

# Graph Learning for Natural Language Understanding

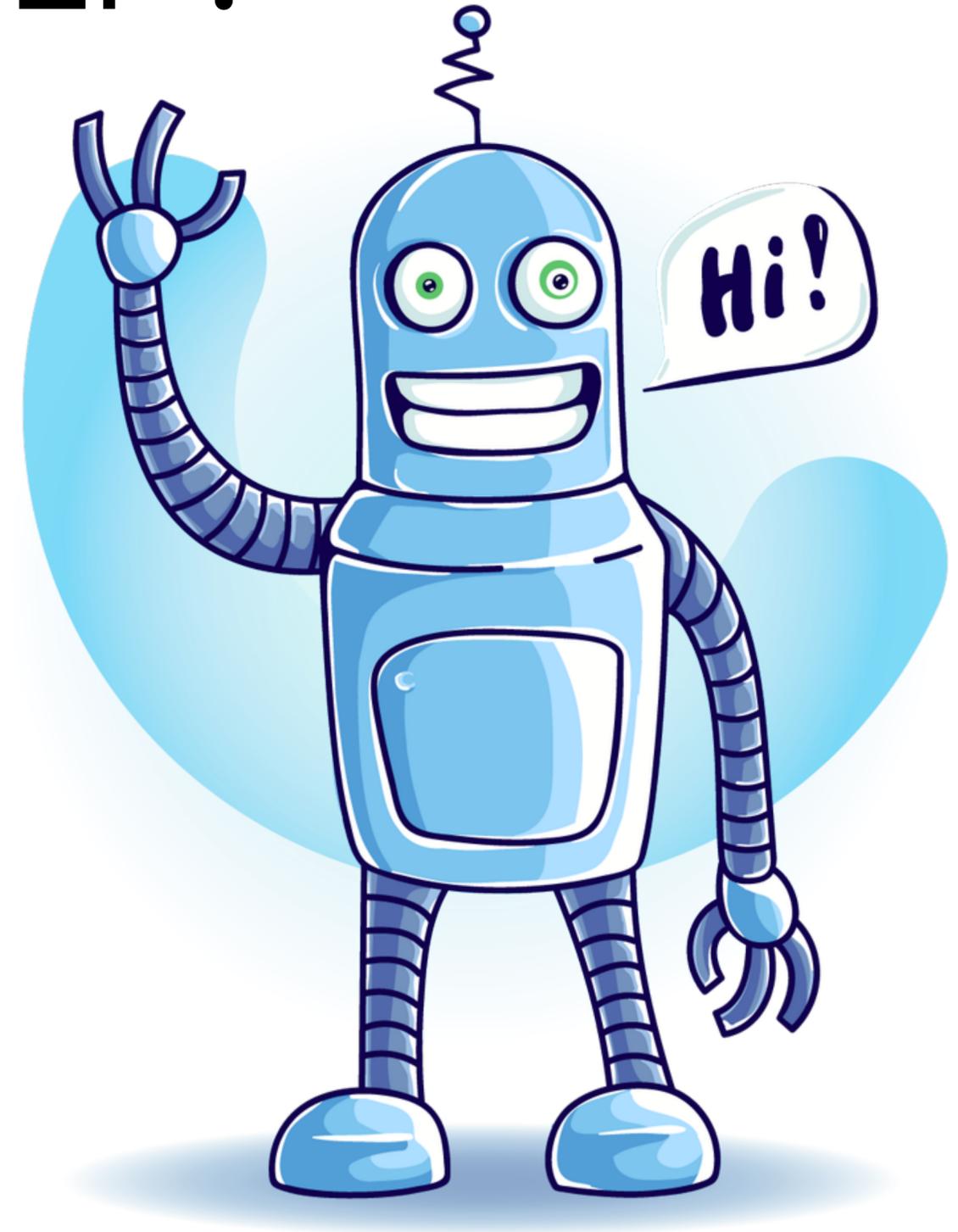
Antoine Bosselut



**EPFL**

# What is NLP?

- Automated understanding of natural language input
- Coherent generation of natural language output

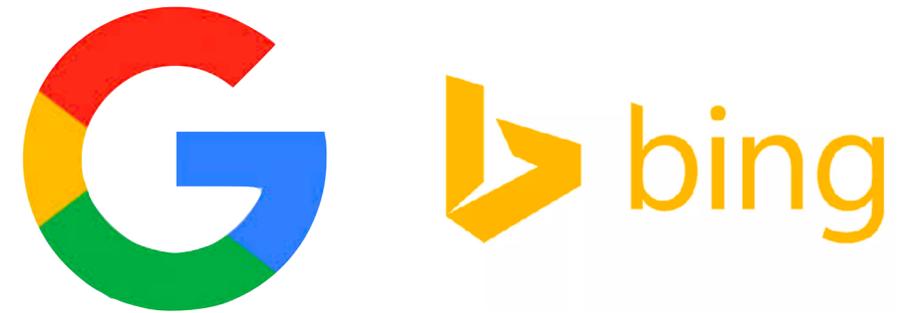


# NLP Applications

Machine Translation:



Question Answering:



Personal Assistants:



Specialized Applications:



Legal Documents

Health Records

Business Intelligence

Customer Research

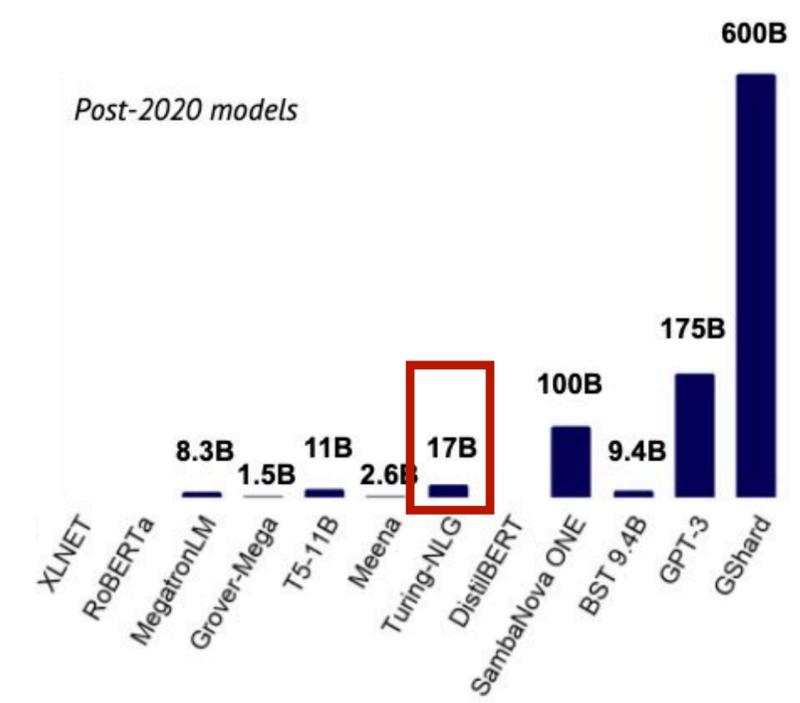
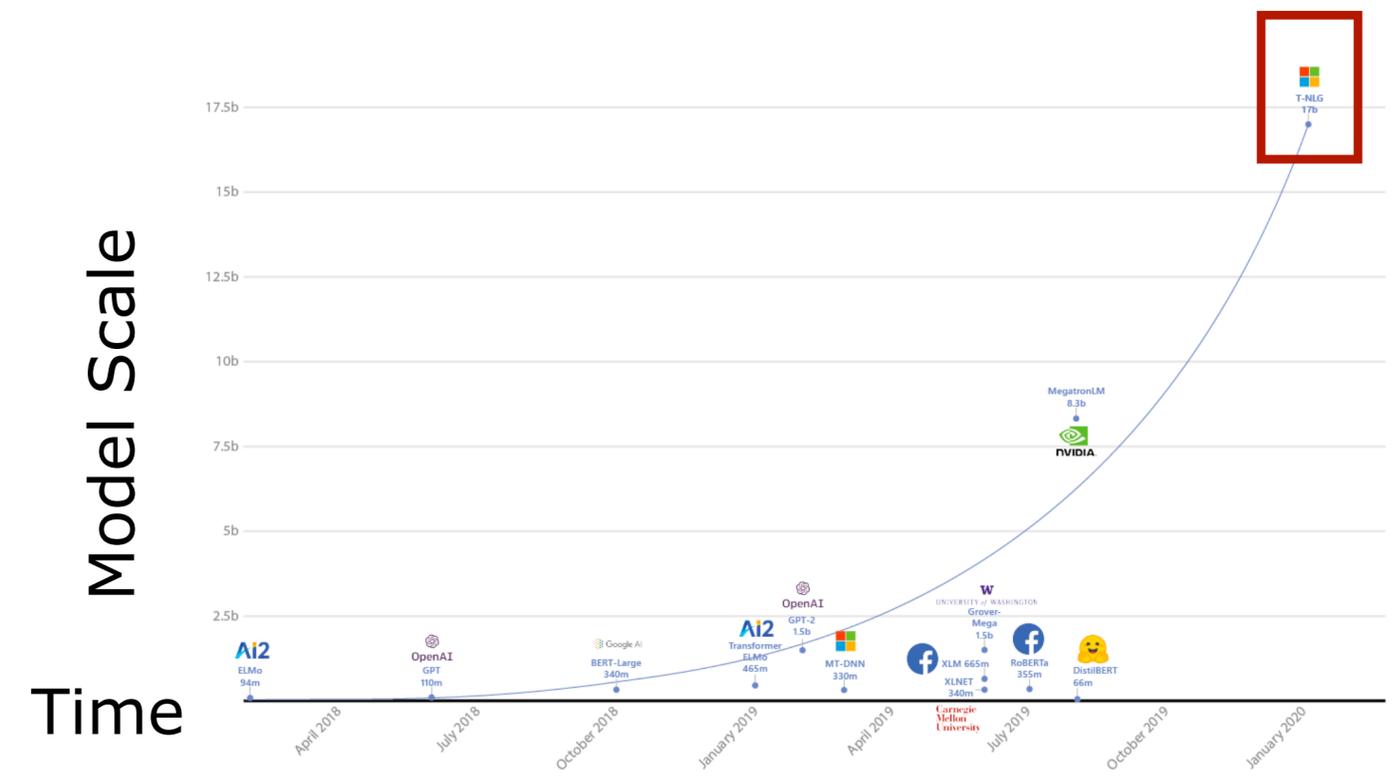
# Modern NLP



**The New York Times**  
**A Breakthrough for A.I. Technology: Passing an 8th-Grade Science Test**

**Vox** **How I'm using AI to write my next novel**

**The New York Times**  
**Meet GPT-3. It Has Learned to Code (and Blog and Argue).**  
 The latest natural-language system generates tweets, pens poetry, summarizes emails, answers trivia questions, translates languages and even writes its own computer programs.



# Trustworthy NLP systems?

DETECT LANGUAGE ENGLISH **FRENCH** SPANISH ▾

J'ai mangé mon avocat pour le déjeuner à la cafétéria. ✕

FRENCH **ENGLISH** SPANISH ▾

I ate my avocado for lunch in the cafeteria. ☆

DETECT LANGUAGE ENGLISH **FRENCH** SPANISH ▾

J'ai mangé mon avocat pour le déjeuner à la cafétéria du palais de justice. ✕

FRENCH **ENGLISH** SPANISH ▾

I ate my lawyer for lunch in the courthouse cafeteria. ☆

## Playground ⓘ



**The girl spilt orange juice on herself and started crying.**

**"Why are you crying?" her dad asked.**

**"Because my clothes are wet", replied the girl.**

**"And why are they wet?"**

"Because I fell in the swimming pool."

