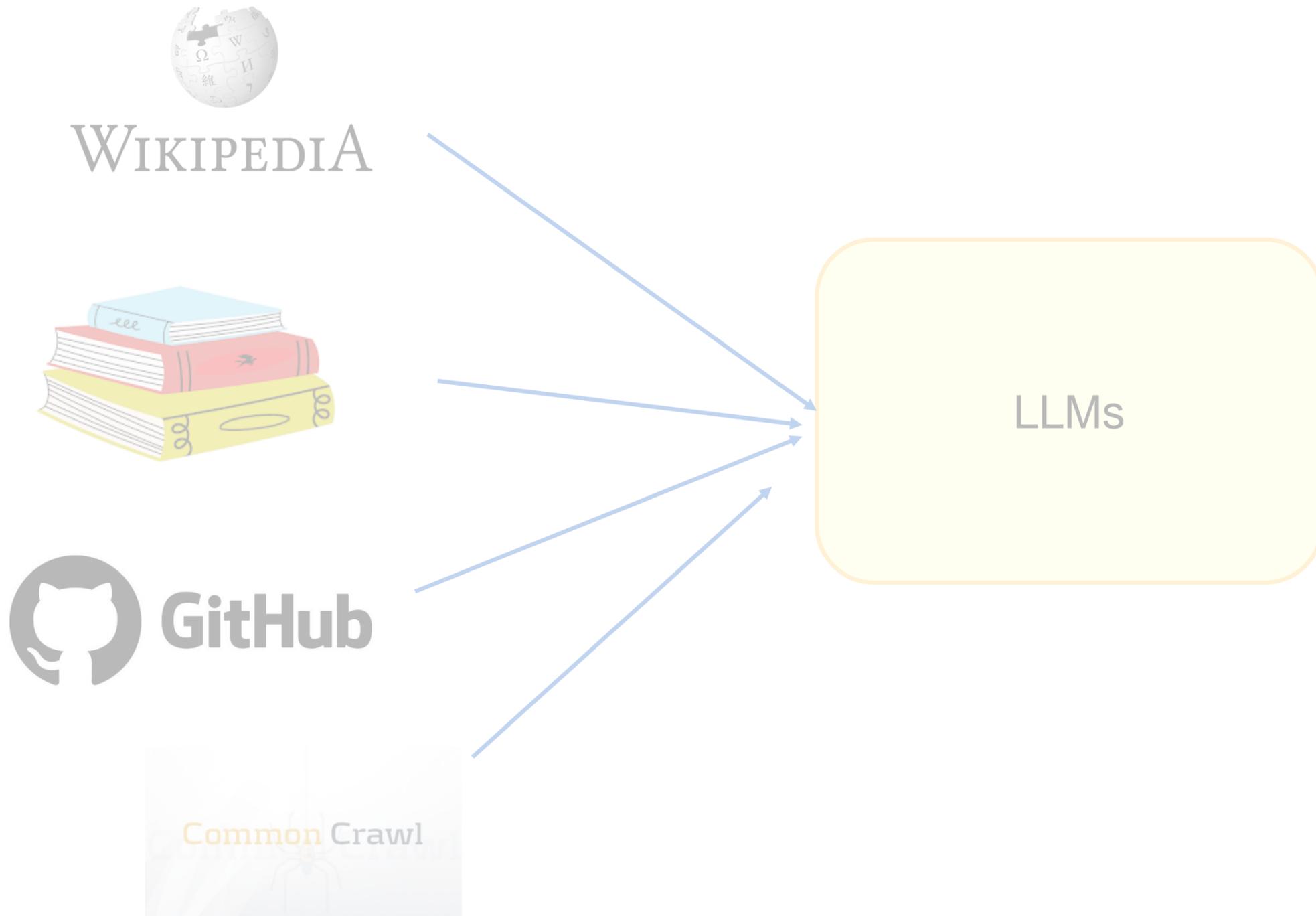


GitHub



Large Language Models (LLMs)

LLMs are pretrained on tons of textual data



Examples of LLMs



Distractor Generation Task

Distractor Generation

Multiple-choice question (**MCQ**) has

Distractor Generation

Multiple-choice question (**MCQ**) has

What is the capital of Belgium?

- A Brussels
- B Antwerp
- C Ghent
- D Amsterdam

Distractor Generation

Multiple-choice question (**MCQ**) has

What is the capital of Belgium?

Question

A Brussels

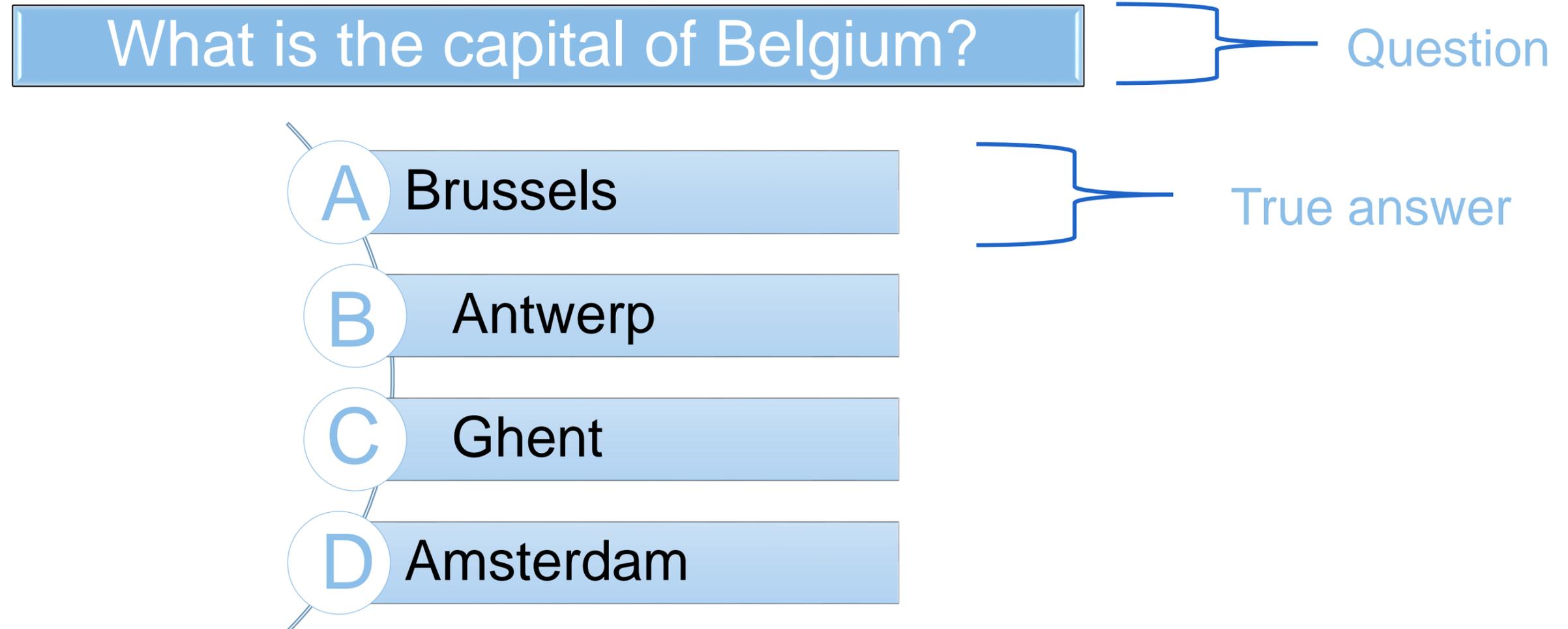
B Antwerp

C Ghent

D Amsterdam

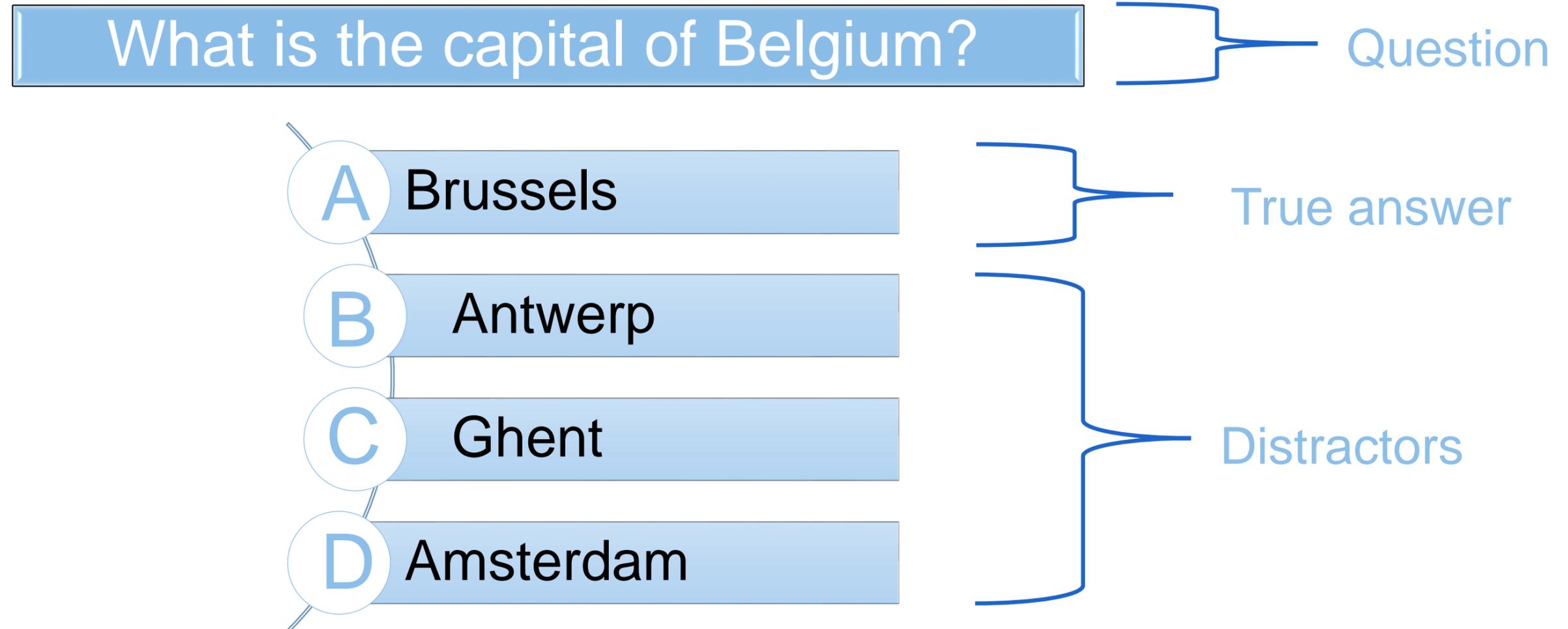
Distractor Generation

Multiple-choice question (**MCQ**) has



Distractor Generation

Multiple-choice question (**MCQ**) has



Quality of Distractors

Effective Distractors

What is the Capital of Belgium?

- A Brussels
- B Antwerp
- C Ghent
- D Amsterdam

Poor Distractors

What is the Capital of Belgium?

- A Brussels
- B Sydney
- C New York
- D Mekelle

Quality of Distractors

Effective Distractors

What is the Capital of Belgium?

- A Brussels
- B Antwerp
- C Ghent
- D Amsterdam

- ✓ Plausible
- ✓ Tests knowledge
- ✓ Good MCQ

Poor Distractors

What is the Capital of Belgium?

- A Brussels
- B Sydney
- C New York
- D Mekelle

- × Implausible
- × Easily dismissed
- × Bad MCQ

Quality of Distractors

Effective Distractors

What is the Capital of Belgium?

A Brussels

B Antwerp

C Ghent

D Amsterdam

Generating distractors → **time-consuming task!**

Quality of Distractors

Effective Distractors

What is the Capital of Belgium?

A Brussels

B Antwerp

C Ghent

D Amsterdam

Generating distractors → **time-consuming task!**

Goal: Can we automate this process to **increase** teachers' efficiency?

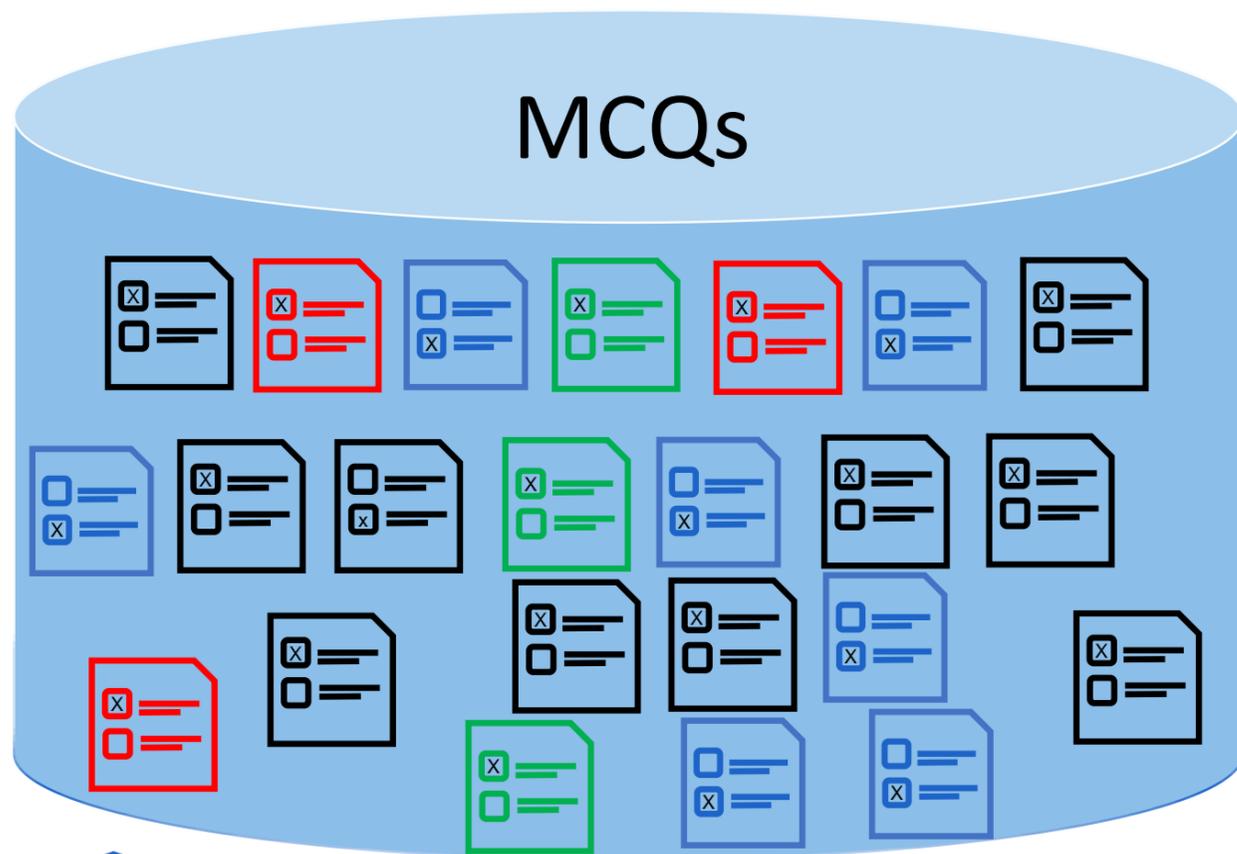
Research question

1. How to **automatically** generate distractors?

Televic Dataset

Real-world dataset provided by industry partner

televic

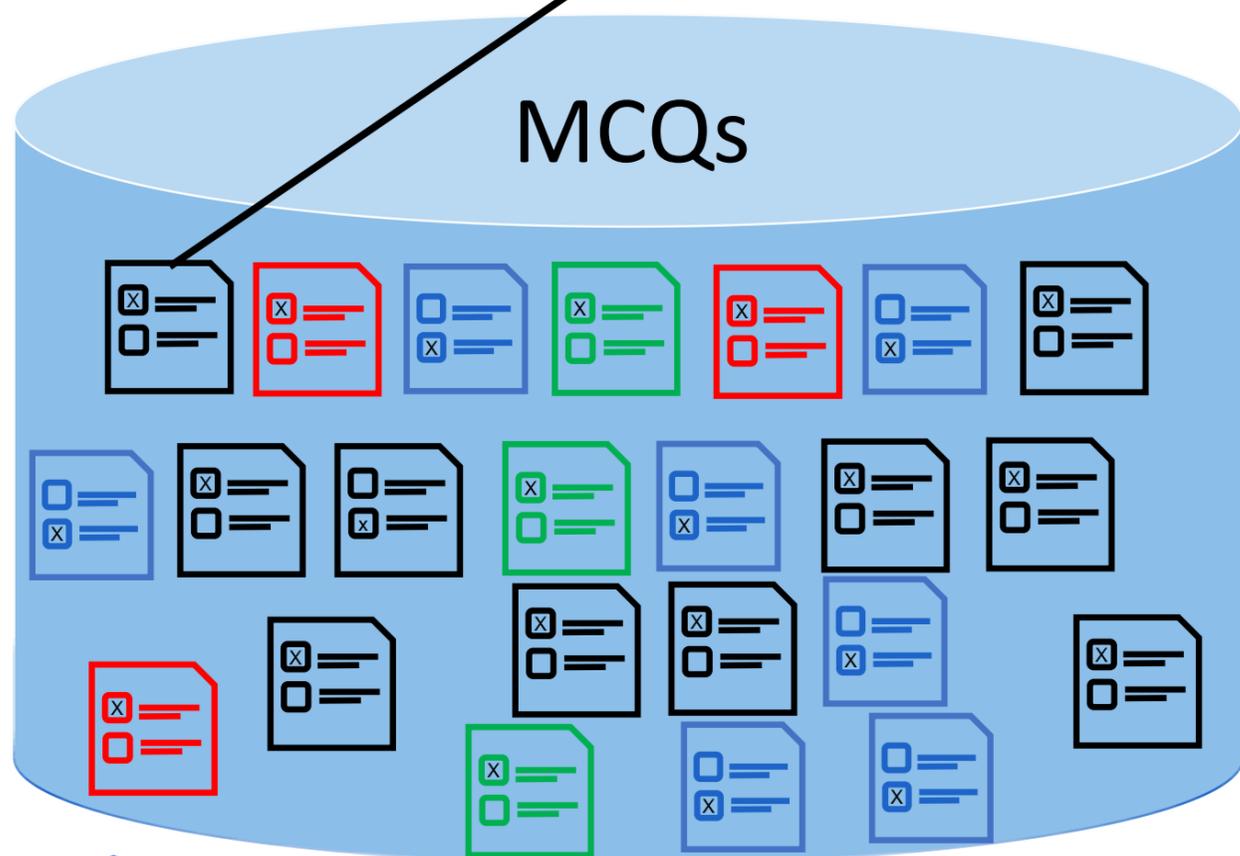


Televic Dataset

The highest mountain in the world is ____

- A. Mont Blanc
- B. Kilimanjaro
- C. **Mount Everest**
- D. Amba Alaje

MCQs



Televic Dataset

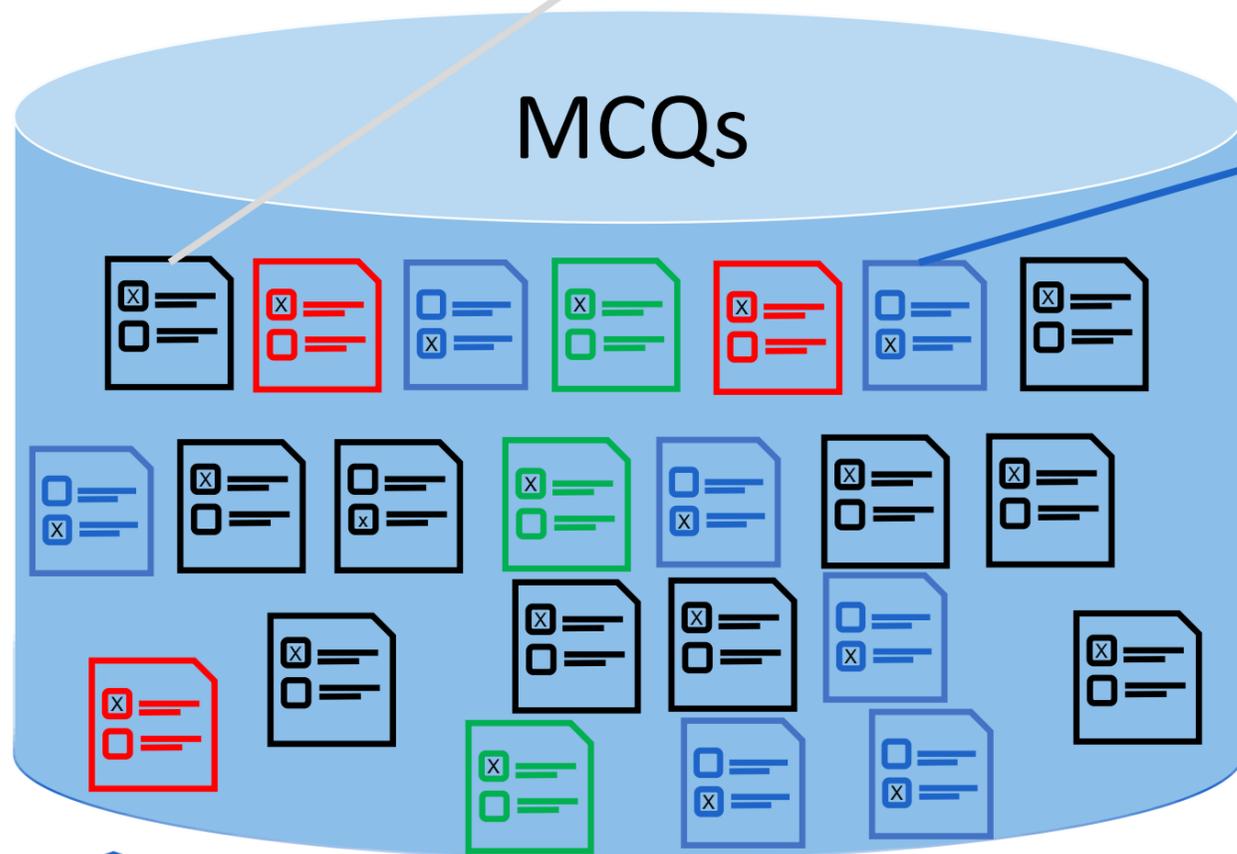
The highest mountain in the world is ____

- A. Mont Blanc
- B. Kilimanjaro
- C. Mount Everest
- D. Amba Alaje

In welk werelddeel ligt Noord-Korea?

- A. Antarctica
- B. Europa
- C. Afrika
- D. Azie

MCQs



Televic Dataset

The highest mountain in the world is ____

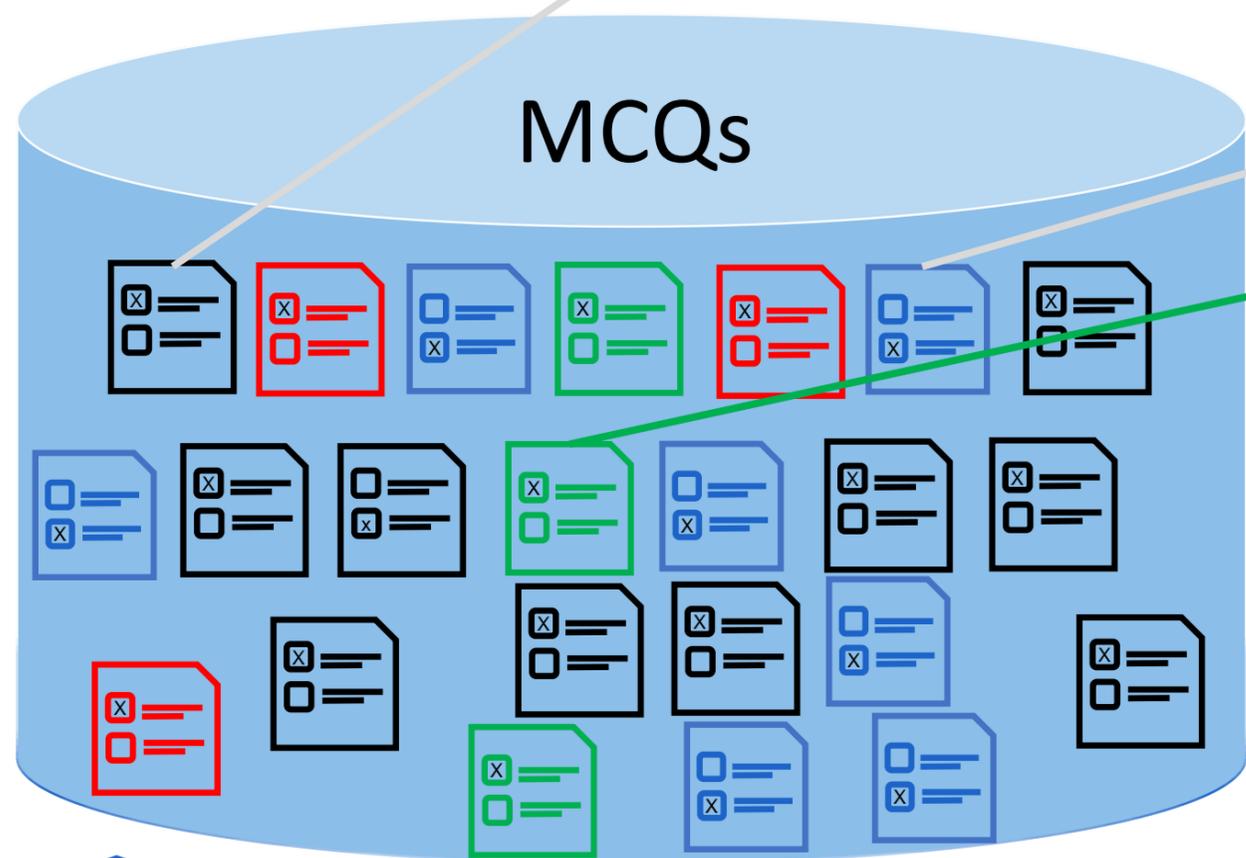
- A. Mont Blanc
- B. Kilimanjaro
- C. Mount Everest
- D. Amba Alaje

In welk werelddeel ligt Noord-Korea?

- A. Antarctica
- B. Europa
- C. Afrika
- D. Azie

By this time next month we ____ the designs.

- A. will have finished
- B. are finishing
- C. will finish
- D. go to finish



Televic Dataset

The highest mountain in the world is ____

- A. Mont Blanc
- B. Kilimanjaro
- C. Mount Everest
- D. Amba Alaje

In welk werelddeel ligt Noord-Korea?

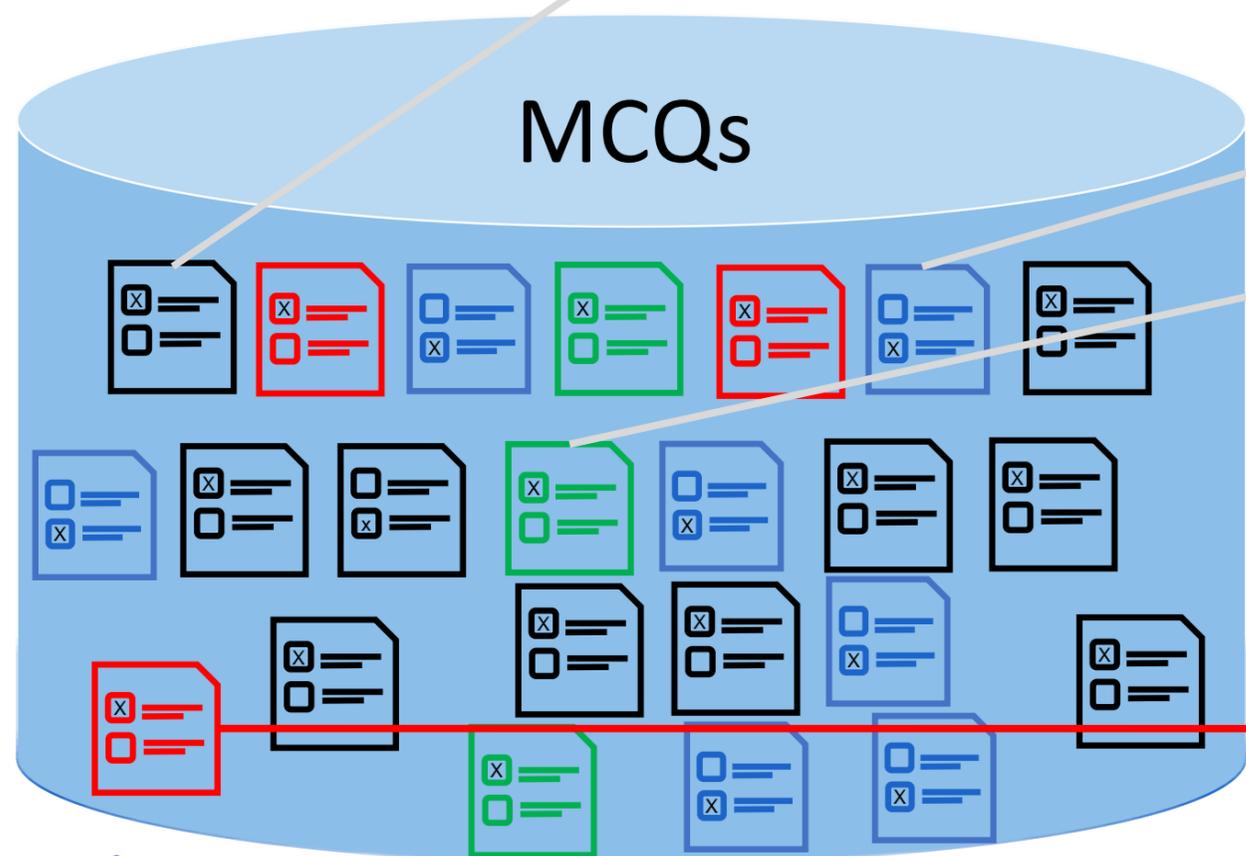
- A. Antarctica
- B. Europa
- C. Afrika
- D. Azie

By this time next month we ____ the designs.