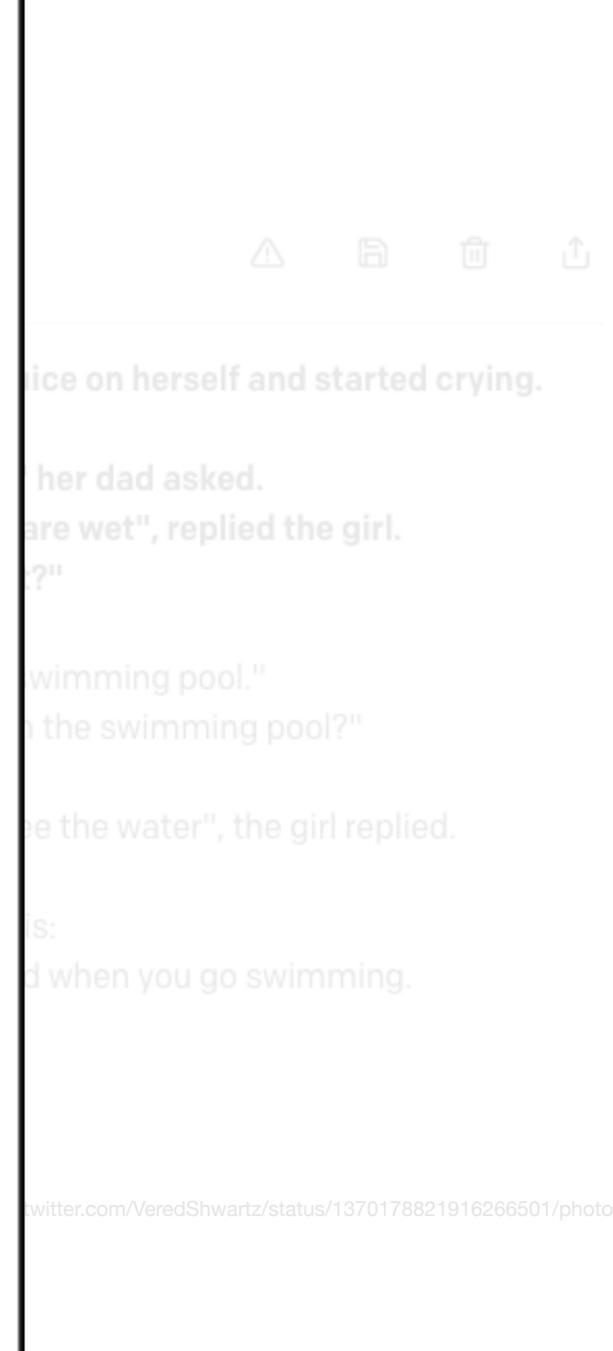
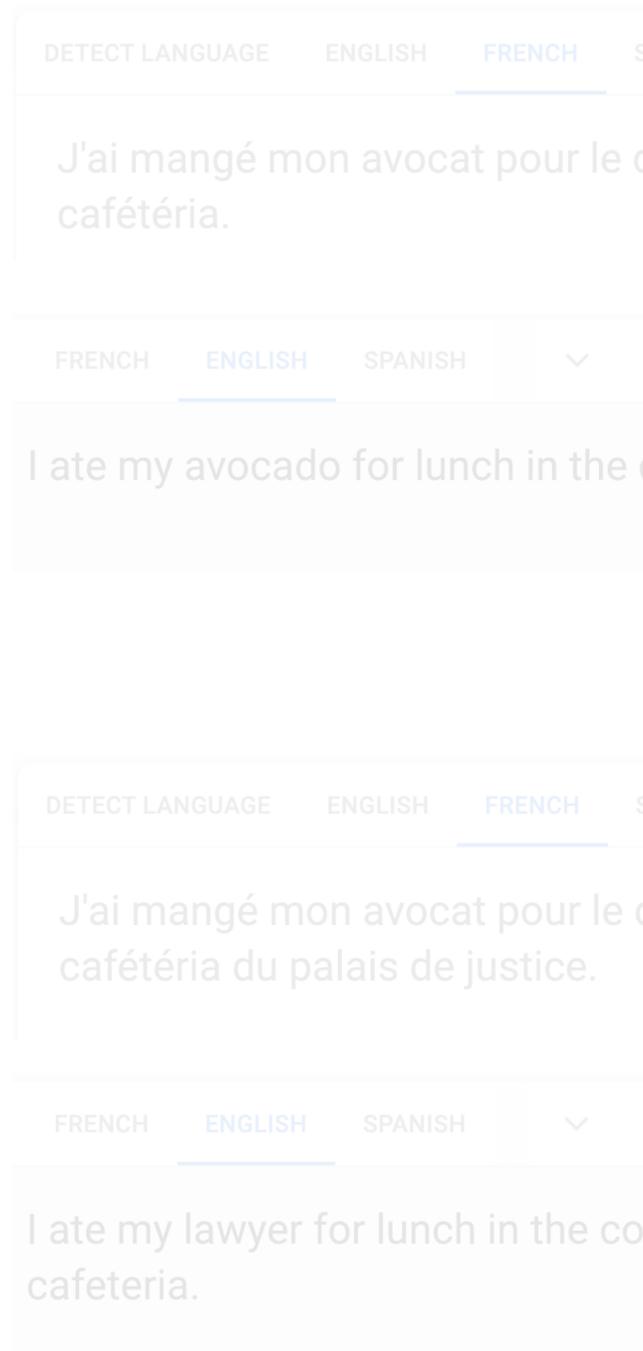
"And why did you fall in the swimming pool?"

"Because I couldn't see the water", the girl replied.

The moral of the story is:

Always wear a blindfold when you go swimming.

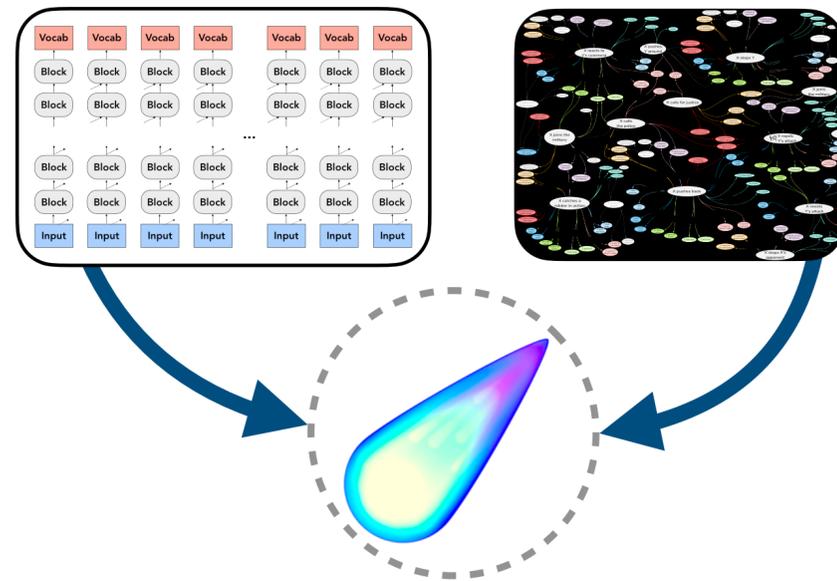
# Trustworthy NLP systems?





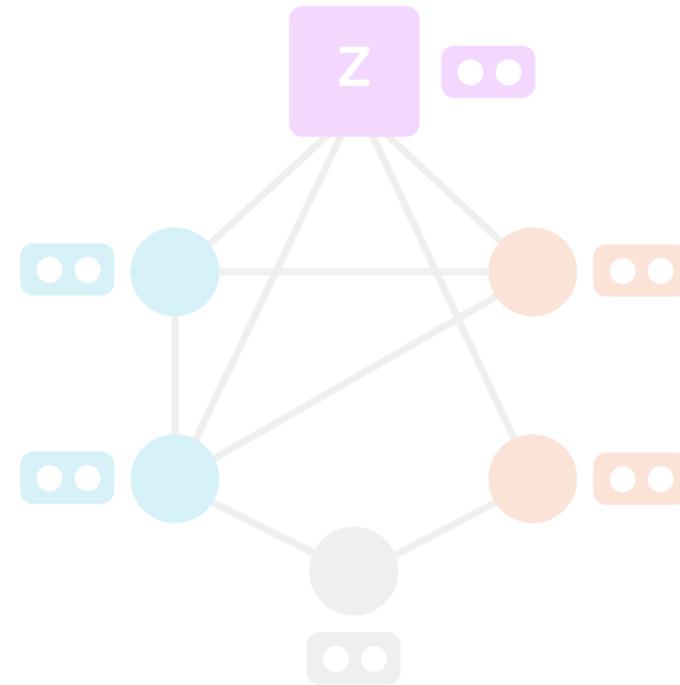
# Outline

## Language Models as KG Representations



**B**RSMCC - ACL 2019  
HBLDS**B**C - AAAI 2021  
**J**BBC - NAACL 2021  
DLLC**B** - AKBC 2021

## Language Models with Graph Reasoning



**B**LCC - AAAI 2021  
YR**B**LL - NAACL 2021  
AL**B**CM - EMNLP 2021  
Z**B**YRML - In Prep

# Language Models as Knowledge Engines

Sentence:  
I wanted to learn to sail, so I bought a |

Predictions:  
14.2% boat  
5.4% sail  
2.6% new  
2.0% small  
1.4% canoe

Sentence:  
I wanted to learn to drive, so I bought a |

Predictions:  
7.5% new  
7.0% car  
1.7% Honda  
1.7% BMW  
1.3% Ford  
← Undo

Sentence:  
I wanted to learn to read, so I bought a |

Predictions:  
17.2% book  
15.2% copy  
3.4% Kindle  
2.4% new  
1.7% few

Sentence:  
I wanted to learn to fly, so I bought a |

Predictions:  
5.3% plane  
3.8% new  
1.6% small  
1.6% Boeing  
1.5% jet  
← Undo

# Language Models as Knowledge Engines

Sentence:	Predictions:	
I wanted to learn to <u>sail</u> , so I bought a ...	<p>Prediction</p> <p>I wanted to learn to dance, so I bought a <b>dress. I had never ...</b></p> <p>I wanted to learn to dance, so I bought a <b>\$10,000 house ...</b></p> <p>I wanted to learn to dance, so I bought a <b>censored</b></p> <p>I wanted to learn to dance, so I bought a <b>US\$10 pair of ...</b></p> <p>I wanted to learn to dance, so I bought a <b>little kit and took part ...</b></p>	<p><b>Predictions:</b></p> <ul style="list-style-type: none"><li>7.5% <b>new</b></li><li>7.0% <u><b>car</b></u></li><li>1.7% <b>Honda</b></li><li>1.7% <b>BMW</b></li><li>1.3% <b>Ford</b></li><li>← Undo</li></ul>
I wanted to learn to <u>read</u> , so I bought a ...	<p><b>Predictions:</b></p> <ul style="list-style-type: none"><li>5.3% <u><b>plane</b></u></li><li>3.8% <b>new</b></li><li>1.6% <b>small</b></li><li>1.6% <b>Boeing</b></li><li>1.5% <b>jet</b></li><li>← Undo</li></ul>	

# Problem

- **Problem:** Language models encode commonsense knowledge **implicitly** from pretraining, but its representation **is not robust or reliable**