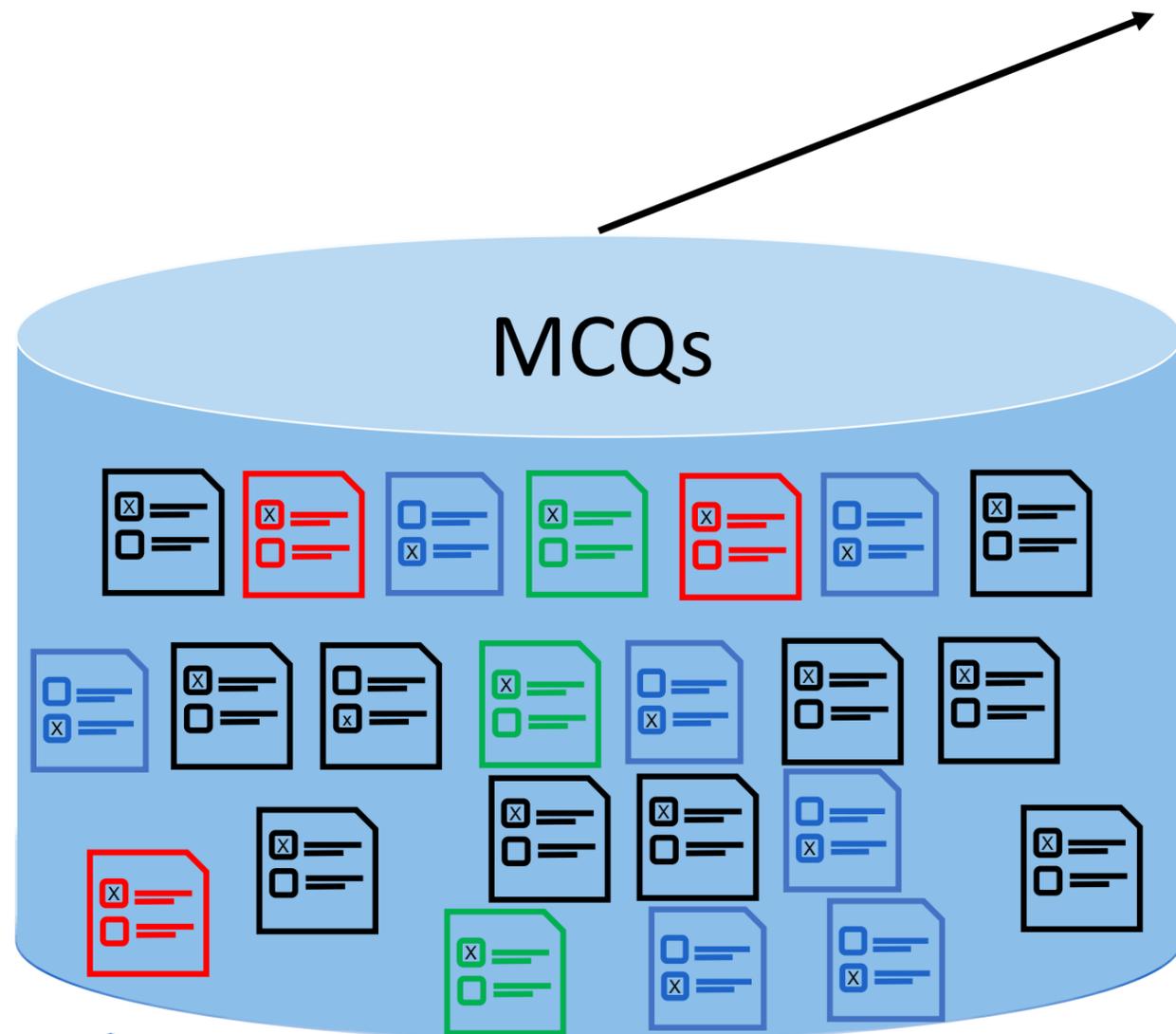
- A. will have finished
- B. are finishing
- C. will finish
- D. go to finish

Weet je nog wat de eerste mensensoort was?

- A. Homo erectus
- B. Homo sapiens
- C. Homo habilis
- D. Homo neanderthalensis



Televic Dataset



Total of
~ 62,000 MCQs

Proposed Solution

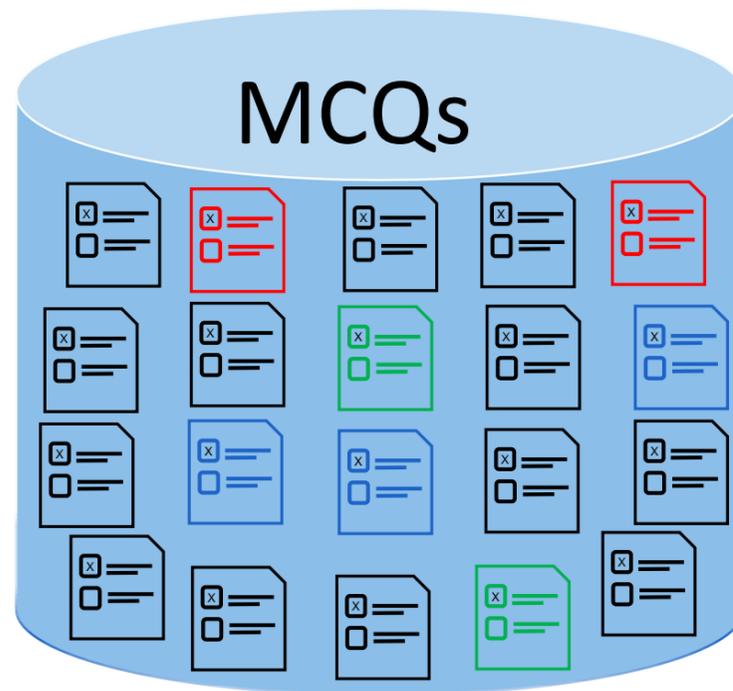
Learn to rank the distractors according to their relevance to a question

Learning to rank distractors

Language Model (**LM**) based model ranks **existing** distractors



Q: The capital of Belgium is ...
A: Brussels



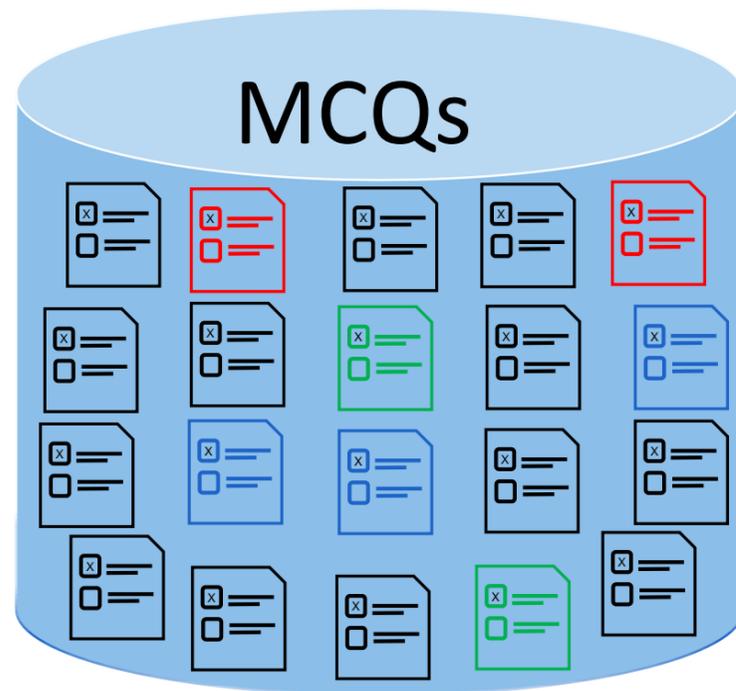
LM Ranking
Model

Learning to rank distractors

Language Model (LM) based model ranks **existing** distractors



Q: The capital of Belgium is ...
A: Brussels



LM Ranking Model

Ranked Distractors

1. Antwerp
2. Ghent
3. Amsterdam
- 4.
-
-
-
77008. will finish
77009. kern
77010. Ribosomen

Our approach

We propose **three** models

Our approach (Q-SIM)

- 1 Question Similarity (Q-SIM) model
 - Questions are **similar** if they share a distractor or an **answer**

Our approach (Q-SIM)

1 Question Similarity (Q-SIM) model

- Questions are **similar** if they share a distractor or an **answer**

What is the capital of the Netherlands?

- A. The Hague C. Amsterdam
B. Rotterdam D. Brussels

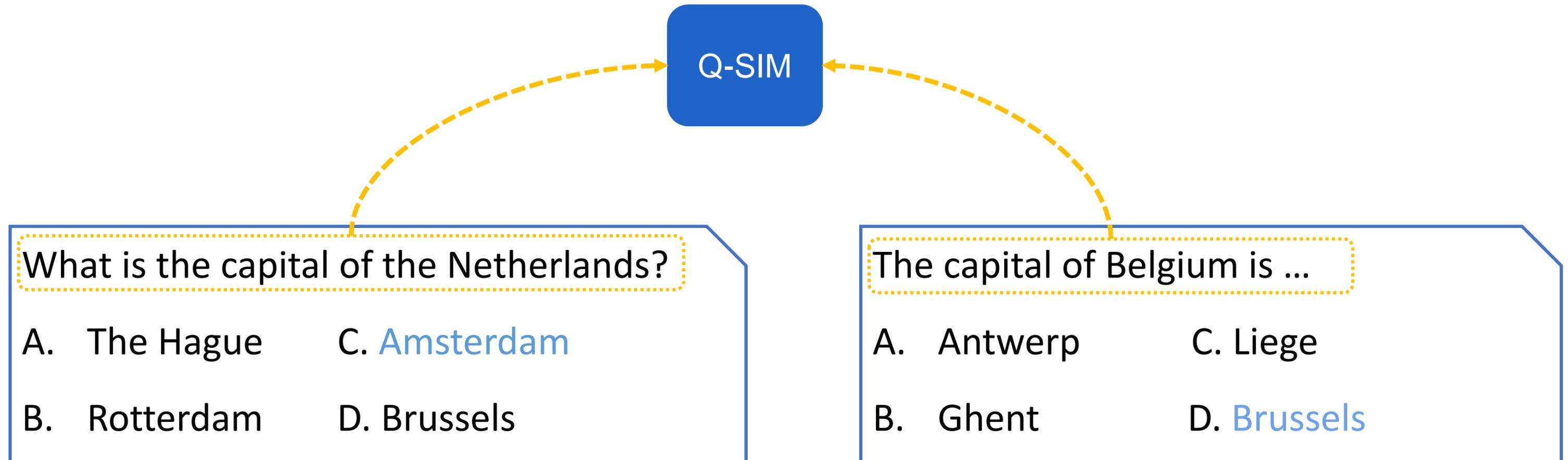
The capital of Belgium is ...

- A. Antwerp C. Liege
B. Ghent D. Brussels

Our approach (Q-SIM)

1 Question Similarity (Q-SIM) model

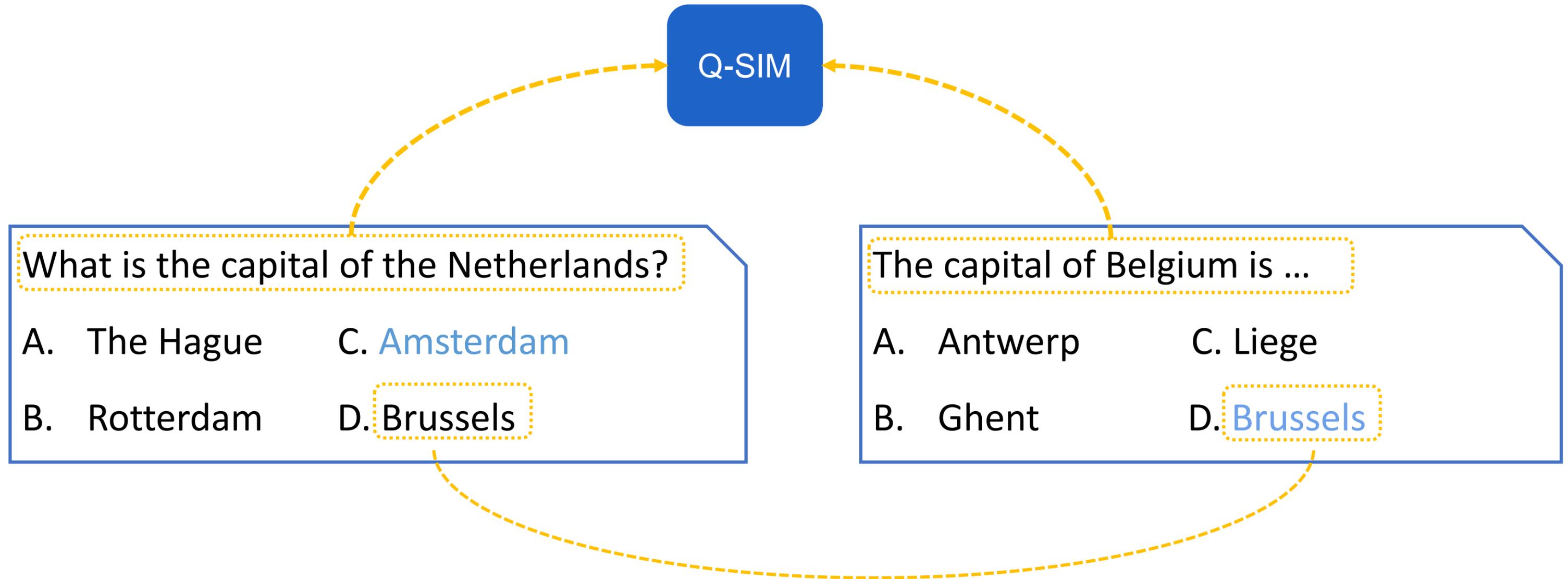
- Questions are **similar** if they share a distractor or an **answer**



Our approach (Q-SIM)

1 Question Similarity (Q-SIM) model

- Questions are **similar** if they share a distractor or an **answer**



Because Brussels is shared

Our approach (Q-SIM)

1 Question Similarity (Q-SIM) model

- Q-SIM clusters **similar** questions together

Q-SIM



Our approach (D-SIM)

2 Distractor Similarity (D-SIM) model

- Distractors are **similar** if they co-occur within the same MCQ

Our approach (D-SIM)

2 Distractor Similarity (D-SIM) model

- Distractors are **similar** if they co-occur within the same MCQ

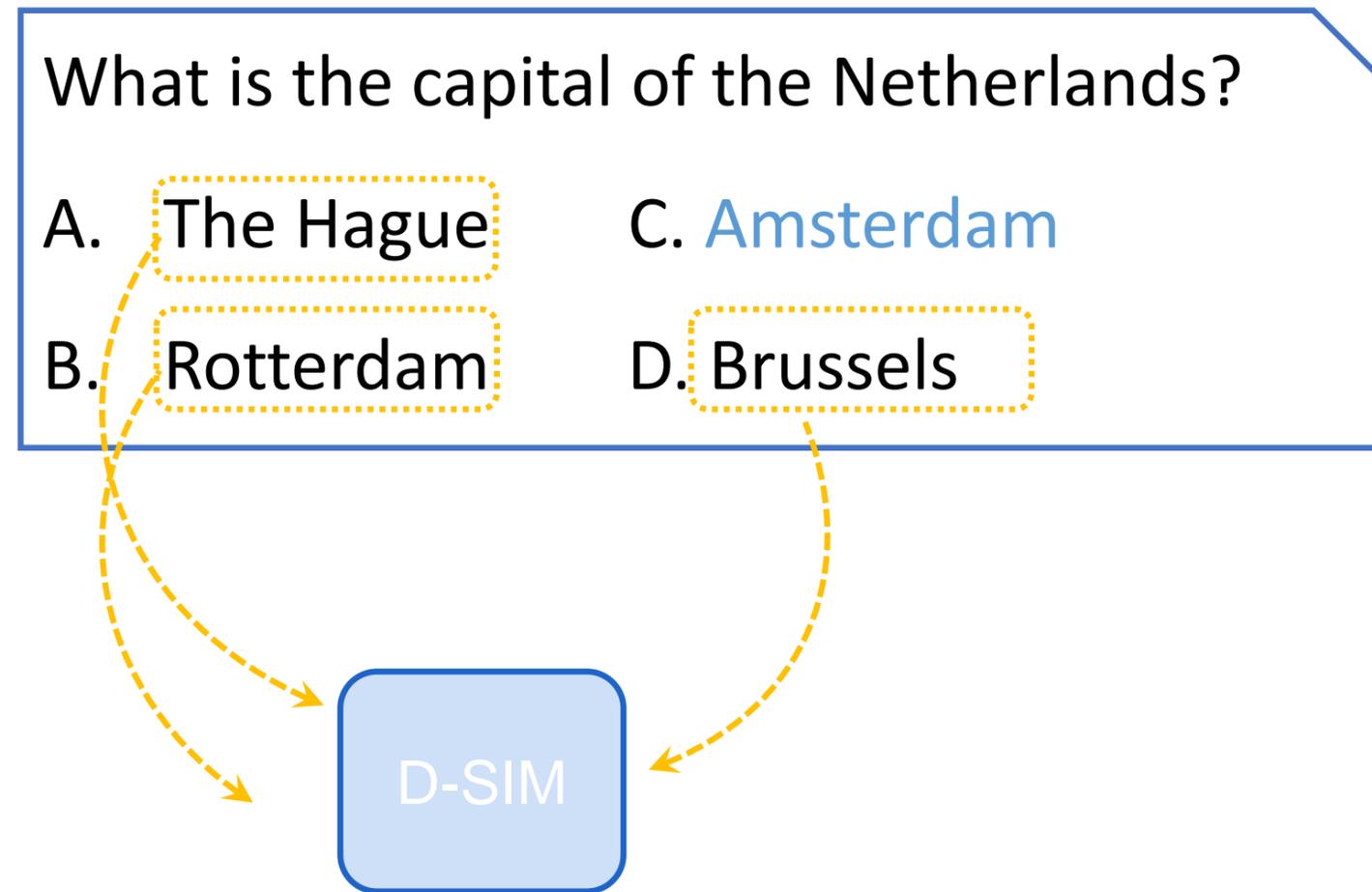
What is the capital of the Netherlands?