**Can we use knowledge graphs to teach language models to more effectively represent knowledge?**

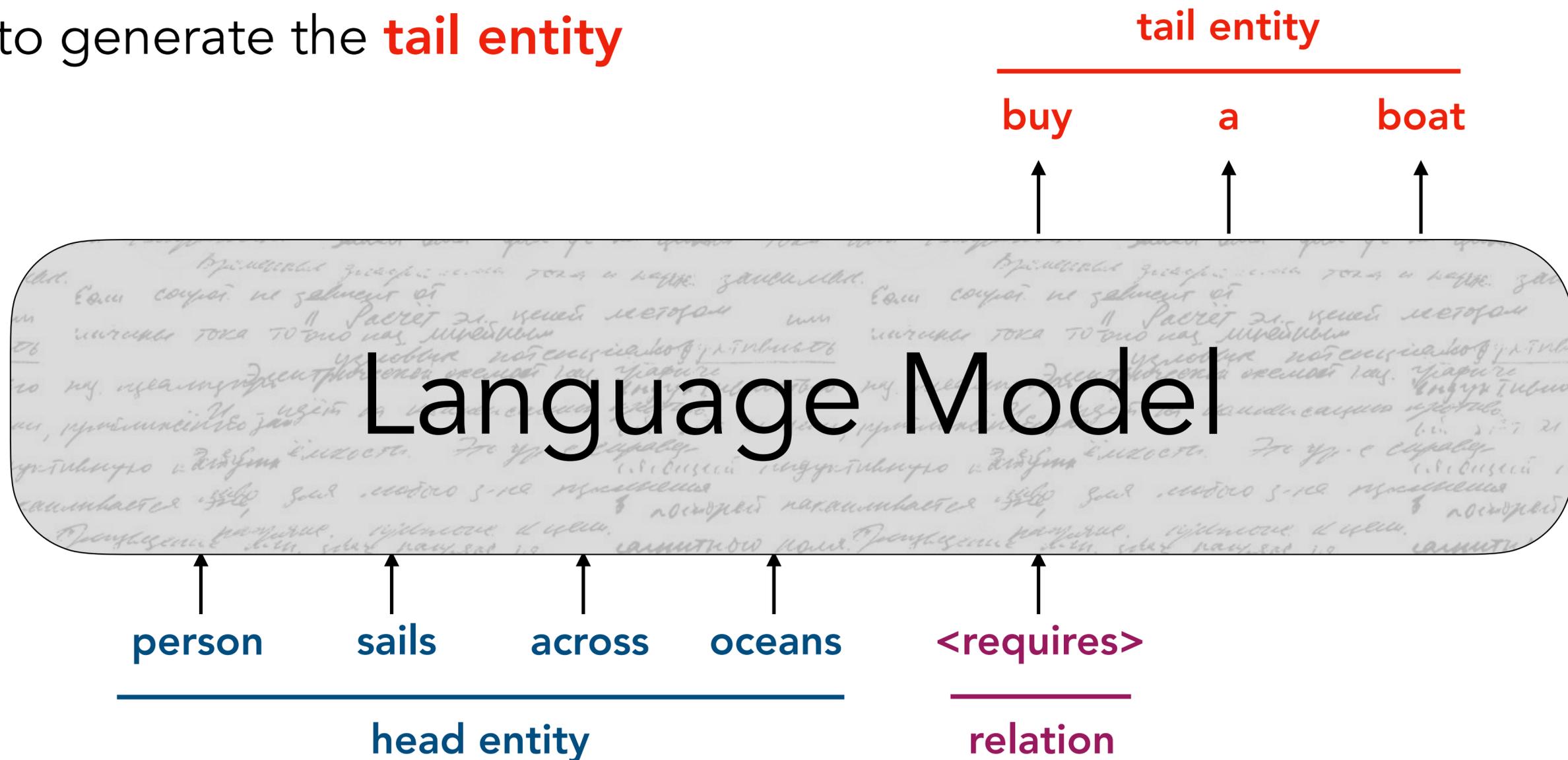
# Learning Structure of Knowledge



# Learning Structure of Knowledge

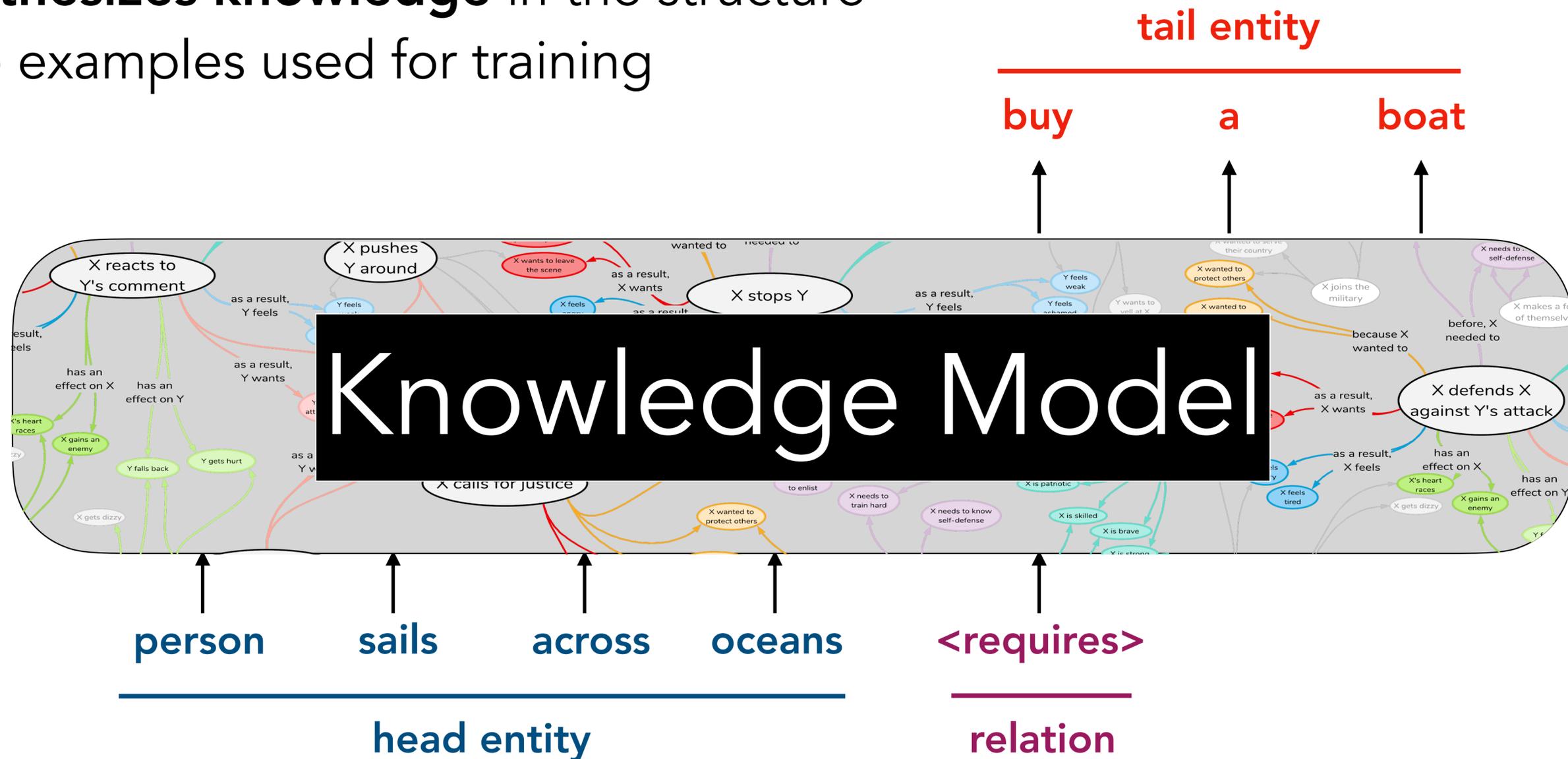
Given a **head entity** and a **relation**,  
learn to generate the **tail entity**

$$\mathcal{L} = - \sum \log P(\text{tail words} | \text{head words, relation})$$



# Learning Structure of Knowledge

Language Model → **Knowledge Model:**  
**Hypothesizes knowledge** in the structure  
of the examples used for training



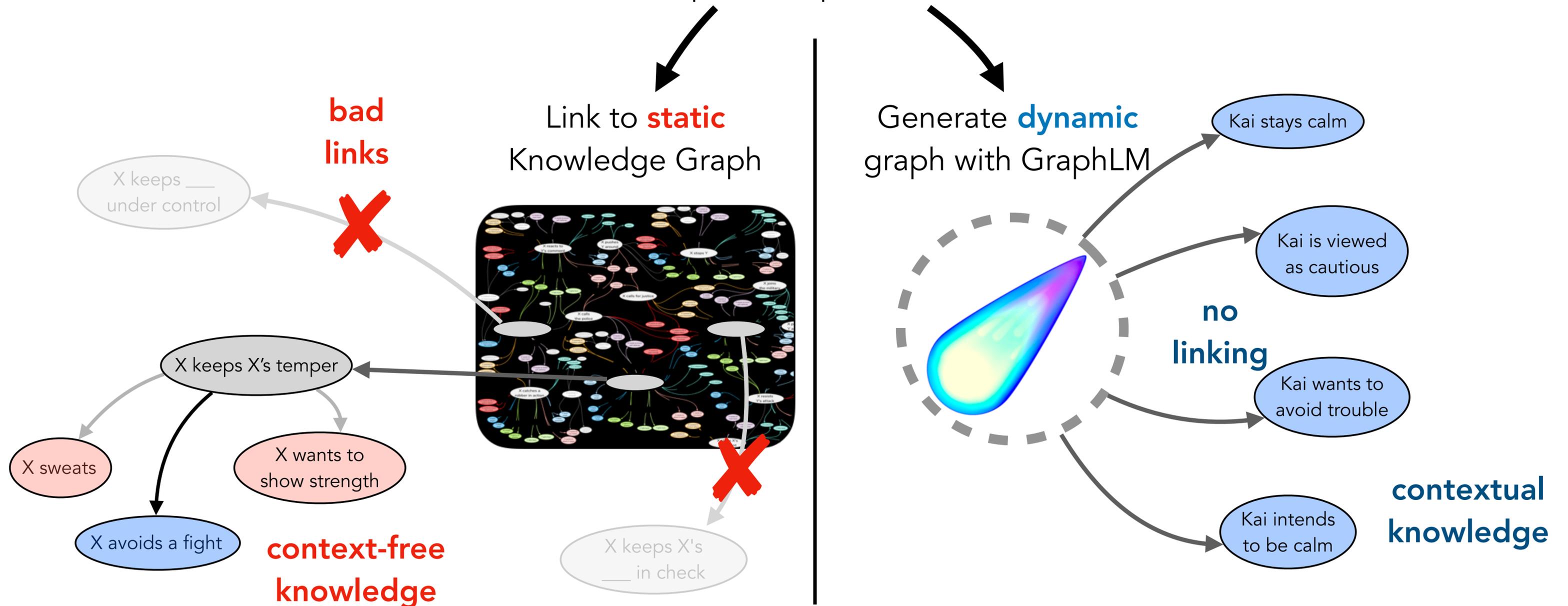


# Solution

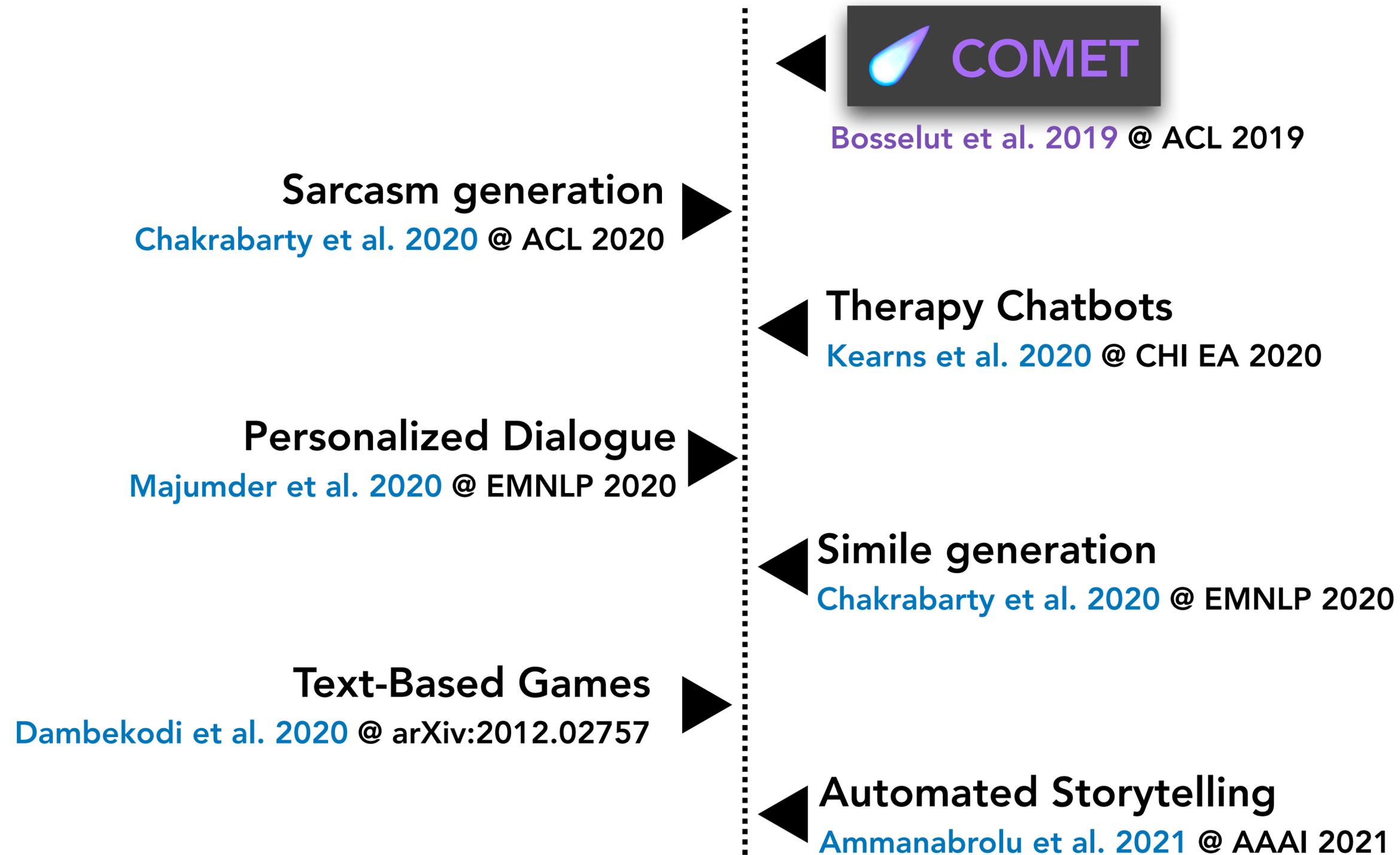
- **Problem:** Language models encode commonsense knowledge **implicitly** from pretraining, but its representation **is not robust or reliable**
- **Solution:** Use knowledge graphs to **explicitly teach commonsense relationships**, allowing LMs to generalise relationships to a broader set of implicitly encoded concepts

# Static vs. Dynamic Knowledge

Kai knew that things were getting out of control and managed to keep his temper in check

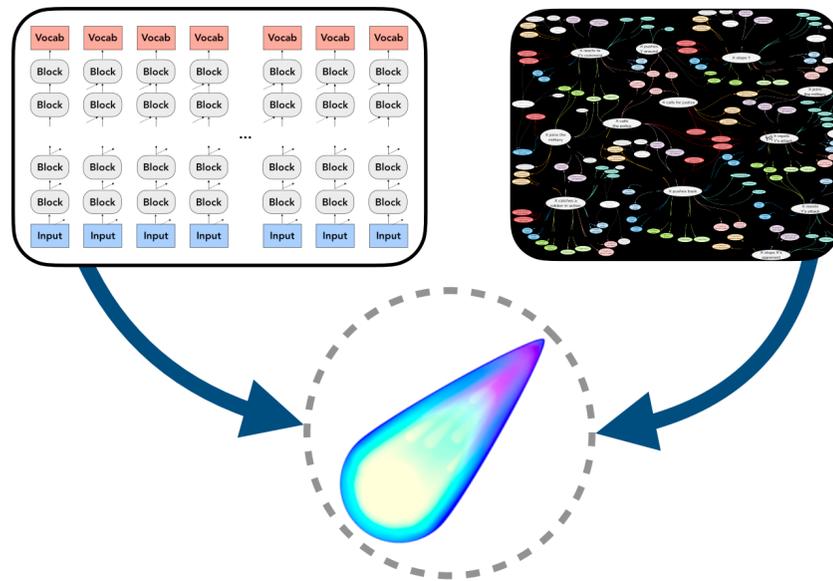


# Applications



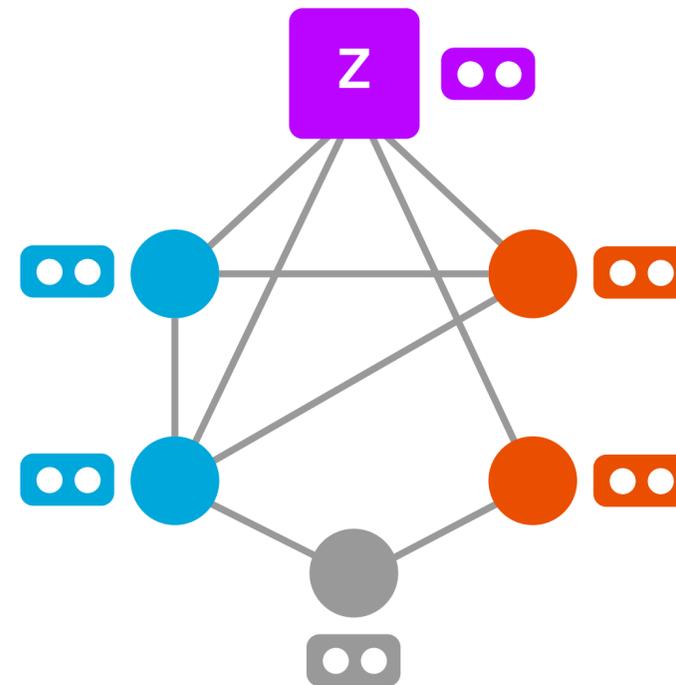
# Outline

## Language Models as Graph Representations



**B**RSMCC - ACL 2019  
HBLDS**B**C - AAAI 2021  
**J**BBC - NAACL 2021  
DLLC**B** - AKBC 2021

## Language Models with Graph Reasoning

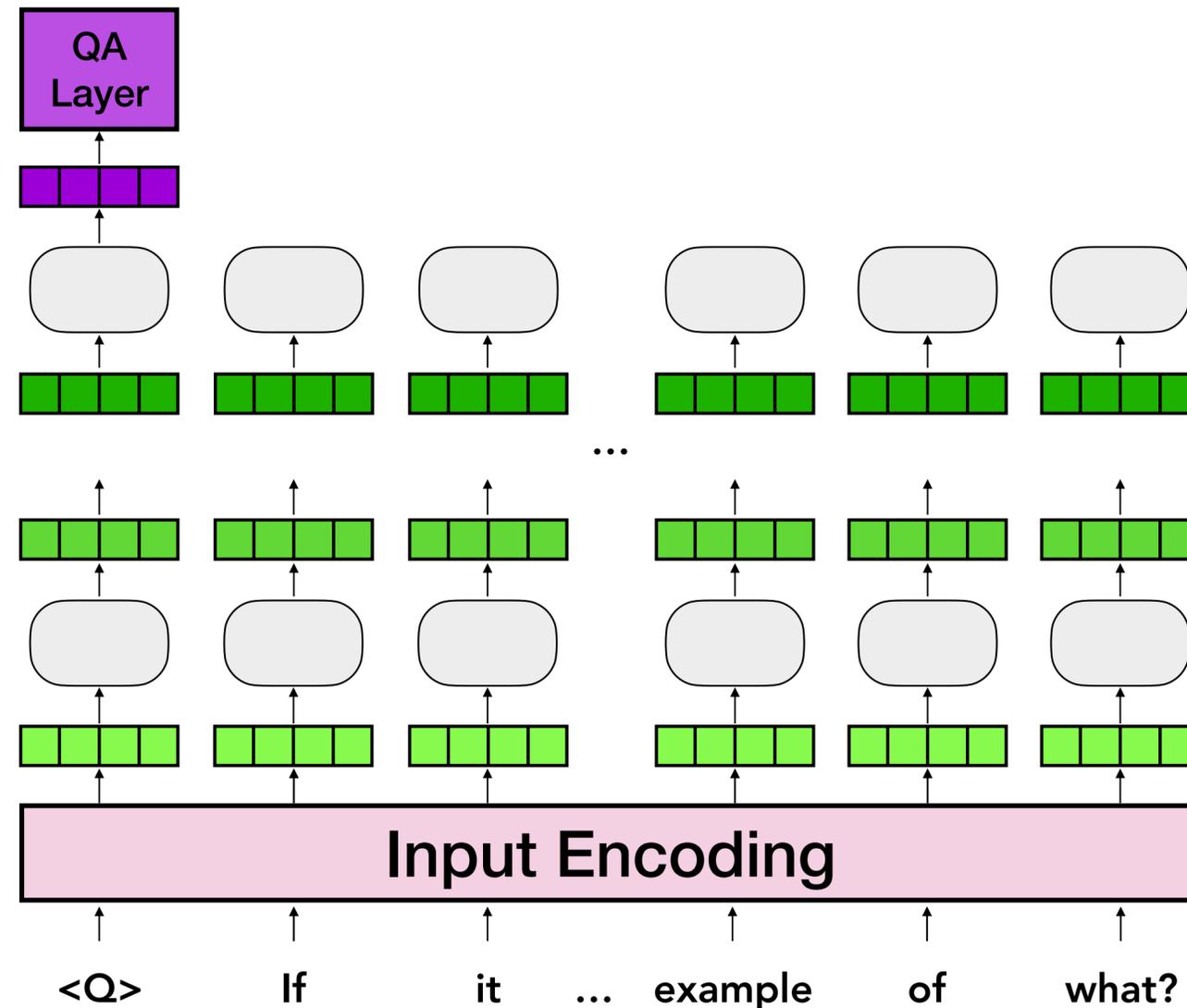


**B**LCC - AAAI 2021  
YR**B**LL - NAACL 2021  
AL**B**CM - EMNLP 2021  
**Z**BYRML - In Prep

# Neural Reasoning

If it is not used for **hair**, a **round brush** is an example of what?

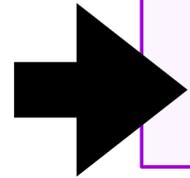
A. **hair brush** B. **bathroom** C. **art supplies\*** D. **shower**



What does the LM think about round brushes?

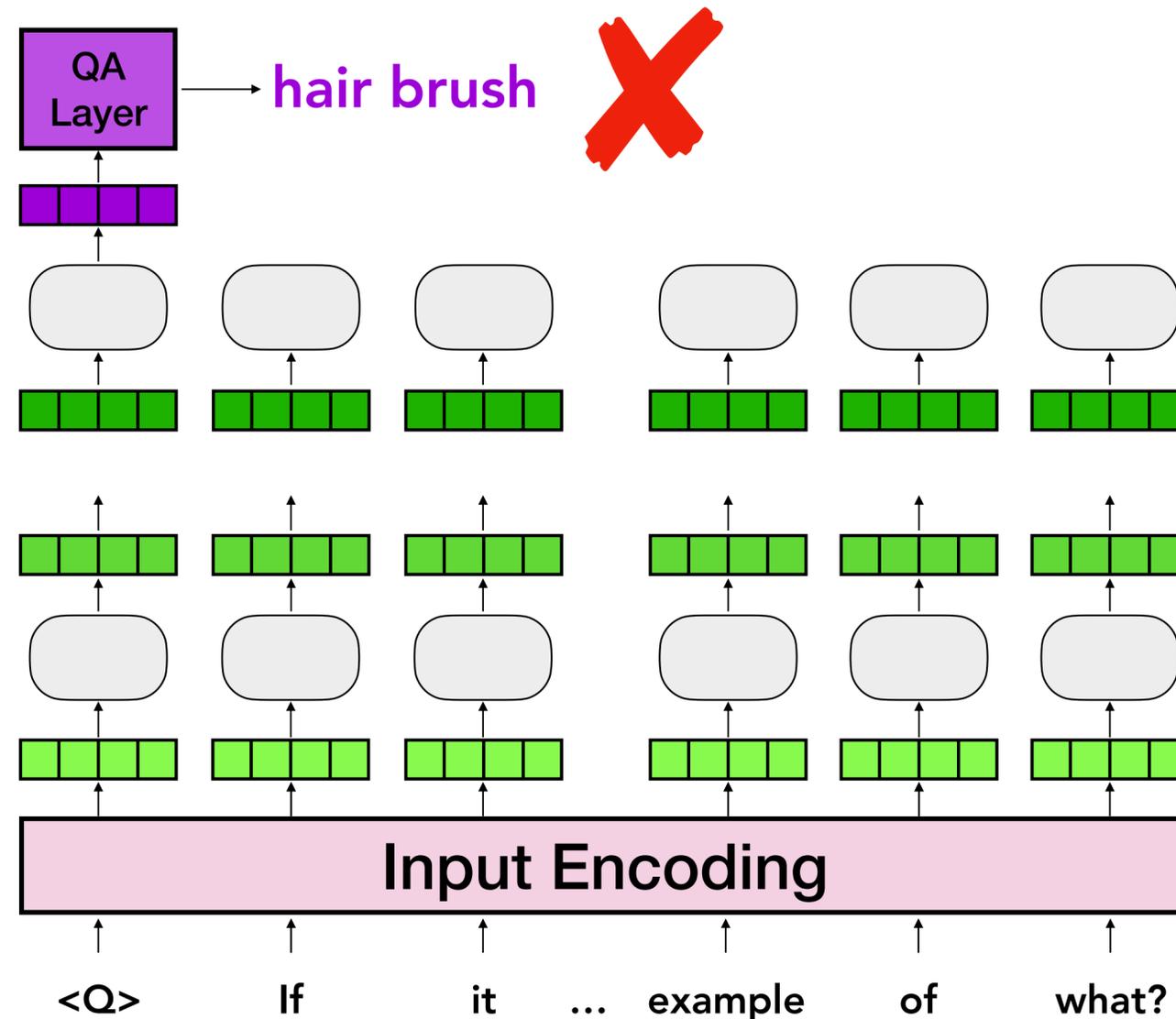
- Is used to
  - paint a picture
  - brush hair with
  - brush the hair
  - paint a portrait

# Neural Reasoning



If it is not used for **hair**, a **round brush** is an example of what?

- A. **hair brush**   B. **bathroom**   C. **art supplies\***   D. **shower**



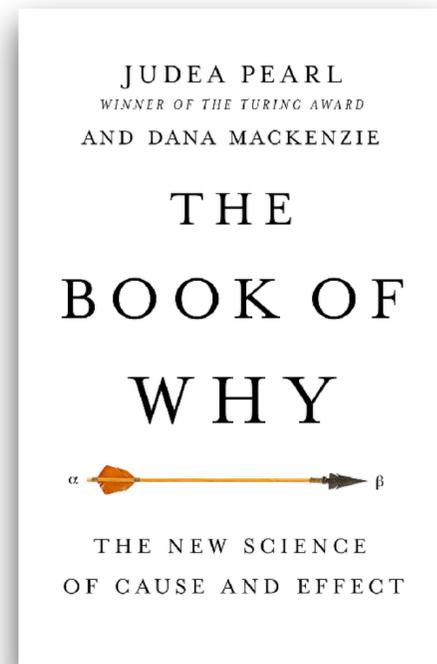
What does the LM think about round brushes?

- Is used to**
  - paint a picture
  - brush hair with
  - brush the hair
  - paint a portrait

# Reasoning with Deep Learning

Deep learning models exploit **biases** (Bolukbasi et al., 2016), **annotation artifacts** (Gururangan et al., 2018), **surface patterns** (Li & Gauthier, 2017), etc.

**They do not learn viable reasoning capabilities**



(Pearl, 2018)

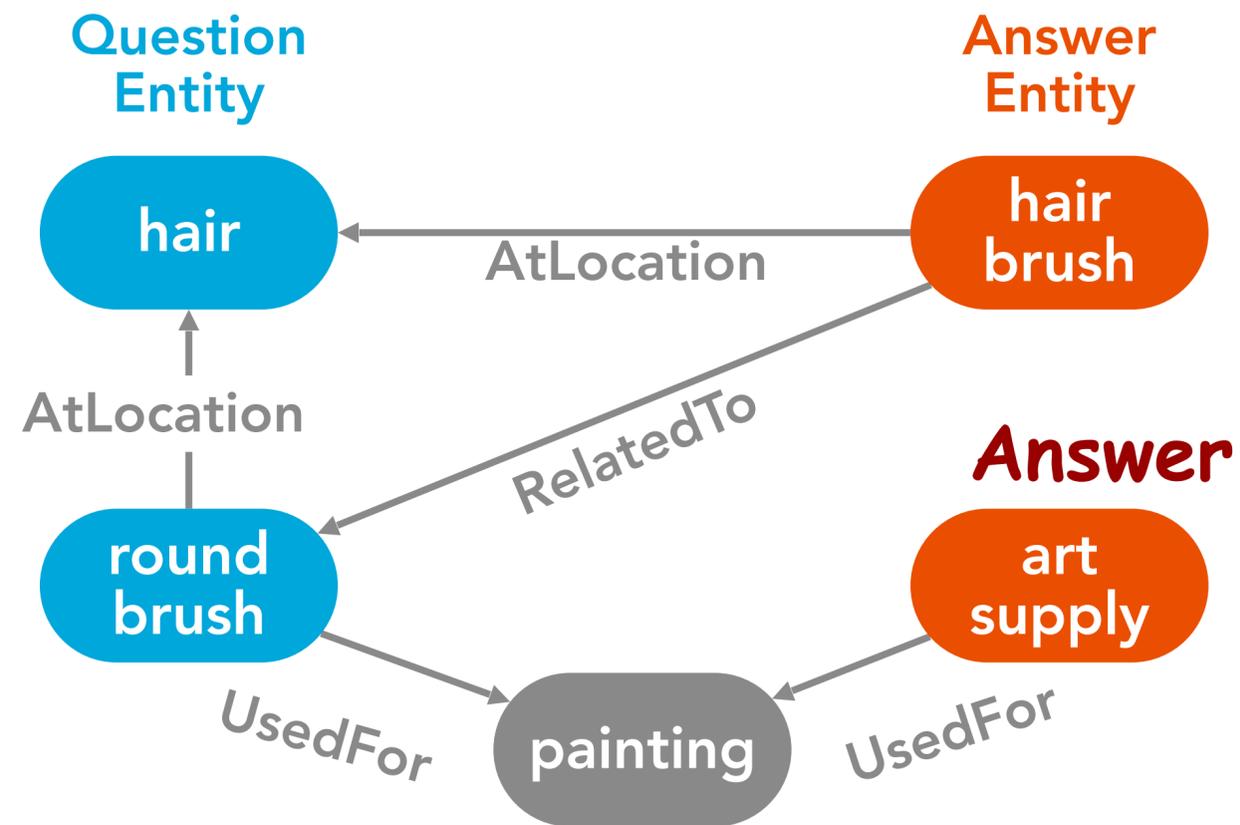
"All the impressive achievements of deep learning amount to just curve fitting"