- A. The Hague C. Amsterdam
B. Rotterdam D. Brussels

Our approach (D-SIM)

2 Distractor Similarity (D-SIM) model

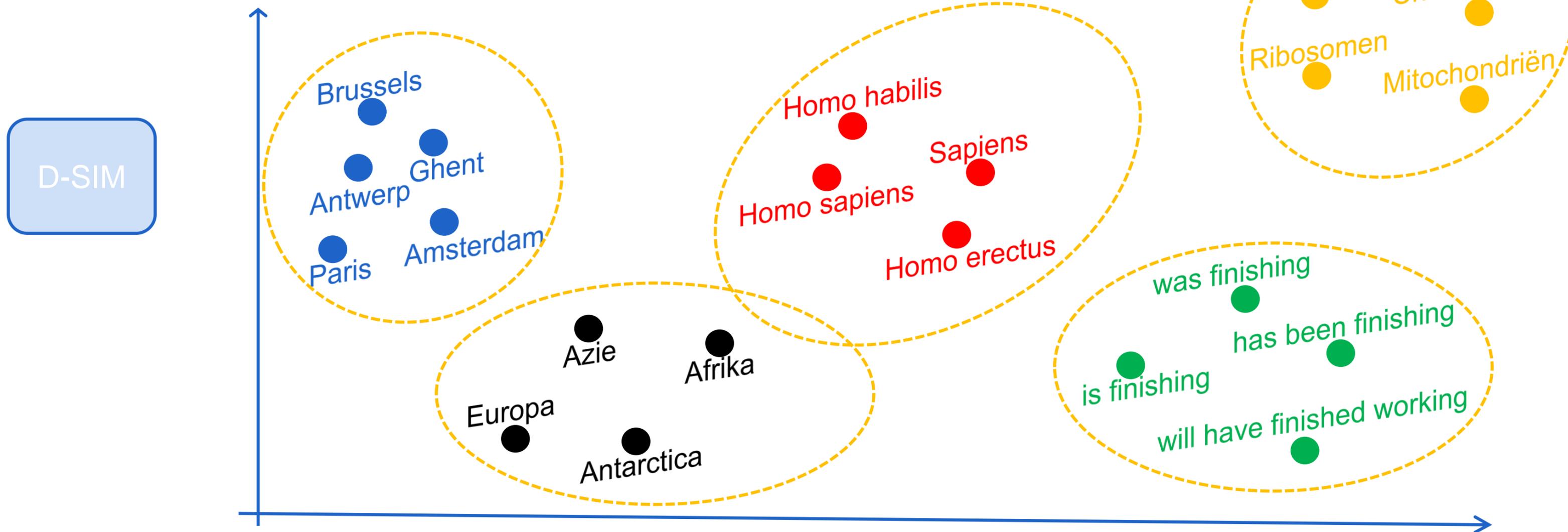
- Distractors are **similar** if they co-occur within the same MCQ



Our approach (D-SIM)

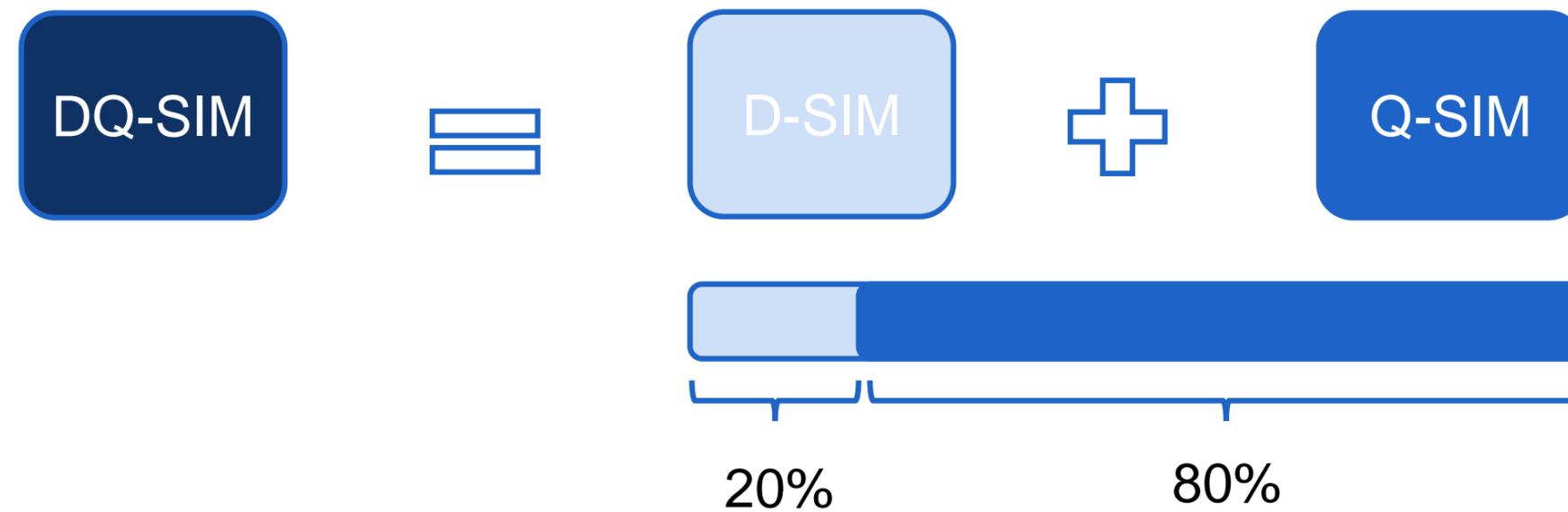
2 Distractor Similarity (D-SIM) model

- D-SIM clusters similar distractors together

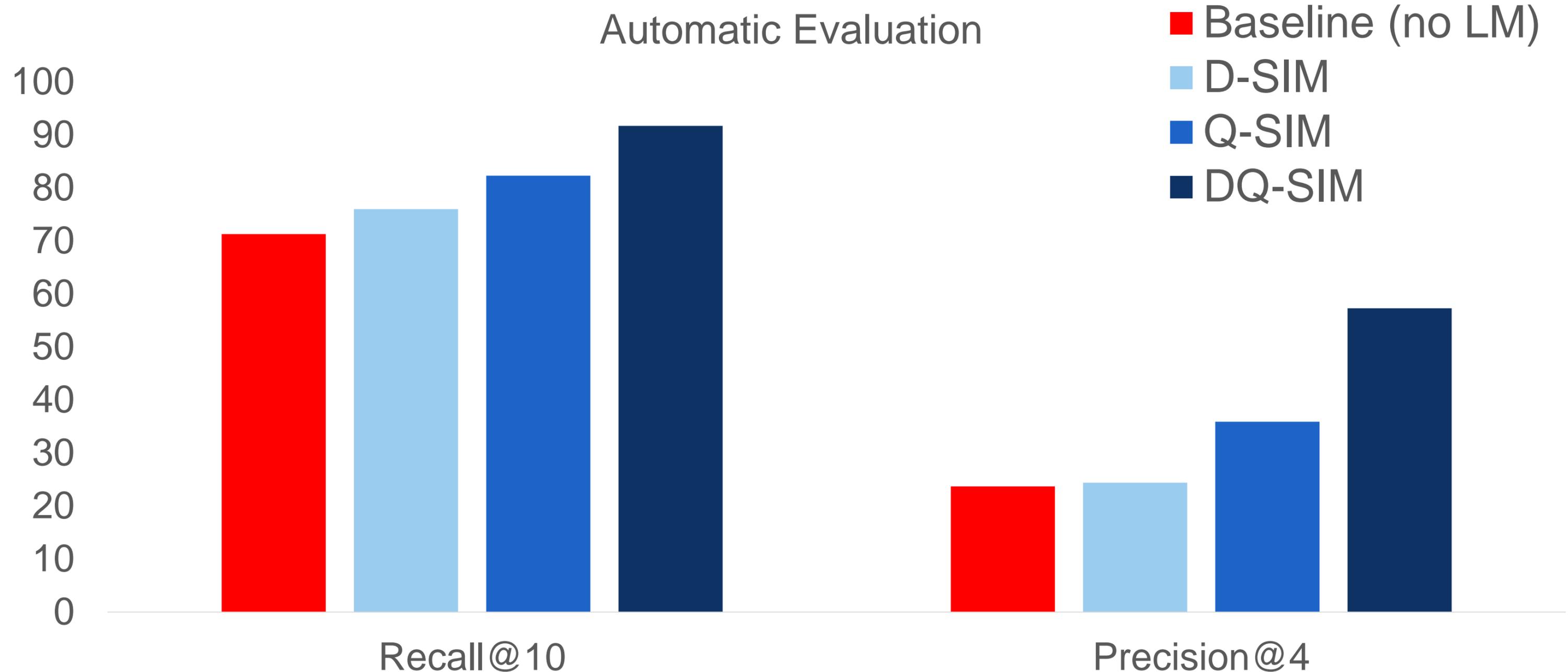


Our approach (DQ-SIM)

- 3 Distractor-Question Similarity (DQ-SIM) model
 - Combines Q-SIM and D-SIM

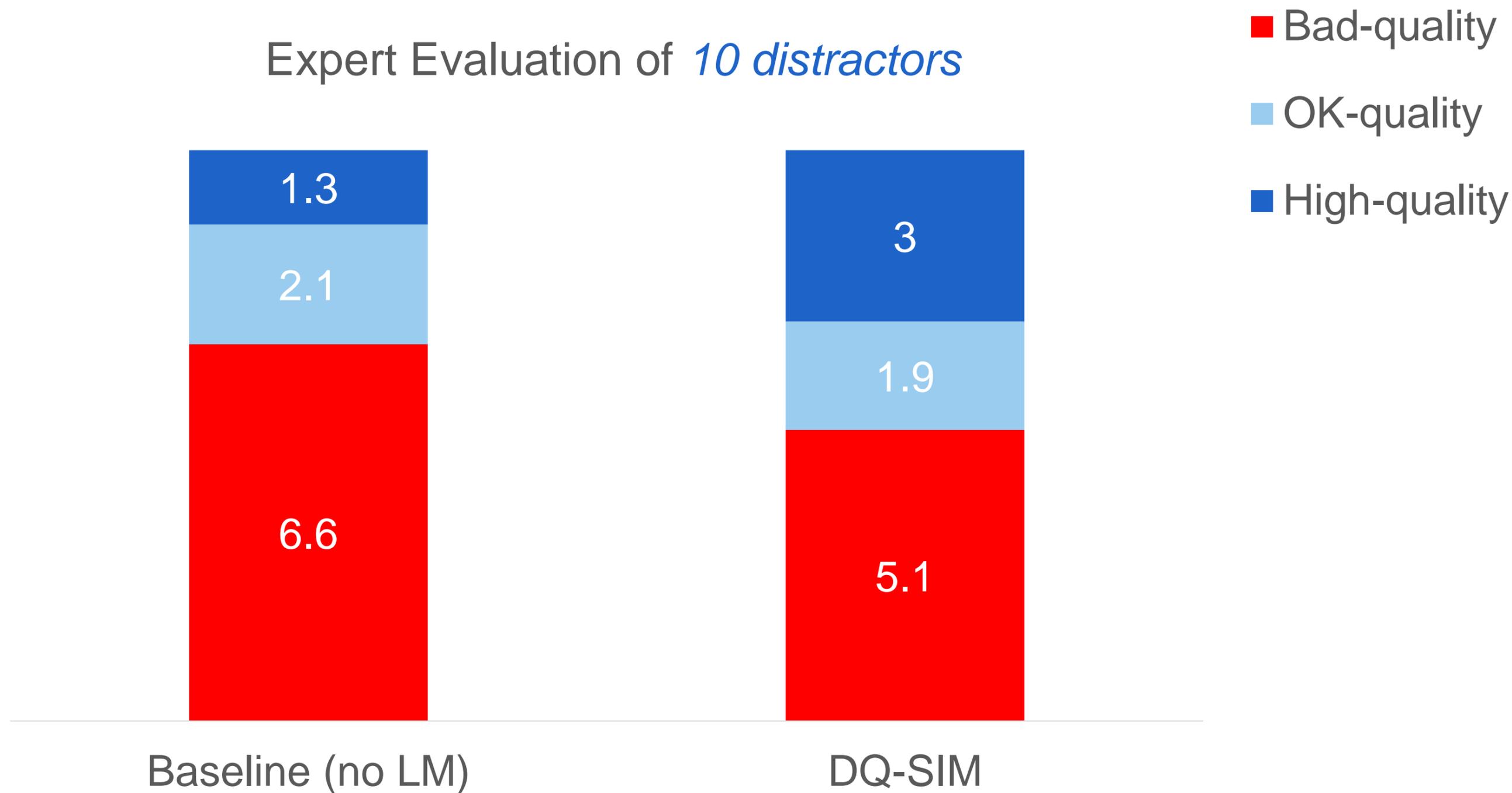


Gain in performance using language models



What do teachers say about our models?

Expert Evaluation of *10 distractors*



What do teachers say about our models?

Expert Evaluation of 10 distractors

3 distractors high-quality out of 10

- Bad-quality
- OK-quality
- High-quality



Baseline (no LM)

DQ-SIM



Our model in Televic's platform

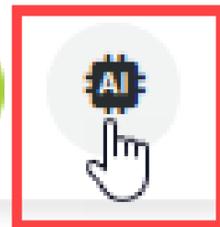
Wat is de meest gebruikte pijnstiller in de Belgische ziekenhuizen?

Selecteer het juiste antwoord.



Paracetamol

Keuze



Our model in Televic's platform

The screenshot displays the Televic platform interface. At the top, the breadcrumb navigation shows the path: Items / FVB / AIDA / Pijnstillers. The current view is 'Voorbeeld' (Preview) for an 'ITEM' type, specifically a 'KEUZE' (Choice) question. The question text is: "Wat is de meest gebruikte pijnstiller in de Beligse z...". Below the question, a text input field contains the answer "Paracetamol". A modal window titled "Afleiders voor 'Paracetamol'" is open, listing several options with checkboxes and a "Gemiddeld" (Average) rating for each. The selected options are aspirine, Ibuprofen, and Morfine. The modal also shows "3 afleiders geselecteerd" (3 derivatives selected) and buttons for "Annuleren" (Cancel) and "Toevoegen" (Add). The background interface includes a toolbar with "Blok toevoegen" (Add block), "1 versie" (1 version), and "0 commentaren" (0 comments). On the right, there are sections for "Geselecteerde keuze" (Selected choice) and "Feedback" (Feedback) for both "Indien juist" (If correct) and "Indien fout" (If incorrect) scenarios, each with a text input field for feedback.

Items / FVB / AIDA / Pijnstillers

Voorbeeld

ITEM BLOK KEUZE

Blok toevoegen 1 versie 0 commentaren

Wat is de meest gebruikte pijnstiller in de Beligse z...
Selecteer het juiste antwoord.