# Knowledge Graphs for Reasoning

**Textual  
Context**

If it is **not** used for **hair**, a **round brush** is an example of what?  
A. **hair brush** B. **bathroom** C. **art supplies\*** D. **shower**



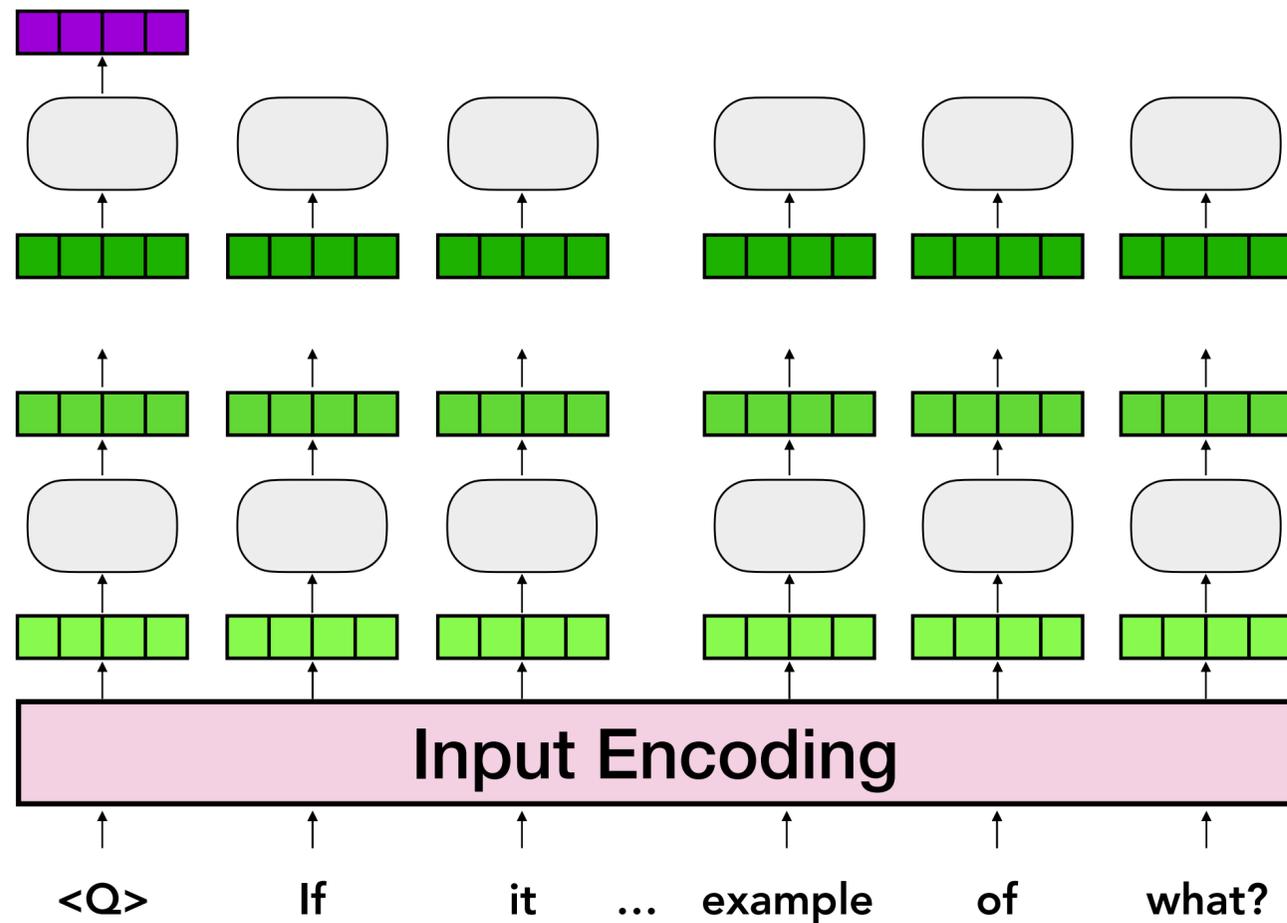
**Knowledge Graph**

**How can we coordinate the reasoning implied  
by language with explicit knowledge from graphs?**

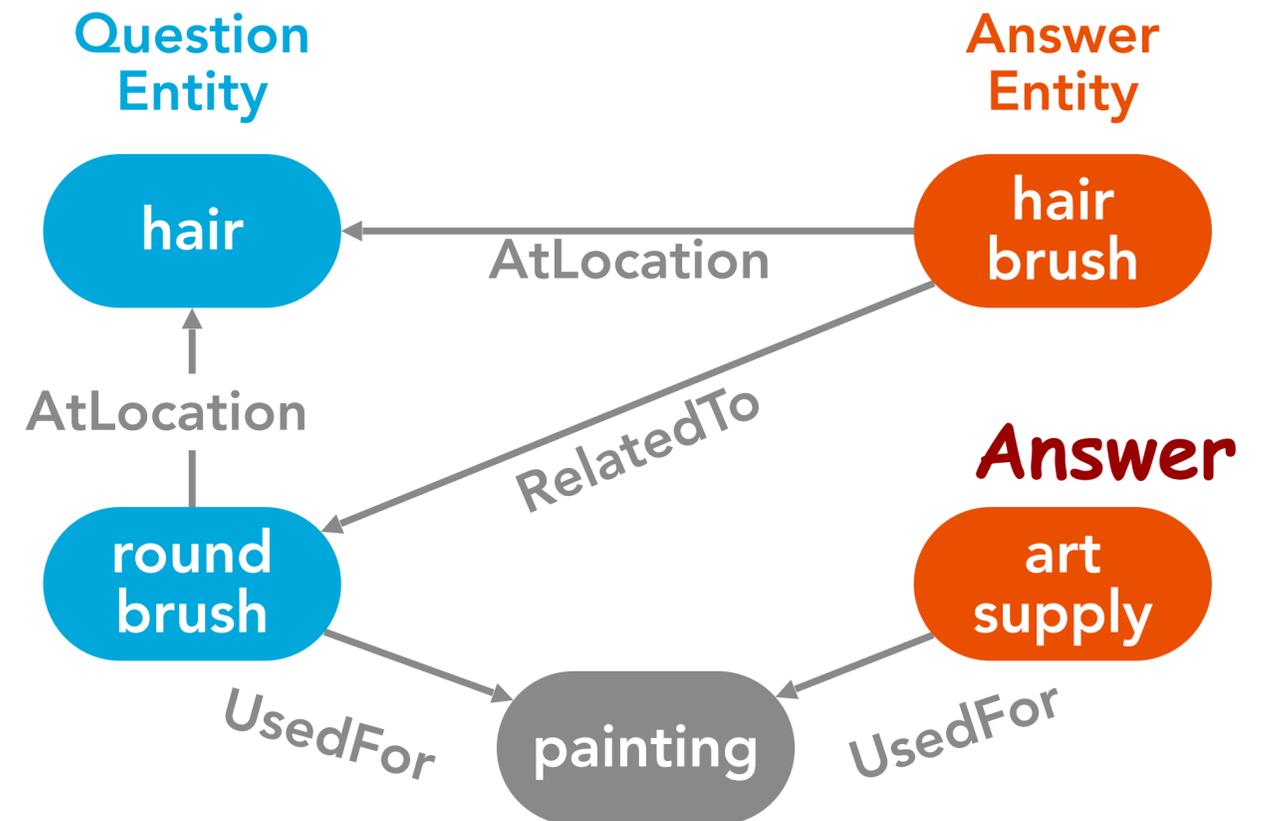
# QA-GNN: Overview

**Textual  
Context**

If it is **not** used for **hair**, a **round brush** is an example of what?  
A. **hair brush** B. **bathroom** C. **art supplies\*** D. **shower**



**Language Model**



**Knowledge Graph**

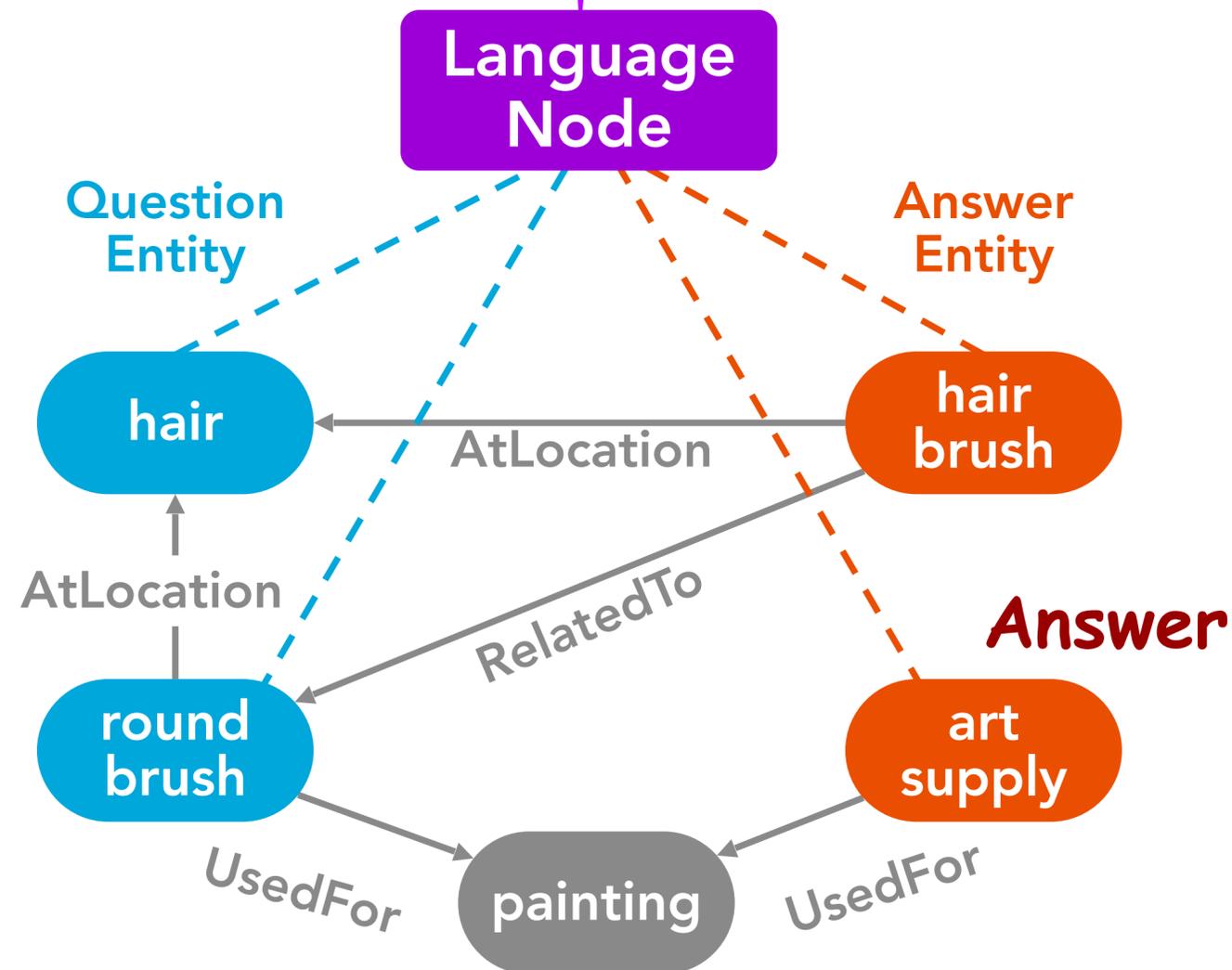
# QA-GNN: Overview

**Textual  
Context**

If it is **not** used for **hair**, a **round brush** is an example of what?