Paracetamol

Keuze

Feedback

- Algemene feedback
- Positieve feedback
- Negatieve feedback

Afleiders voor 'Paracetamol'

<input checked="" type="checkbox"/>	aspirine	Gemiddeld
<input type="checkbox"/>	Apirine	Gemiddeld
<input type="checkbox"/>	Acetylsali zuur	Gemiddeld
<input type="checkbox"/>	asperine	Gemiddeld
<input type="checkbox"/>	Asprinie	Gemiddeld
<input type="checkbox"/>	NSAID	Gemiddeld
<input checked="" type="checkbox"/>	Ibuprofen	Gemiddeld
<input checked="" type="checkbox"/>	Morfine	Gemiddeld

3 afleiders geselecteerd Annuleren Toevoegen

Geselecteerde keuze

Indien juist

Feedback

Typ hier uw inhoud

Indien fout

Feedback

Typ hier uw inhoud

min. 0 max. 1

Keuze 1 / 1

Our model in Televic's platform

The screenshot displays the Televic platform interface for editing a quiz item. The breadcrumb path is "/ Items / FVB / AIDA / Pijnstillers". The current item is a "Keuze" (Choice) type, labeled "Voorbeeld". The question is "Wat is de meest gebruikte pijnstiller in de Belgische ziekenhuizen?" (What is the most commonly used painkiller in Belgian hospitals?). The instruction is "Selecteer het juiste antwoord." (Select the correct answer.). The question is displayed in a rich text editor with a toolbar. Below the question, there are four radio button options: Paracetamol (selected), Aspirine, Morfine, and iboprufen. A green "Keuze" button is visible at the bottom left of the question area, and "min. 0" and "max. 1" are shown at the bottom right. On the right side, a notification box states "Afleidings toegevoegd" (Derivatives added) with a green checkmark, indicating that 3 derivatives were successfully added. Below the notification, there is a section for "Media" with the text "Dit blok heeft nog geen media" (This block has no media yet) and a "Opmaak" (Format) section.

Items / FVB / AIDA / Pijnstillers

Voorbeeld

ITEM BLOK KEUZE

+ Blok toevoegen 1 versie 0 commentaren

Wat is de meest gebruikte pijnstiller in de Belgische ziekenhuizen?
Selecteer het juiste antwoord.

B I U x_2 x^2 f_x

Paracetamol
 Aspirine
 Morfine
 iboprufen

+ Keuze min. 0 max. 1

Feedback

Algemene feedback

Positieve feedback

Negatieve feedback

Afleidings toegevoegd
3 afleidings zijn met succes toegevoegd

Dit blok heeft nog geen media

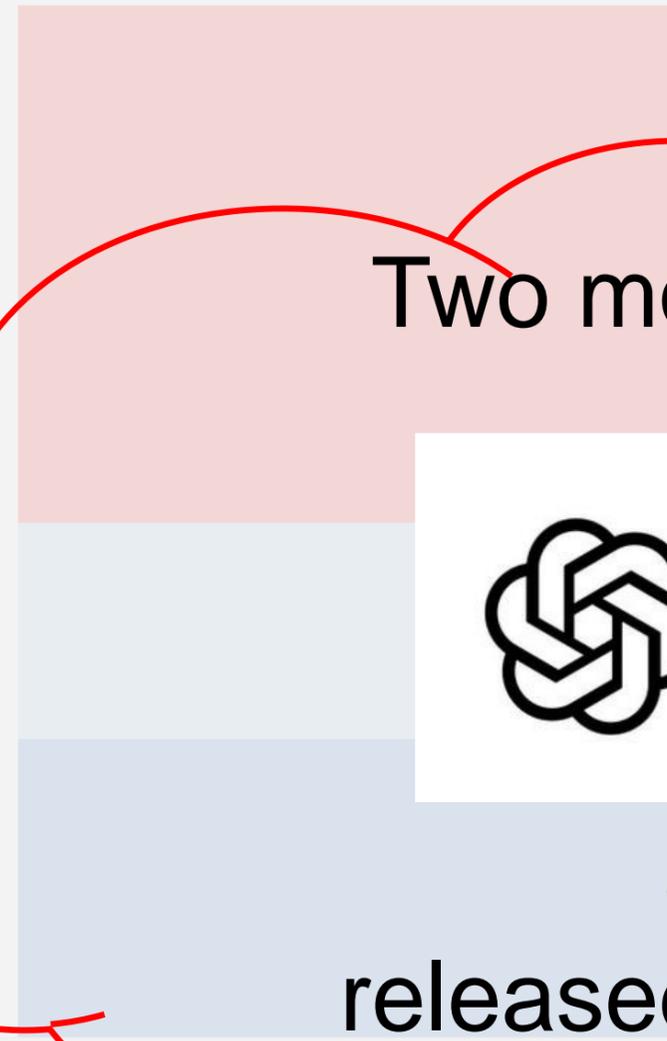
> Opmaak

Users say about our models?

Expert Evaluation



80%
70%
60%
50%
40%
30%
20%
10%
0%



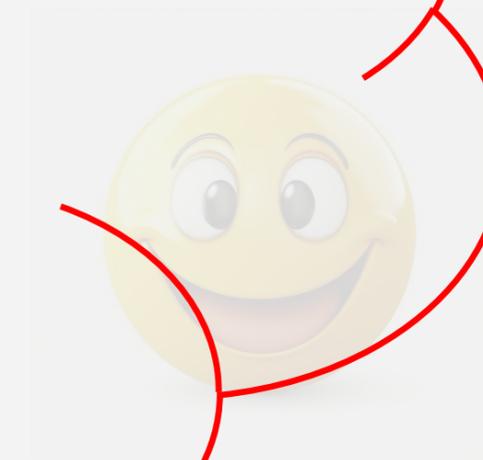
Two months later



released ChatGPT

3 distractors high-quality out of 10

televisic



■ High-quality ■ OK-quality ■ Bad-quality

Distractor Generation

Using Generative language models

Research questions

1. How to automatically generate distractors?

2. How to design a system that generates **free-form** distractors?

Free-form Distractor Generation

- **Recap:** the previous methods **CANNOT** generate brand new distractors

Free-form Distractor Generation

- **Recap:** the previous methods **CANNOT** generate brand new distractors
- **Solution:** Ask LLMs such as ChatGPT to generate distractors (***Zero-ChatGPT***)

Free-form Distractor Generation

- **Recap:** the previous methods **CANNOT** generate brand new distractors
- **Solution:** Ask LLMs such as ChatGPT to generate distractors (***Zero-ChatGPT***)

Original Question and Correct Answer:

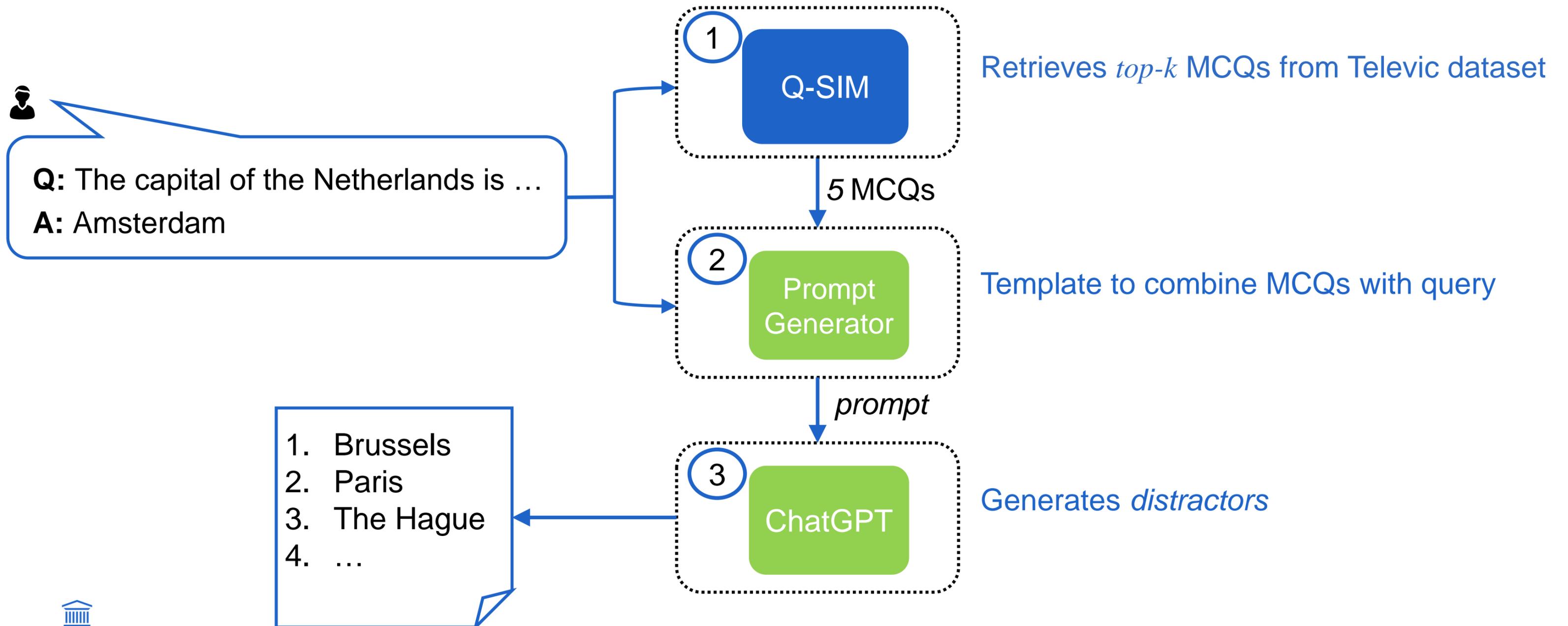
The capital of the Netherlands is ...
Amsterdam

Zero-shot prompt input:

Generate 10 plausible but incorrect answers for the following question.
Question: The capital of the Netherlands is ...
Answer: Amsterdam

Our approach (Dynamic-ChatGPT)

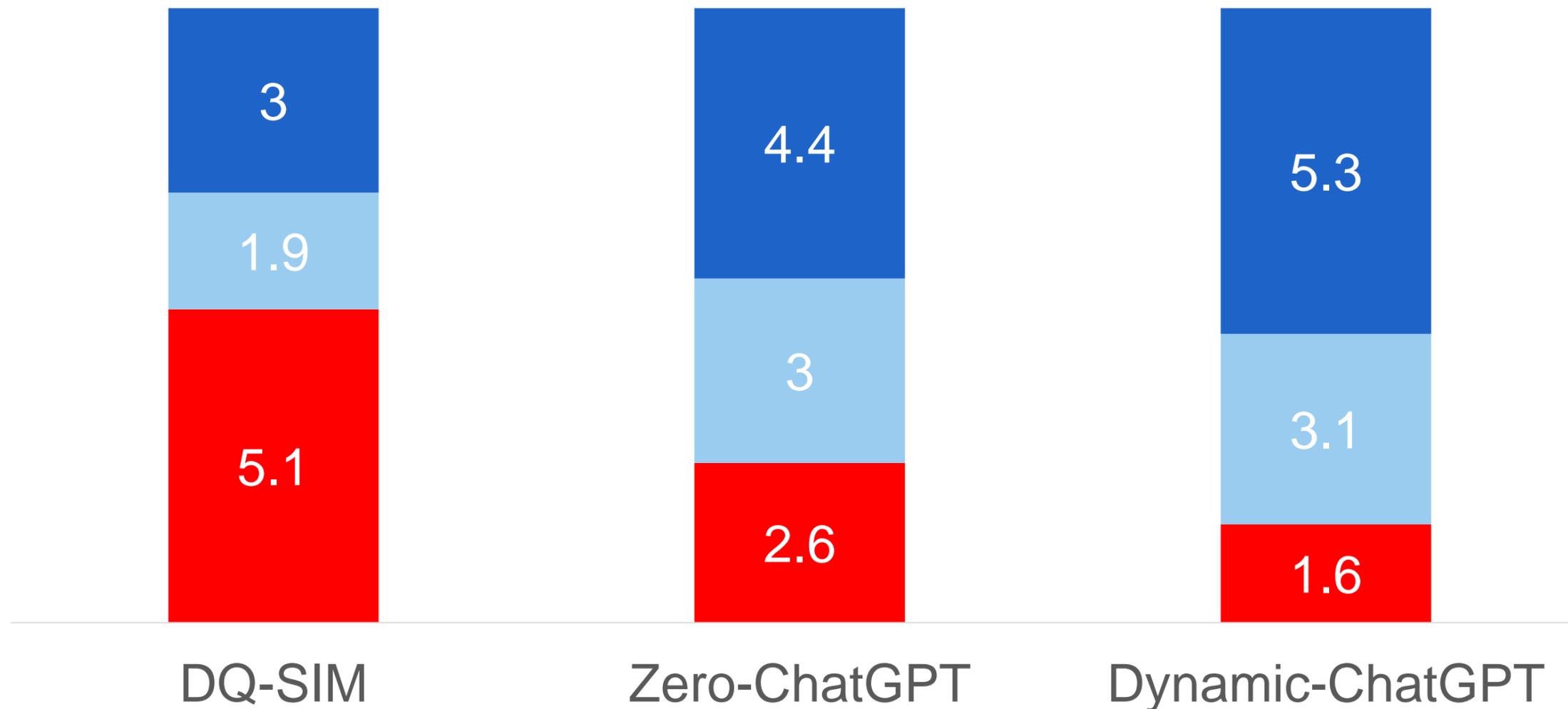
- We proposed a pipeline approach with three components



Performance gain on quality of generated distractors

Expert Evaluation of *10 distractors*

- Bad-quality
- OK-quality
- High-quality



Gap-filling Grammar

Exercise Generation

Research questions

1. How to automatically generate distractors?

2. How to design a system that generates free-form distractors?

3. How to adapt a language model to **generate gap-fill** grammar exercises?

Gap-filling Grammar Exercise Generation

Gap-filling grammar exercise important in **formative** assessment

Gap-filling Grammar Exercise Generation

new tense



Gap-filling Grammar Exercise Generation

new tense



John arrived from the airport yesterday at 8:00. He checked into the hotel at 9:00, and met his colleagues at 10:00. He will present his work tomorrow.

Gap-filling Grammar Exercise Generation

new tense



Past tense

John arrived from the airport yesterday at 8:00. He checked into the hotel at 9:00, and met his colleagues at 10:00. He will present his work tomorrow.

Gap-filling Grammar Exercise Generation

new tense



Past tense

John arrived from the airport yesterday at 8:00. He checked into the hotel at 9:00, and met his colleagues at 10:00. He will present his work tomorrow.

Input

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.

Gap-filling Grammar Exercise Generation

new tense



Past tense

John arrived from the airport yesterday at 8:00. He checked into the hotel at 9:00, and met his colleagues at 10:00. He will present his work tomorrow.

Input

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.

LM-based
Exercise
Generator

Output

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.

Gap-filling Grammar Exercise Generation

new tense



Future tense

John arrived from the airport yesterday at 8:00. He checked into the hotel at 9:00, and met his colleagues at 10:00. He will present his work tomorrow.



Input

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.



LM-based
Exercise
Generator

Gap-filling Grammar Exercise Generation

new tense



Future tense

John arrived from the airport yesterday at 8:00. He checked into the hotel at 9:00, and met his colleagues at 10:00. He will present his work tomorrow.



Input

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.

LM-based
Exercise
Generator

Output

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.

Gap-filling Grammar Exercise Generation

combine tenses



Past + Future

John arrived from the airport yesterday at 8:00. He checked into the hotel at 9:00, and met his colleagues at 10:00. He will present his work tomorrow.



Input

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.



LM-based
Exercise
Generator

Gap-filling Grammar Exercise Generation

combine tenses



Past + Future

John arrived from the airport yesterday at 8:00. He checked into the hotel at 9:00, and met his colleagues at 10:00. He will present his work tomorrow.

Input

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.

LM-based
Exercise
Generator

Out put

Guardiola admitted there was a deliberate ploy in this game to use Debruyne to unlock Arsenal. It worked to devastating effect.

City will play Chelsea in next week.

Our approach

- 1 Given input text, extract spans (e.g. [size of 4](#))

Input

City will play Chelsea in next week.

Our approach

- 1 Given input text, extract spans (e.g. [size of 4](#))



City

The extracted span "City" is shown in a separate grey box below the input text.

Our approach

- 1 Given input text, extract spans (e.g. [size of 4](#))



City will

City

Our approach

- 1 Given input text, extract spans (e.g. [size of 4](#))

Input
City will play Chelsea in next week.

City will

City

City will play

Our approach

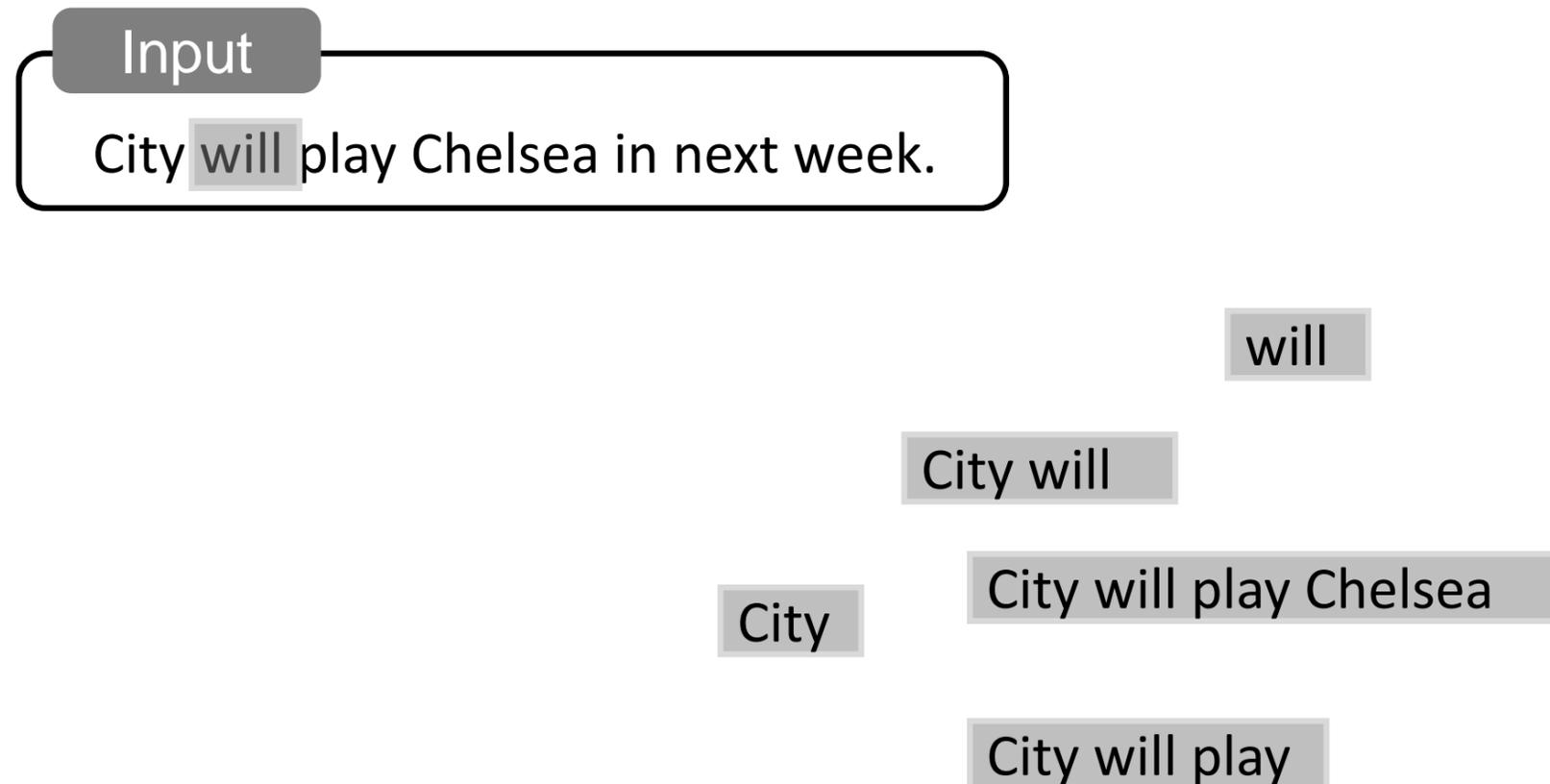
- 1 Given input text, extract spans (e.g. [size of 4](#))

Input
City will play Chelsea in next week.

City will
City City will play Chelsea
City will play

Our approach

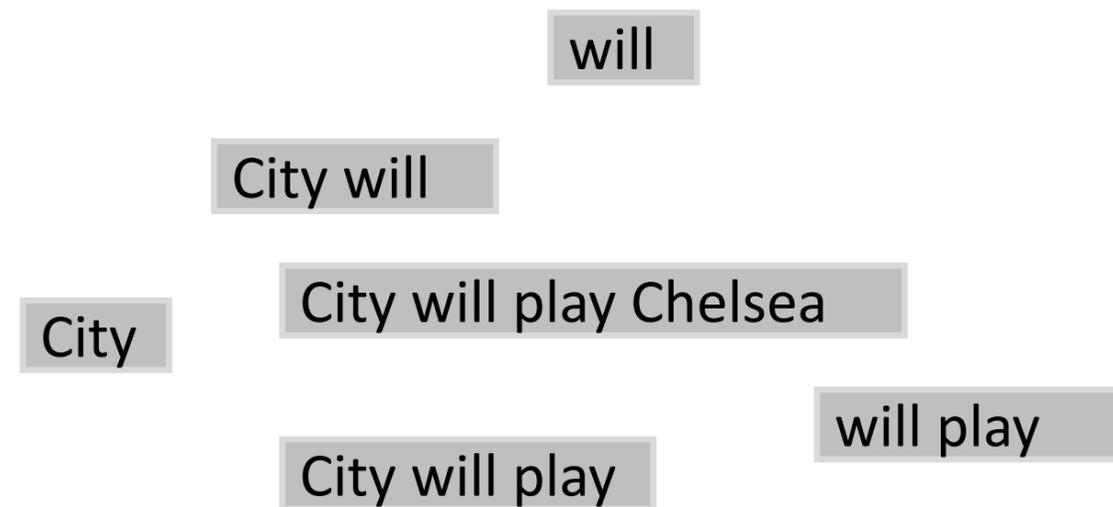
- 1 Given input text, extract spans (e.g. [size of 4](#))



Our approach

- 1 Given input text, extract spans (e.g. [size of 4](#))

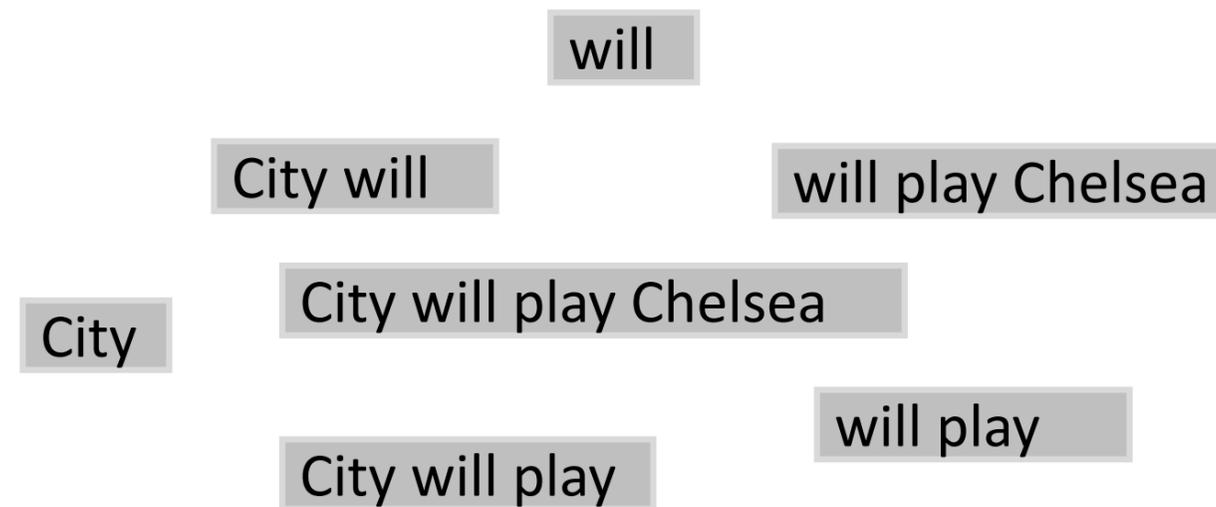
Input
City will play Chelsea in next week.



Our approach

- 1 Given input text, extract spans (e.g. [size of 4](#))

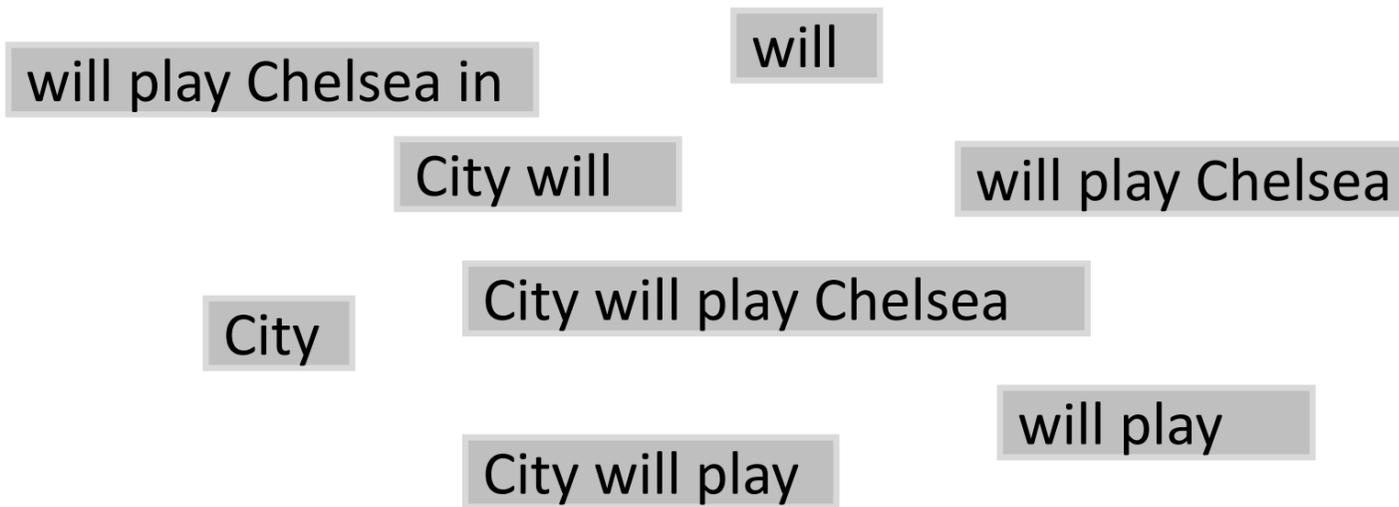
Input
City will play Chelsea in next week.



Our approach

- 1 Given input text, extract spans (e.g. [size of 4](#))

Input
City will play Chelsea in next week.



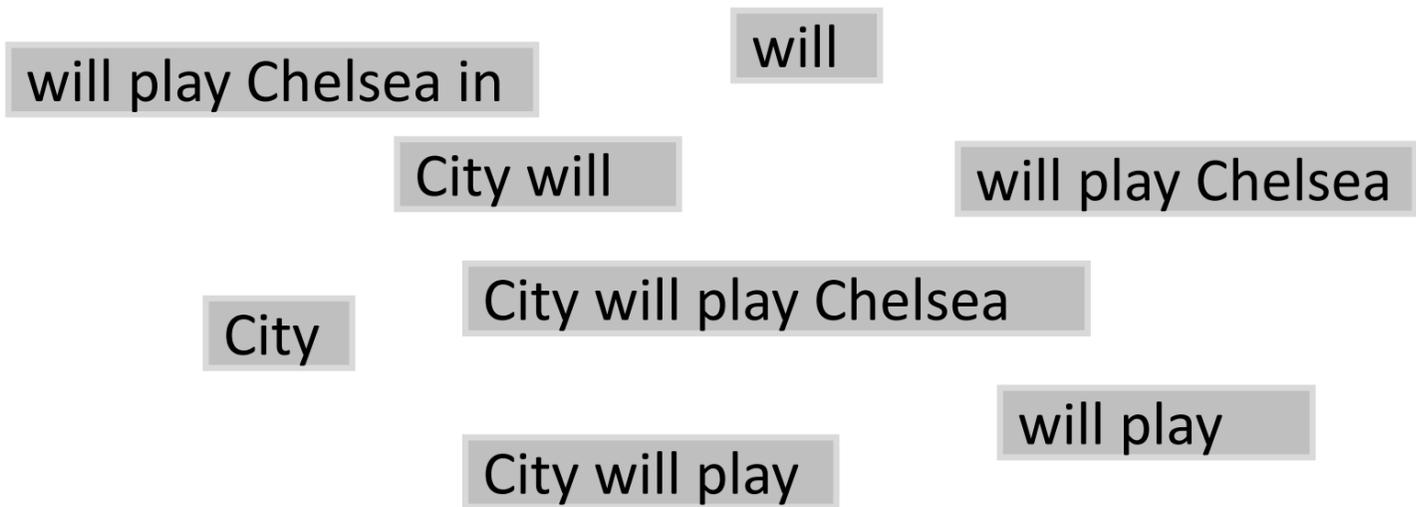
Our approach

1 Given input text, extract spans (e.g. *size of 4*)

2 Decide if each span is a suitable gap based on example

Input
City will play Chelsea in next week.

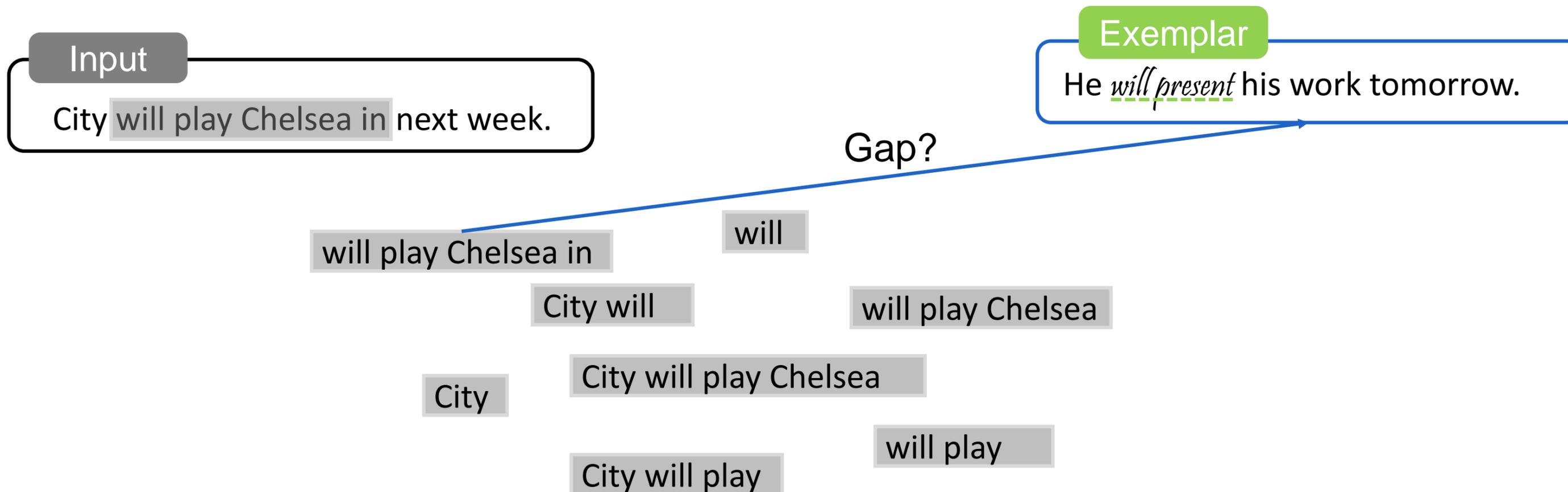
Exemplar
He will present his work tomorrow.