A. **hair brush** B. **bathroom** C. **art supplies\*** D. **shower**

**Knowledge  
Graph**



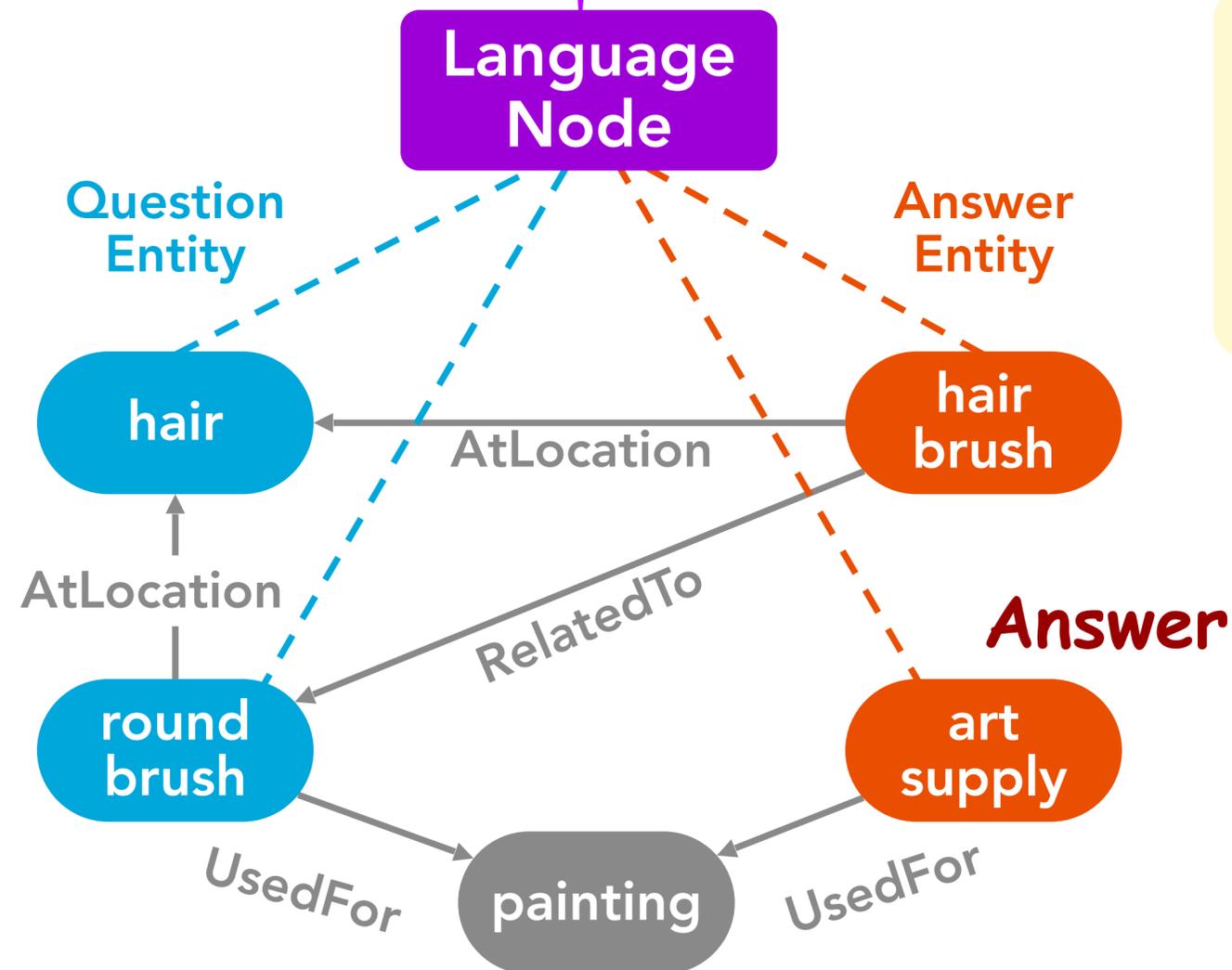
# QA-GNN: Overview

**Textual  
Context**

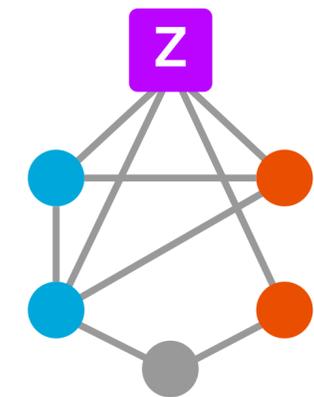
If it is **not** used for **hair**, a **round brush** is an example of what?

A. **hair brush** B. **bathroom** C. **art supplies\*** D. **shower**

**Knowledge  
Graph**



**Joint graph** representation that brings language and knowledge together in a shared semantic space



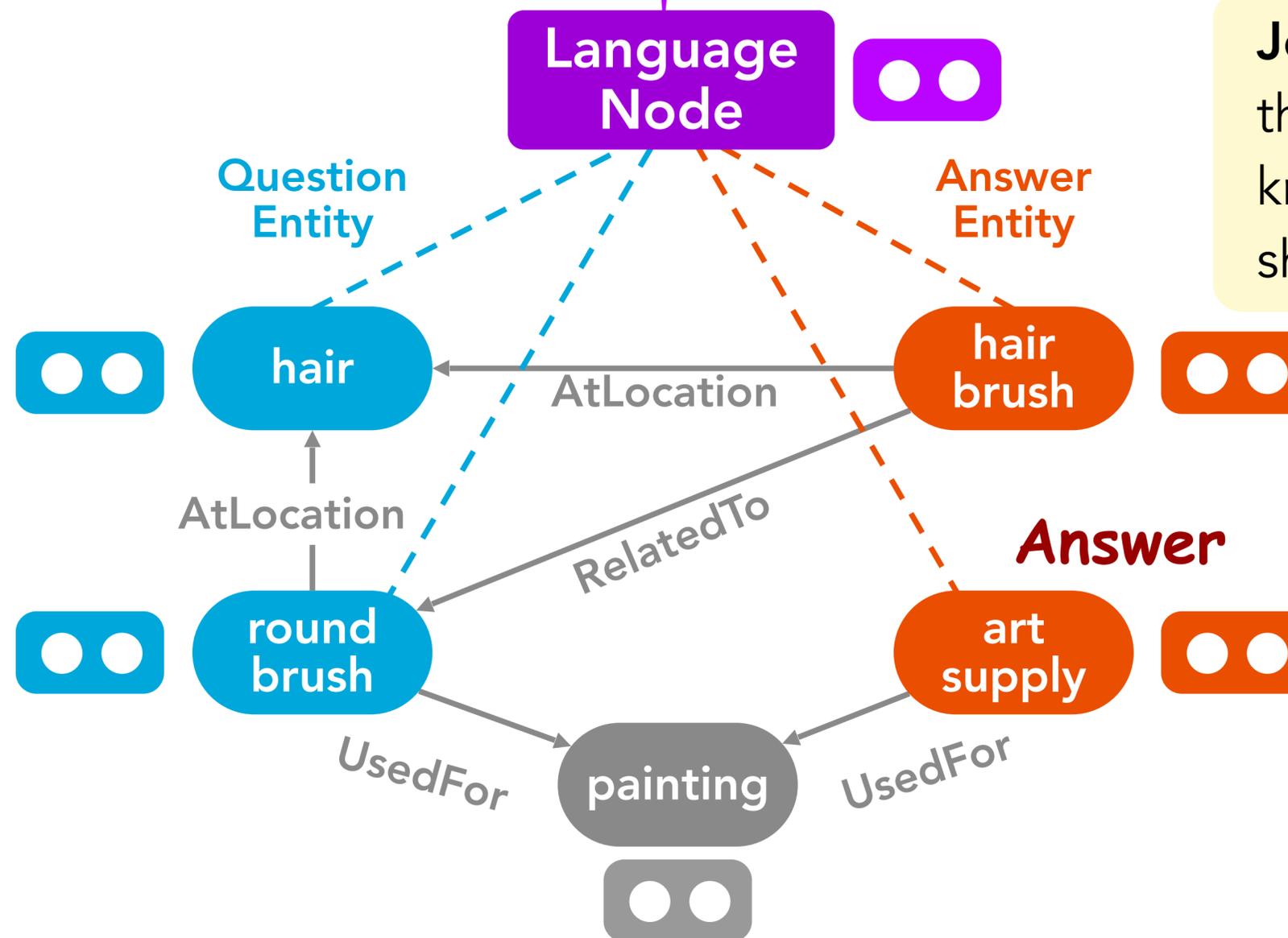
# QA-GNN: Overview

**Textual  
Context**

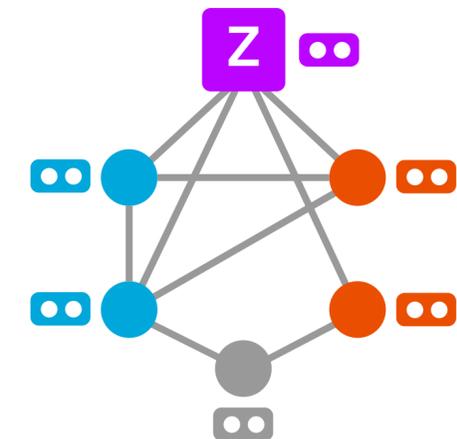
If it is **not** used for **hair**, a **round brush** is an example of what?

A. **hair brush** B. **bathroom** C. **art supplies\*** D. **shower**

**Knowledge  
Graph**



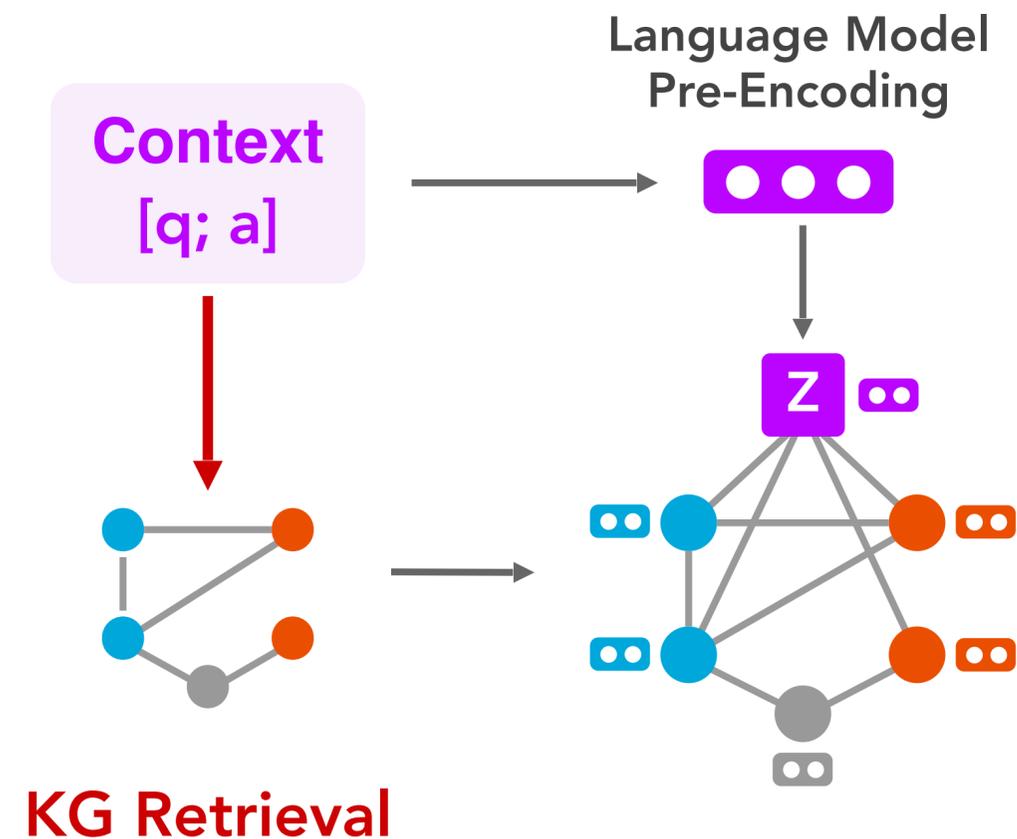
**Joint graph** representation that brings language and knowledge together in a shared semantic space



# QA-GNN: Overview

If it is not used for **hair**, a **round brush** is an example of what?

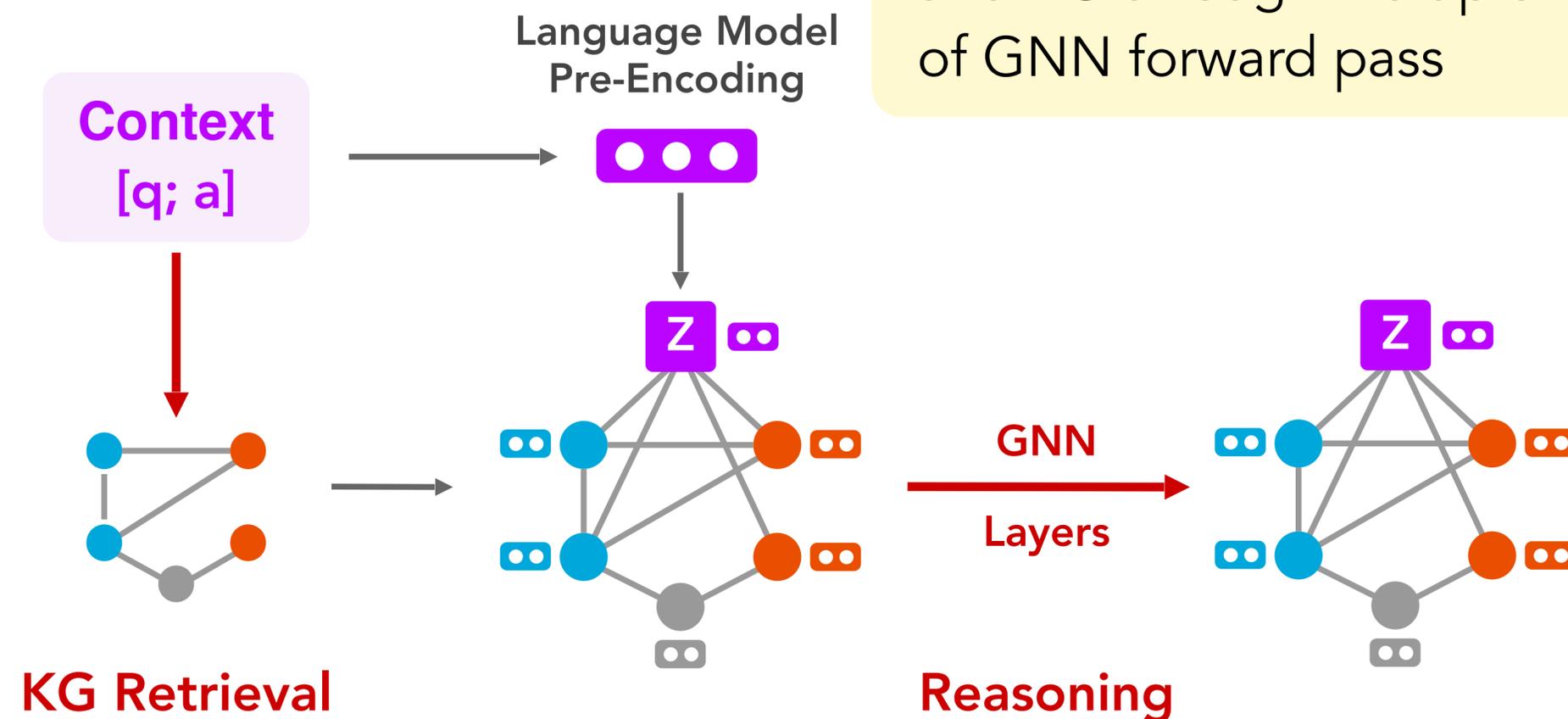
- A. **hair brush**   B. **bathroom**   C. **art supplies\***   D. **shower**