Our approach

1 Given input text, extract spans (e.g. *size of 4*)

2 Decide if each span is a suitable gap based on example



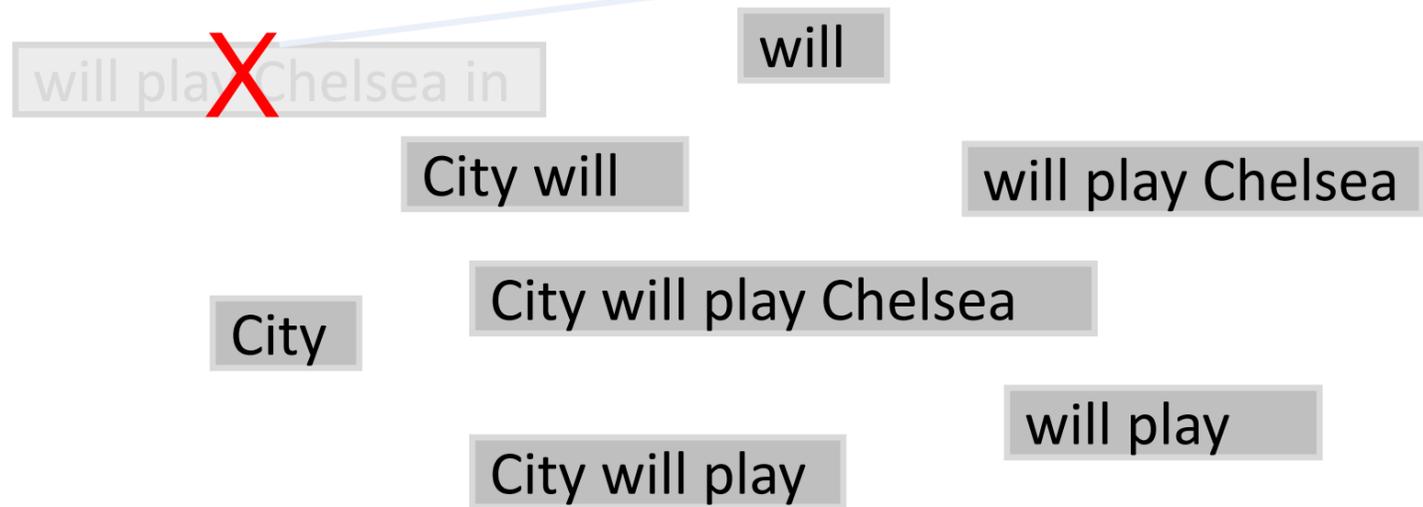
Our approach

1 Given input text, extract spans (e.g. *size of 4*)

2 Decide if each span is a suitable gap based on example

Input
City will play Chelsea in next week.

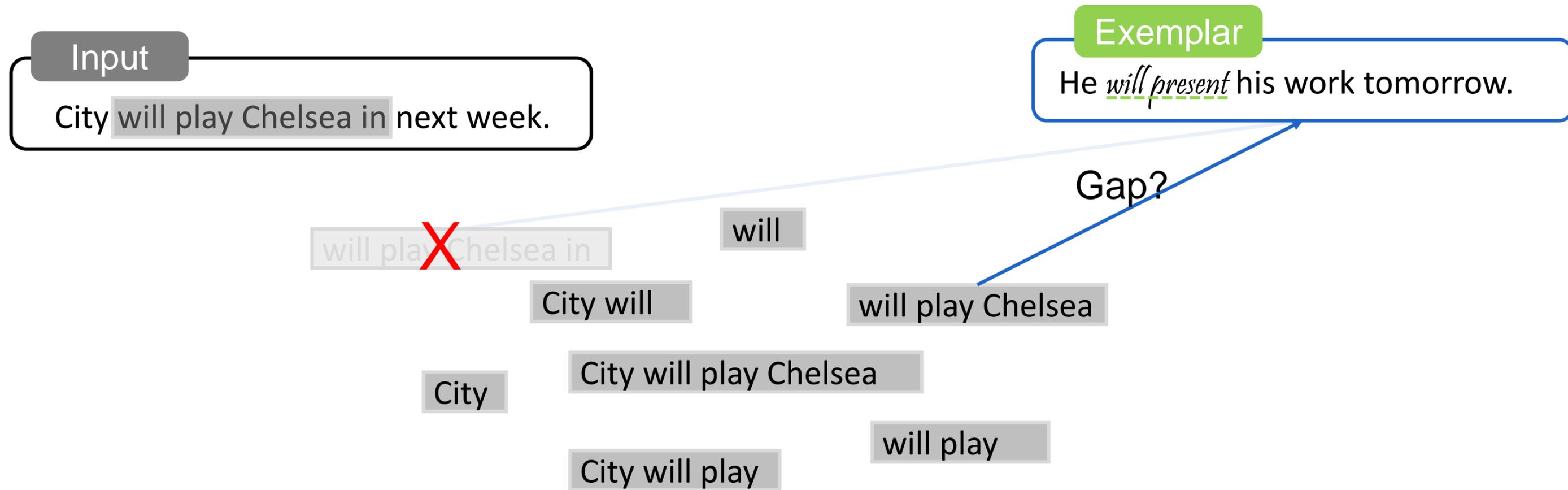
Exemplar
He will present his work tomorrow.



Our approach

1 Given input text, extract spans (e.g. size of 4)

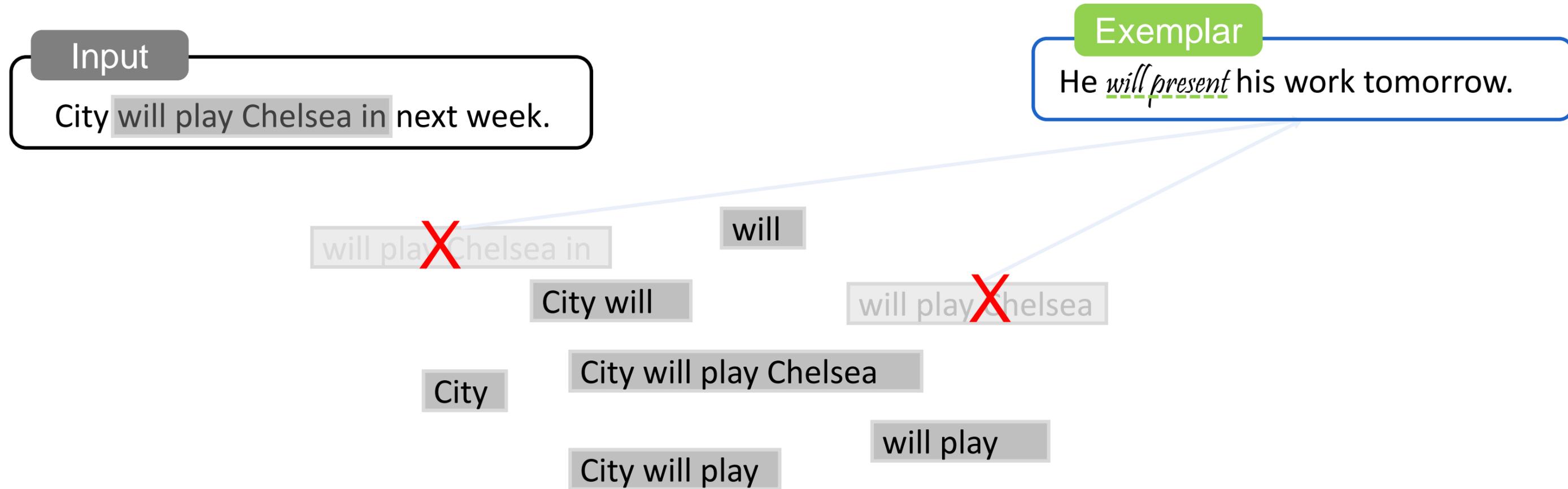
2 Decide if each span is a suitable gap based on example



Our approach

1 Given input text, extract spans (e.g. size of 4)

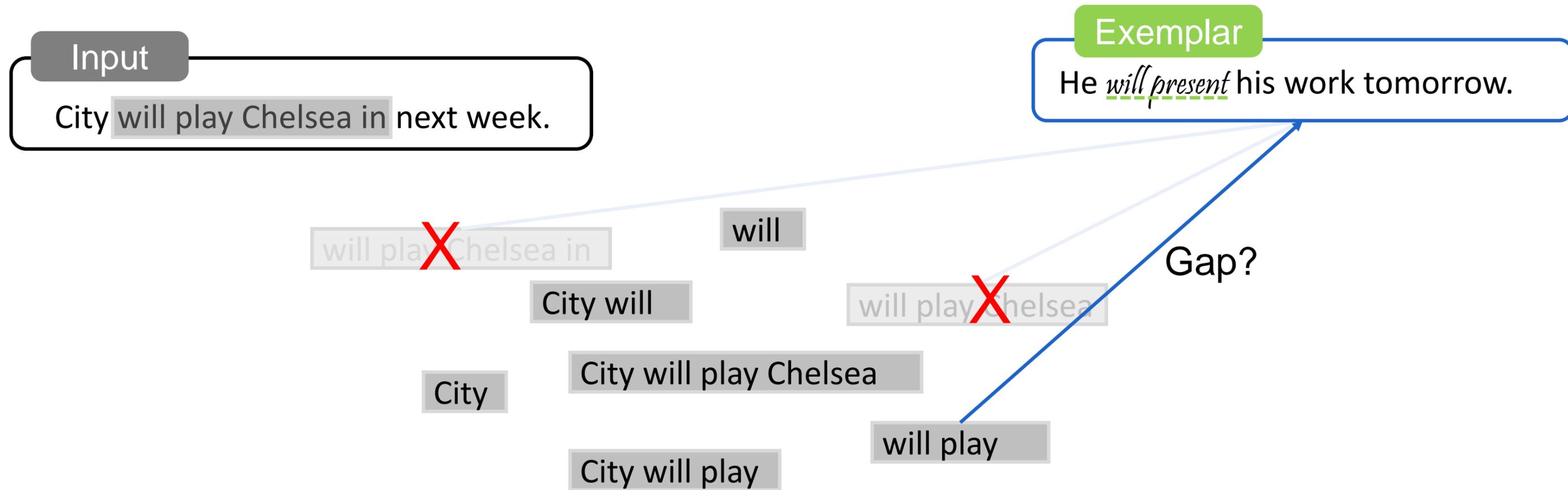
2 Decide if each span is a suitable gap based on example



Our approach

1 Given input text, extract spans (e.g. size of 4)

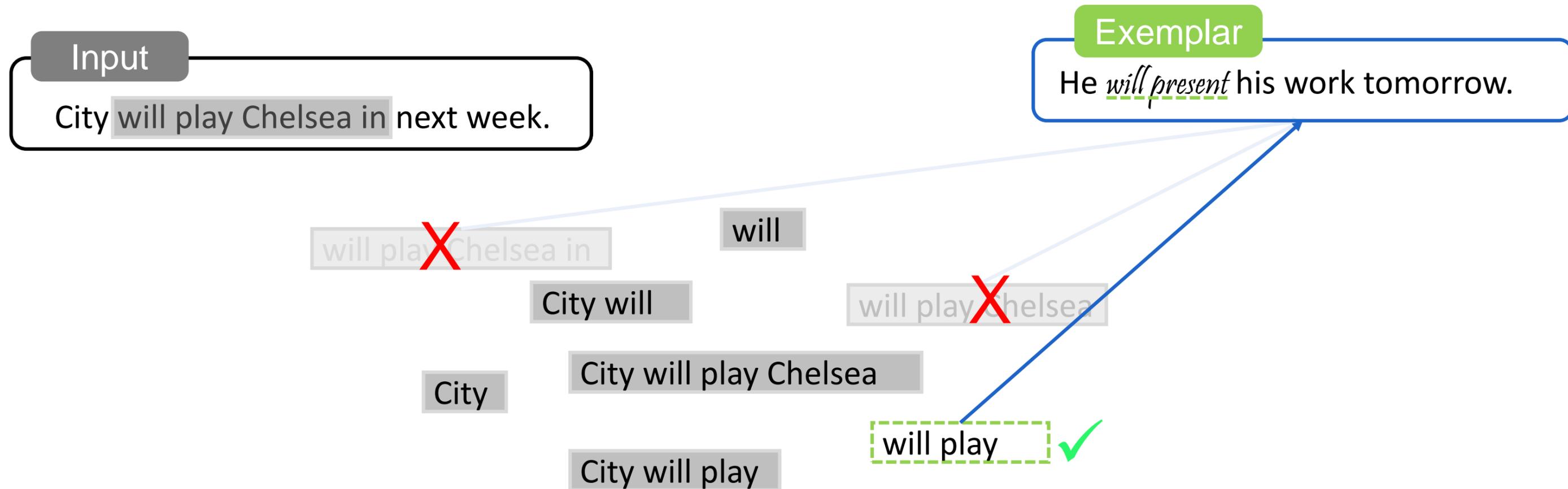
2 Decide if each span is a suitable gap based on example



Our approach

1 Given input text, extract spans (e.g. size of 4)

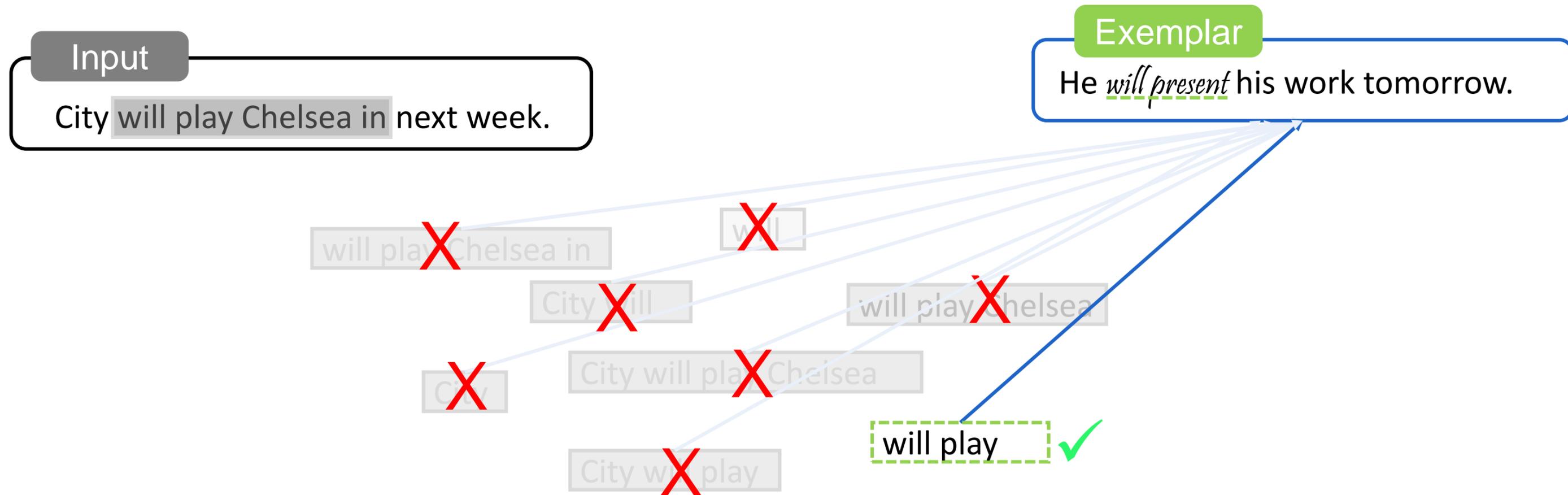
2 Decide if each span is a suitable gap based on example



Our approach

1 Given input text, extract spans (e.g. size of 4)

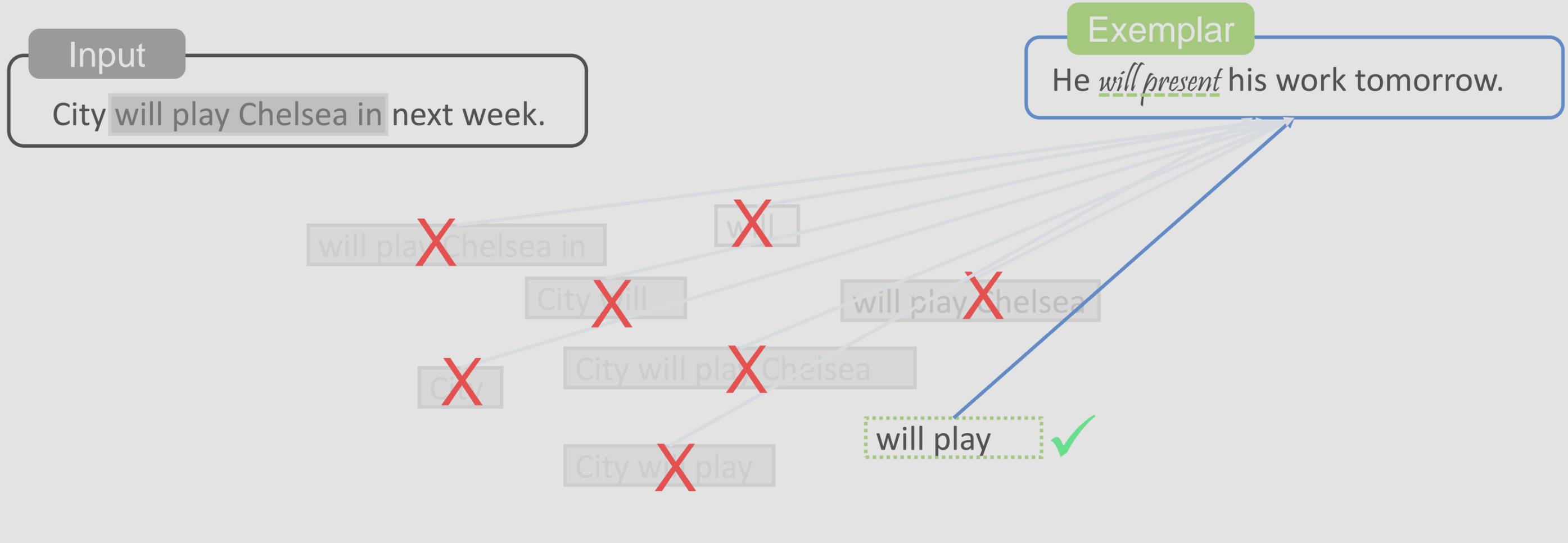
2 Decide if each span is a suitable gap based on example



Our approach

1 Given input text, extract spans (e.g. *size of 4*)

2 Decide if each span is a suitable gap based on example



New Exercise

City will play Chelsea in next week.

GF2 Grammar Dataset

A real-world corpus of existing exercises on mixed grammar topics

- ~ 800 exercise documents
- French grammar exercises

Example 1

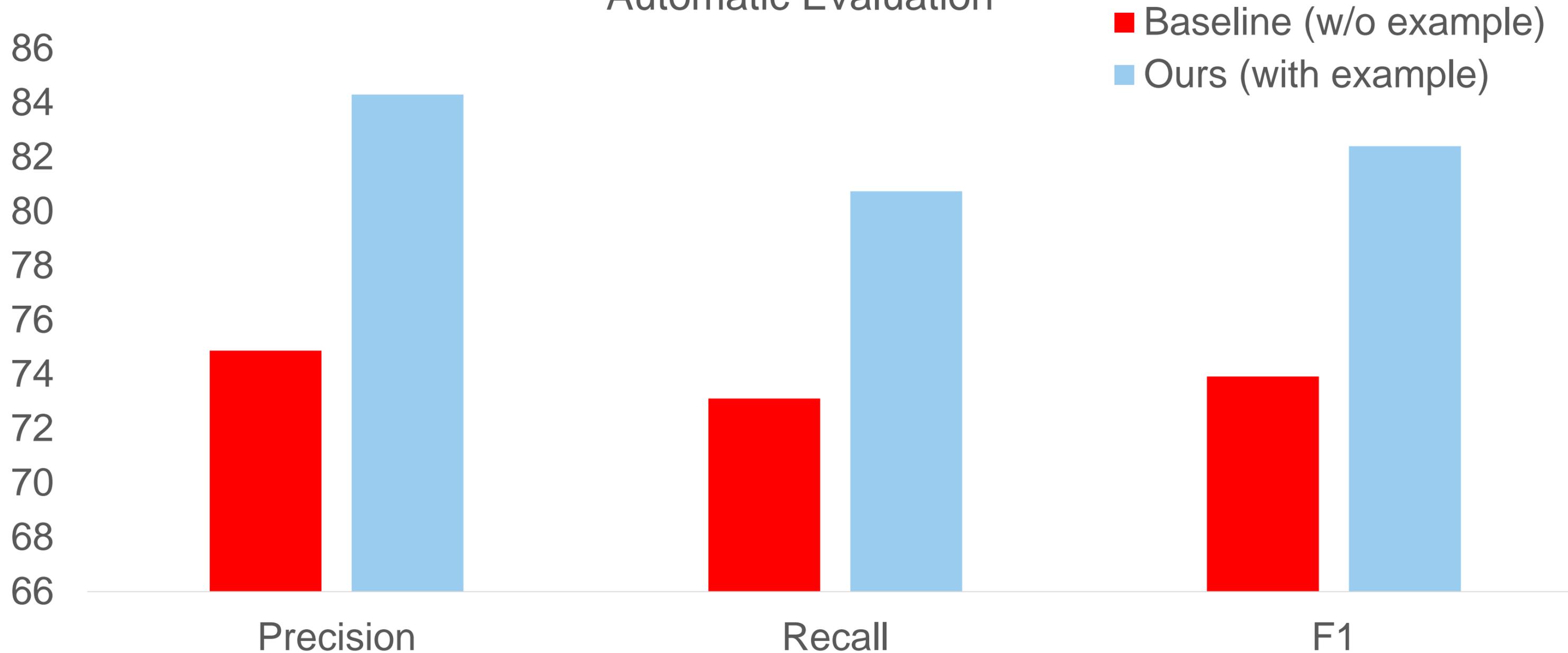
- 1 Vous travaillerez beaucoup?
1 Will you work a lot?
- 2 En ne mangeant plus de bonbons,
tu maigriras vite!
2 By not eating sweets, you will lose
weight quickly!
- 3 J'espère que mon équipe favorite
ne perdra plus aucun match.
3 I hope my favorite team won't lose any
more games.
- 4 Maxime m'a promis, qu'il ne
mentira plus jamais.
4 Maxime promised me that he will never
lie again.
- 5 Maman préparera des spaghettis ce soir.
5 Mum will make spaghetti tonight.

Example 2

A l'âge de 27 ans, le Californien David Blancarte
At the age of 27, Californian David Blancarte had
a eu un grave accident de scooter. Quand il s'est
a serious scooter accident. When he woke up
réveillé à l'hôpital, il ne sentait plus ses
in the hospital, he no longer felt his
jambes. On lui a expliqué qu'il ne pourrait
legs. It was explained to him that he couldn't
plus marcher. C' était une vraie catastrophe
walk anymore. It was a real disaster for him!
pour lui! Pendant une longue période de
During a long period of rehabilitation, he learned
revalidation, il a appris à se déplacer en chaise.
to move around in a wheelchair.
roulante. ...

Gain in performance using our approach

Automatic Evaluation



Take-away Messages

Burden on educators

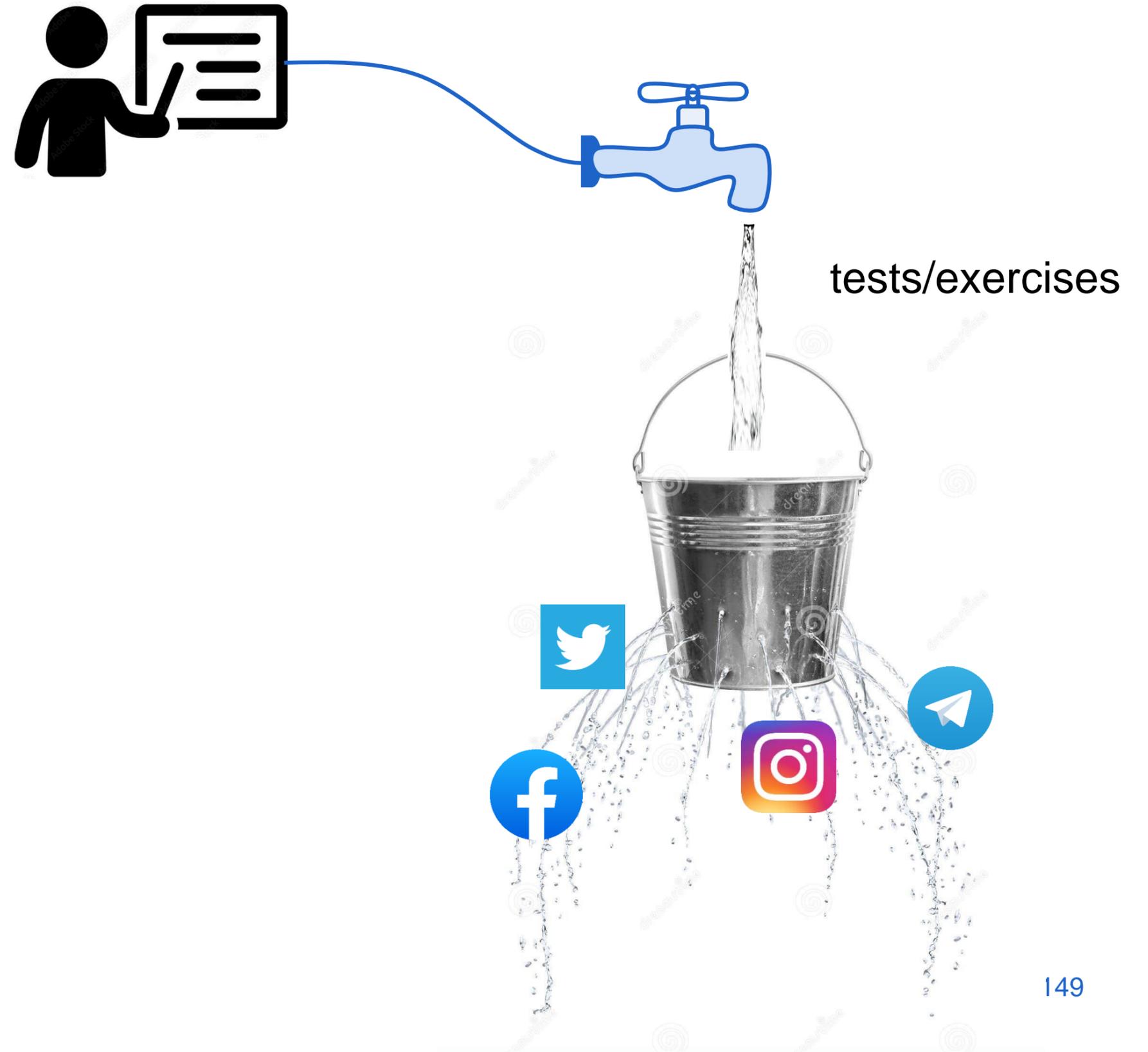


Burden on educators

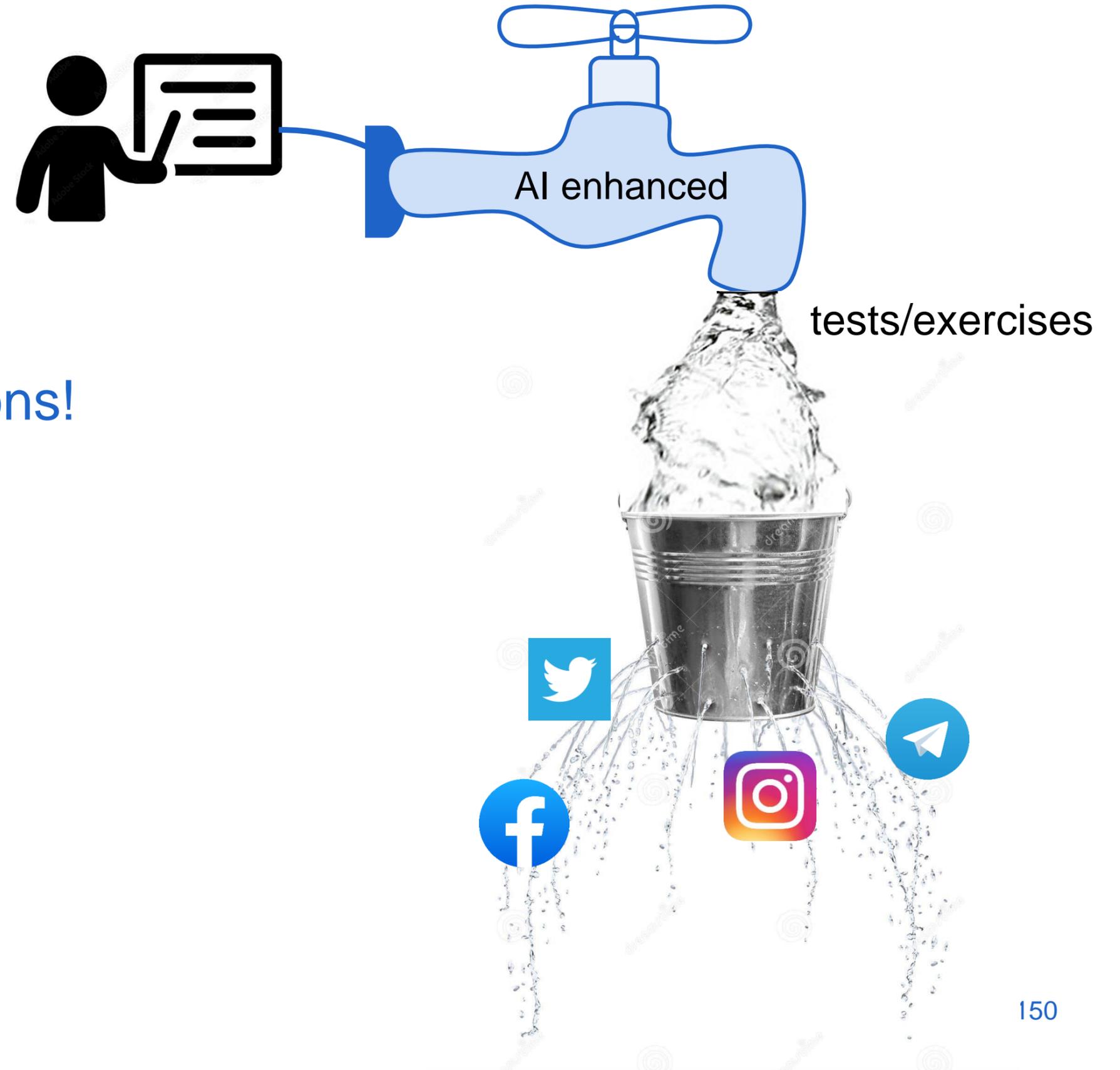
Tiny bit less burden!



Single-use dilemma



Single-use dilemma



Distractor Generation

- Smartly combining locally fine-tuned models with LLMs such as ChatGPT improves distractor generation

Distractor Generation

- Smartly combining locally fine-tuned models with LLMs such as ChatGPT improves distractor generation
- Teacher evaluations: critical for assessing quality of AI educational tools

Gap-filling exercise generation

- Our example-aware gap-filling exercise generator better than baseline that doesn't integrate examples

Gap-filling exercise generation

- Our example-aware gap-filling exercise generator better than baseline that doesn't integrate examples
- Facilitates targeted **formative** assessment
 - Teachers can use it for introducing new grammar exercises
 - Students can practice on grammar exercises they want to master

Ethical and practical consideration

- Deployment of AI tools in education:
 - Rigorous critical evaluation of AI models and human supervision
 - Emphasis on privacy of students and educators!

Ethical and practical consideration

- Deployment of AI tools in education:
 - Rigorous critical evaluation of AI models and human supervision
 - Emphasis on privacy of students and educators!
- Address data bias in LLMs

Ethical and practical consideration

- Deployment of AI tools in education:
 - Rigorous critical evaluation of AI models and human supervision
 - Emphasis on privacy of students and educators!
- Address data bias in LLMs
- Consider environmental impact during the development of these tools!

Thanks!



Let's do a quick poll: Personal Trivia

What is my **last** name?

A

Kiros

C

Semere

B

Bitew

D

Kiros Bitew

Let's do a quick poll: Personal Trivia

What is my **last** name?

A

BMW

C

Gent

B

Bitew

D

Kiros Bitew