# QA-GNN: Overview

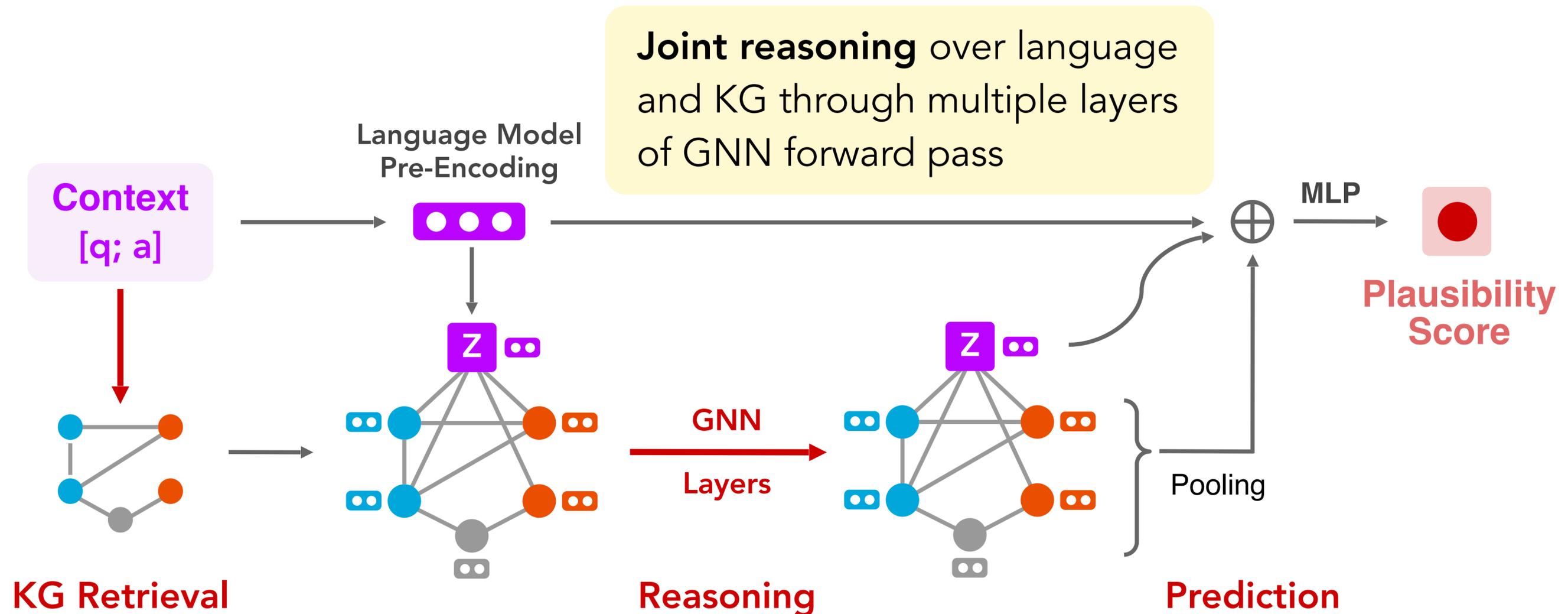
If it is not used for **hair**, a **round brush** is an example of what?

A. **hair brush** B. **bathroom** C. **art supplies\*** D. **shower**



# QA-GNN: Overview

If it is not used for **hair**, a **round brush** is an example of what?  
A. **hair brush** B. **bathroom** C. **art supplies\*** D. **shower**



# Solution

- **Problem:** Machine learning models **shortcut reasoning** by learning to exploit easier statistical patterns (e.g., biases)
- **Solution:** **Jointly model** language and knowledge in a **shared graph structure**, enabling **interaction** between modalities

# Experimental Setup

**CommonsenseQA:** reasoning with commonsense knowledge

What do people typically do while playing guitar?

- (A) cry
- (B) hear sounds
- (C) singing**
- (D) arthritis
- (E) making music

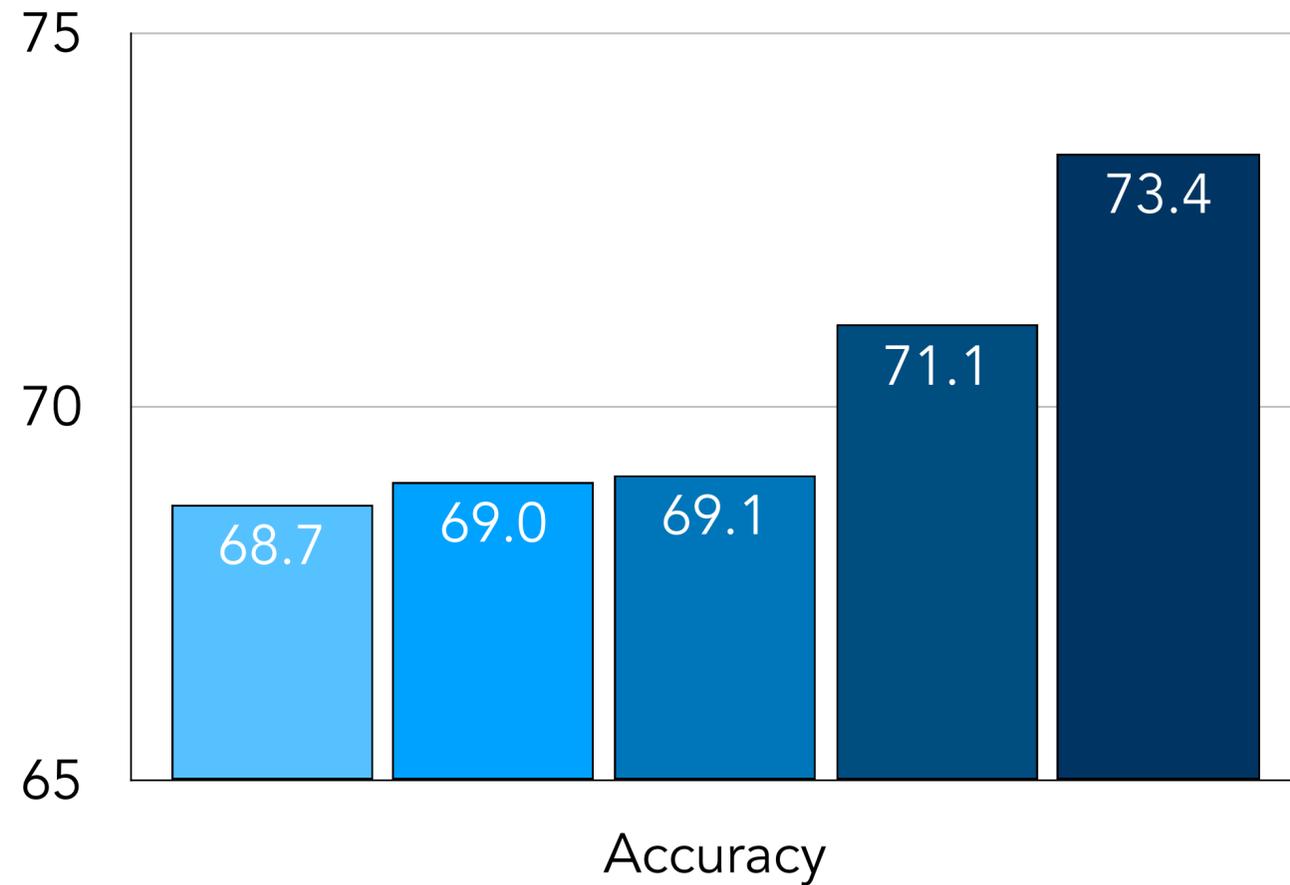
**OpenBookQA:** reasoning with elementary science knowledge

Which of these would let the most heat travel through?

- (A) a new pair of jeans
- (B) a steel spoon in a cafeteria**
- (C) a cotton candy at a store
- (D) a calvin klein cotton hat

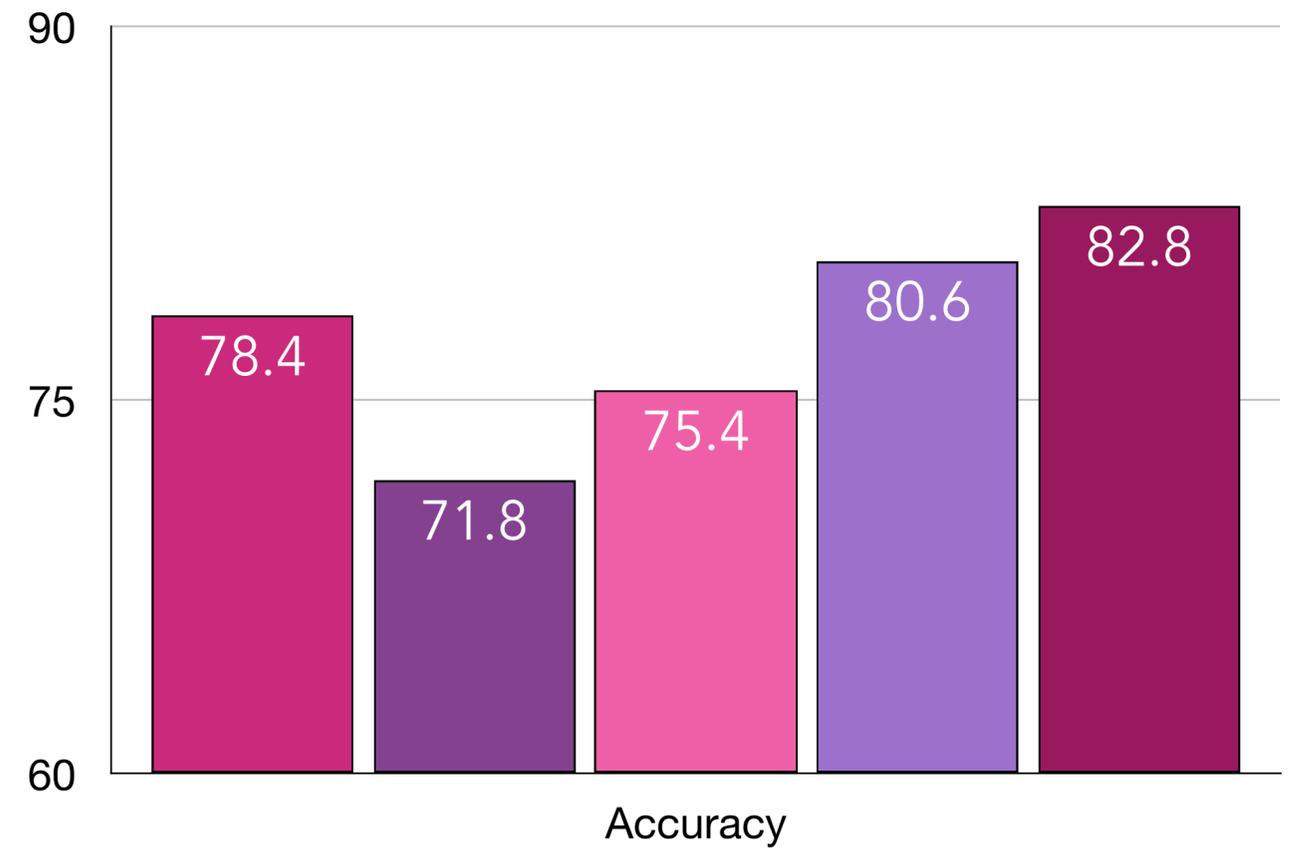
# Experimental Results

## CommonsenseQA



- RoBERTa (Liu, 2019)
- ReINet (Santoro, 2017)
- QA-GNN (Ours)
- KagNet (Lin, 2019)
- MHGRN (Feng, 2020)

## OpenbookQA



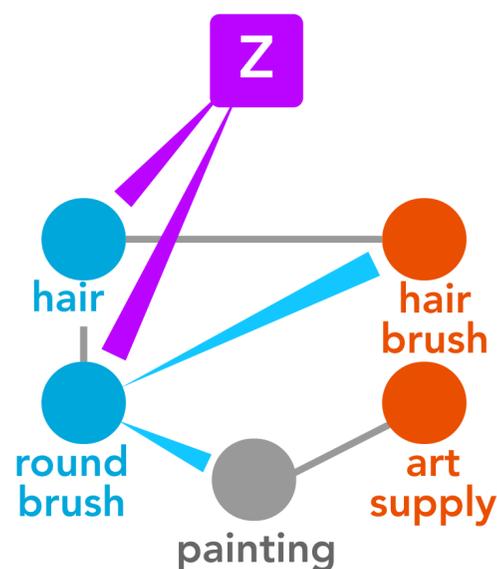
- AristoRoBERTa (Liu, 2019)
- ReINet (Santoro, 2017)
- QA-GNN (Ours)
- GconAttn (Wang, 2019)
- MHGRN (Feng, 2020)

# Structured Reasoning - Negation

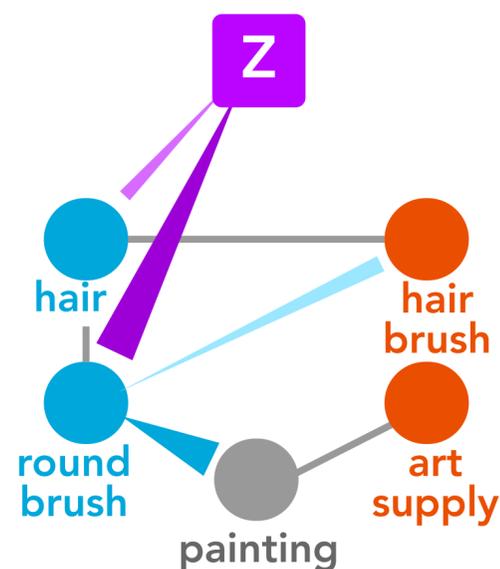
## Original Question

If it is **not** used for **hair**, a **round brush** is an example of what?

A. **hair brush** B. **art supplies\***



GNN 1st Layer



GNN Final Layer

(  
A. hair brush (#1)  
B. art supplies (#2)  
RoBERTa Prediction  
)

A. hair brush (#2)

**B. art supplies (#1)**

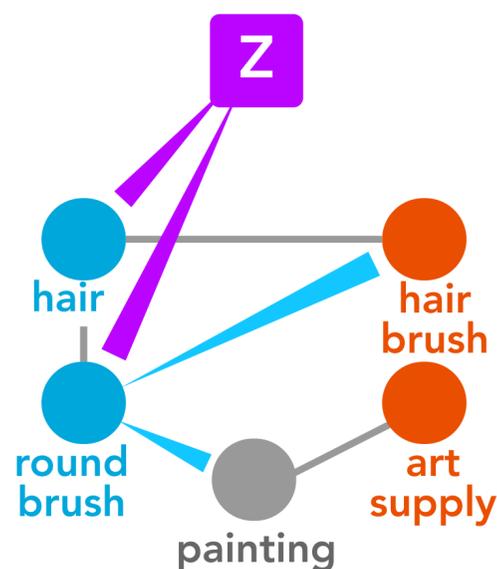
Our GNN Prediction

The attention weight by **Z** over **hair** decreases and increases on **round brush**, The attention weight by **round brush** over **painting** increases.

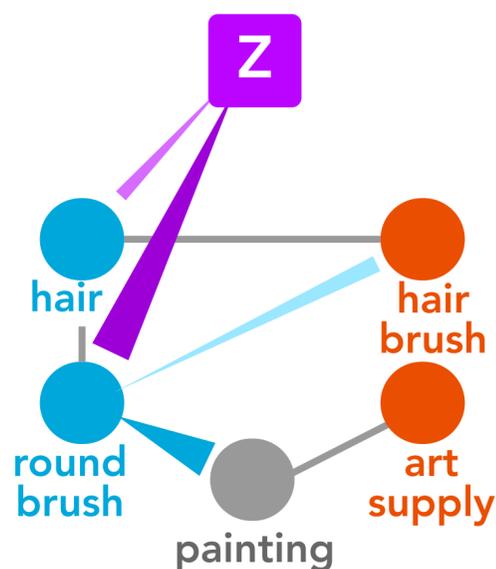
# Structured Reasoning - Negation

## Original Question

If it is **not** used for **hair**, a **round brush** is an example of what?  
A. **hair brush** B. **art supplies\***



GNN 1st Layer



GNN Final Layer

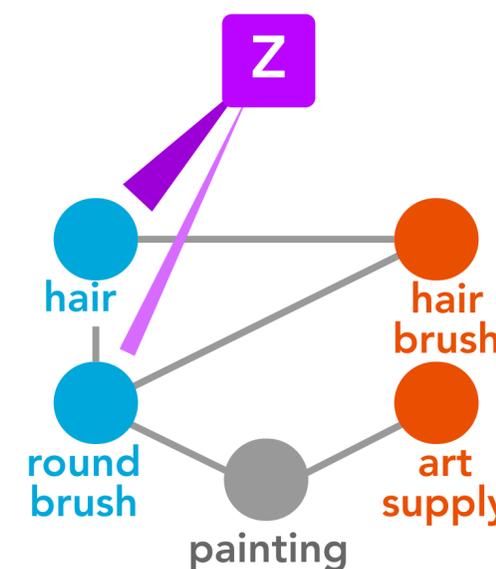
(  
A. hair brush (#1)  
B. art supplies (#2)  
RoBERTa Prediction  
)

A. hair brush (#2)  
B. **art supplies (#1)**

Our GNN Prediction

## Negation Removed

If it is used for **hair**, a **round brush** is an example of what?  
A. **hair brush** B. **art supplies**



GNN Final Layer

(  
A. hair brush (#1)  
B. art supplies (#2)  
RoBERTa Prediction  
)

**A. hair brush (#1)**  
B. art supplies (#2)

Our GNN Prediction

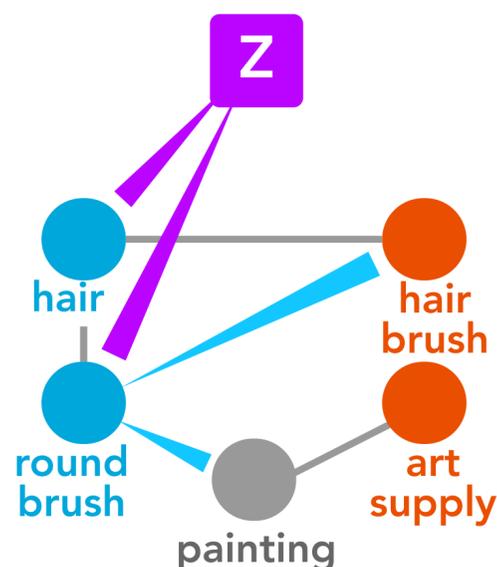
The attention weight by **Z** over **hair** decreases and increases on **round brush**, The attention weight by **round brush** over **painting** increases.

The attention weight of **Z** on **hair** now increases in the final layer of the GNN.

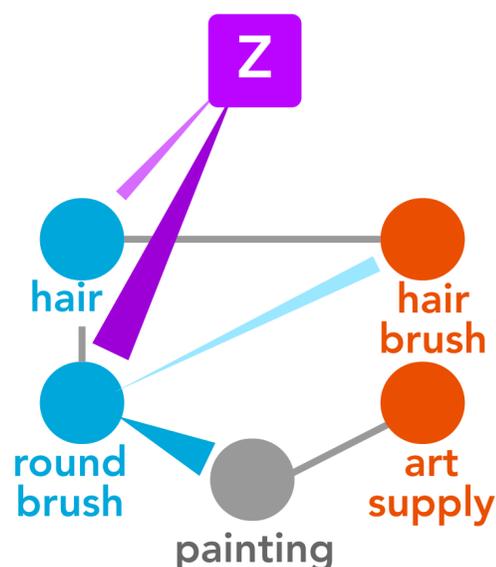
# Structured Reasoning - Negation

## Original Question

If it is **not** used for **hair**, a **round brush** is an example of what?  
A. **hair brush** B. **art supplies\***



GNN 1st Layer



GNN Final Layer

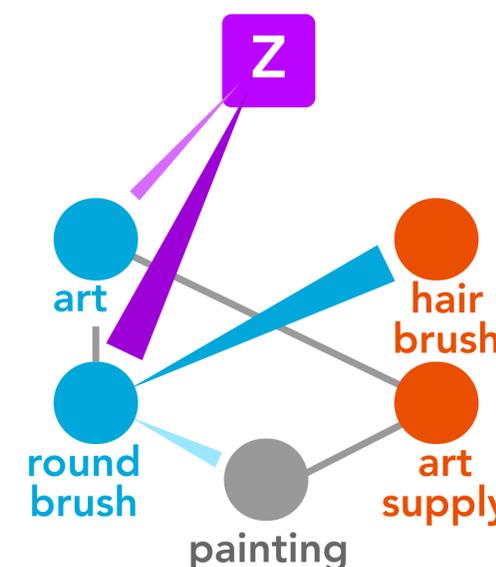
(  
A. hair brush (#1)  
B. art supplies (#2)  
RoBERTa Prediction  
)

A. hair brush (#2)  
B. **art supplies (#1)**

Our GNN Prediction

## Entity Changed (hair → art)

If it is **not** used for **art**, a **round brush** is an example of what? A. **hair brush** B. **art supplies**



GNN Final Layer

(  
A. hair brush (#1)  
B. art supplies (#2)  
RoBERTa Prediction  
)

A. **hair brush (#1)**  
B. art supplies (#2)

Our GNN Prediction

The attention weight by **Z** over **hair** decreases and increases on **round brush**, The attention weight by **round brush** over **painting** increases.

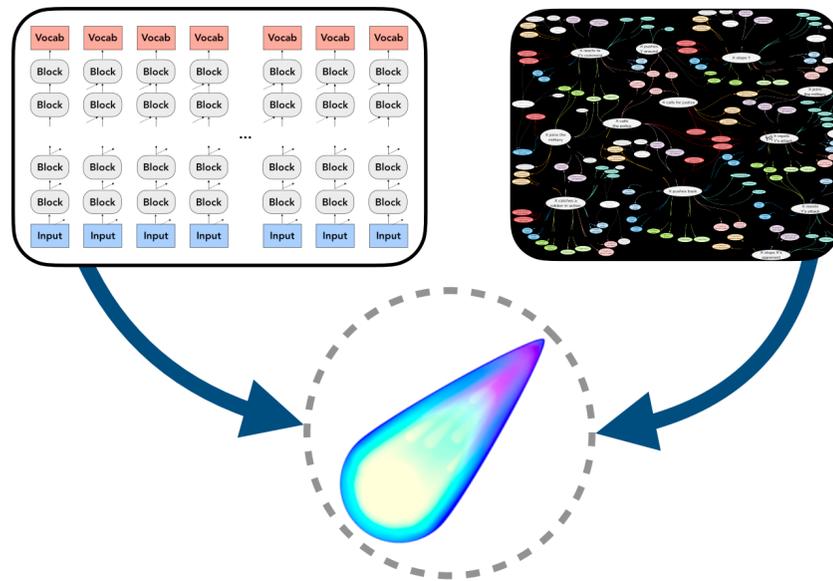
The attention weight of **Z** on **art** becomes low in the final layer of the GNN. **Round brush** attends to **hair brush** now.

# Impact

- **Problem:** Machine learning and symbolic methods **not individually robust** enough for language-based reasoning
- **Solution:** **Jointly model** language and knowledge in a **shared graph structure**, enabling **interaction** between modalities
- **Impact:** Knowledge graph becomes a **scaffold for neural reasoning** and natural language understanding

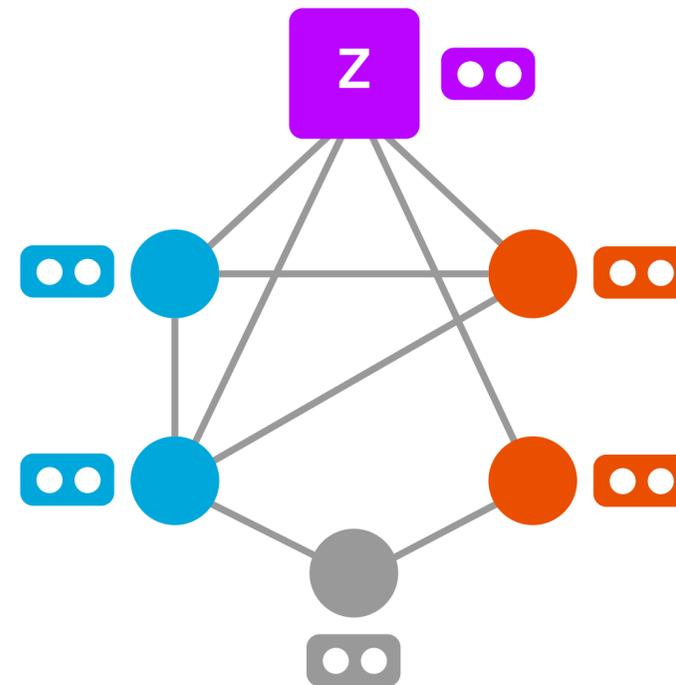
# Outline

## Language Models as Graph Representations



**B**RSMCC - ACL 2019  
HBLDS**B**C - AAAI 2021  
**J**BBC - NAACL 2021  
DLLC**B** - AKBC 2021

## Language Models with Graph Reasoning



**B**LC - AAAI 2021  
YR**B**LL - NAACL 2021  
AL**B**CM - EMNLP 2021  
**Z**BYRML - In Prep

# Final Thoughts

- **Representation: Learning knowledge relationships** from KG examples enables pretrained language models to **generalise those relationships** to implicitly encoded knowledge learned during pretraining
- **Reasoning: Interfacing language models with graph structure** allows language models to **learn scaffolds** for robust neural reasoning

# References

- **Antoine Bosselut**, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Çelikyilmaz, Yejin Choi. *COMET: Commonsense Transformers for Automatic Knowledge Graph Construction*. ACL 2019.  
<http://arxiv.org/abs/1906.05317>
- Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, **Antoine Bosselut**, Yejin Choi (2021). *(Comet-)Atomic 2020: On Symbolic and Neural Commonsense Knowledge Graphs*. AAAI 2021.  
<https://arxiv.org/abs/2010.05953>
- Michihiro Yasunaga, Hongyu Ren, **Antoine Bosselut**, Percy Liang, Jure Leskovec (2021). *QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering*. NAACL 2021.  
<https://arxiv.org/abs/2104.06378>
- Additional work in progress!

# Collaborators

