Neural Language Models as Commonsense Representation Engines

Antoine Bosselut

Language Models

The New York Times

Finally, a Machine That Can Finish Your Sentence

Completing someone else's though
But new systems are starting to
language

THE NEW YORKER

The Next Word

will predictive text take us?

Text by John Seabrook

The New York Times

A Breakthrough for A.I. Technology: Passing an 8th-Grade

Vox

How I'm using AI to write my next novel

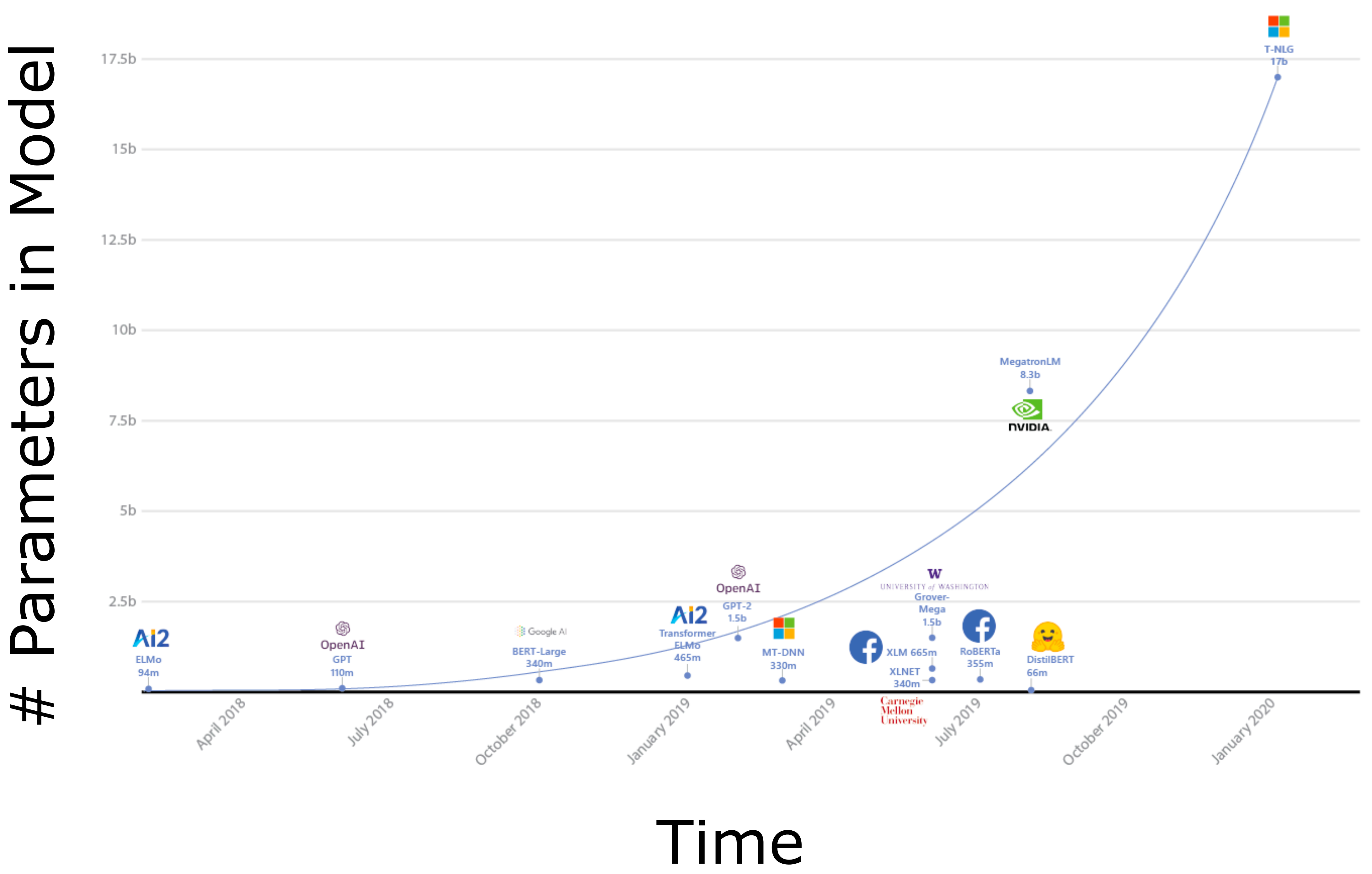
The New York

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.

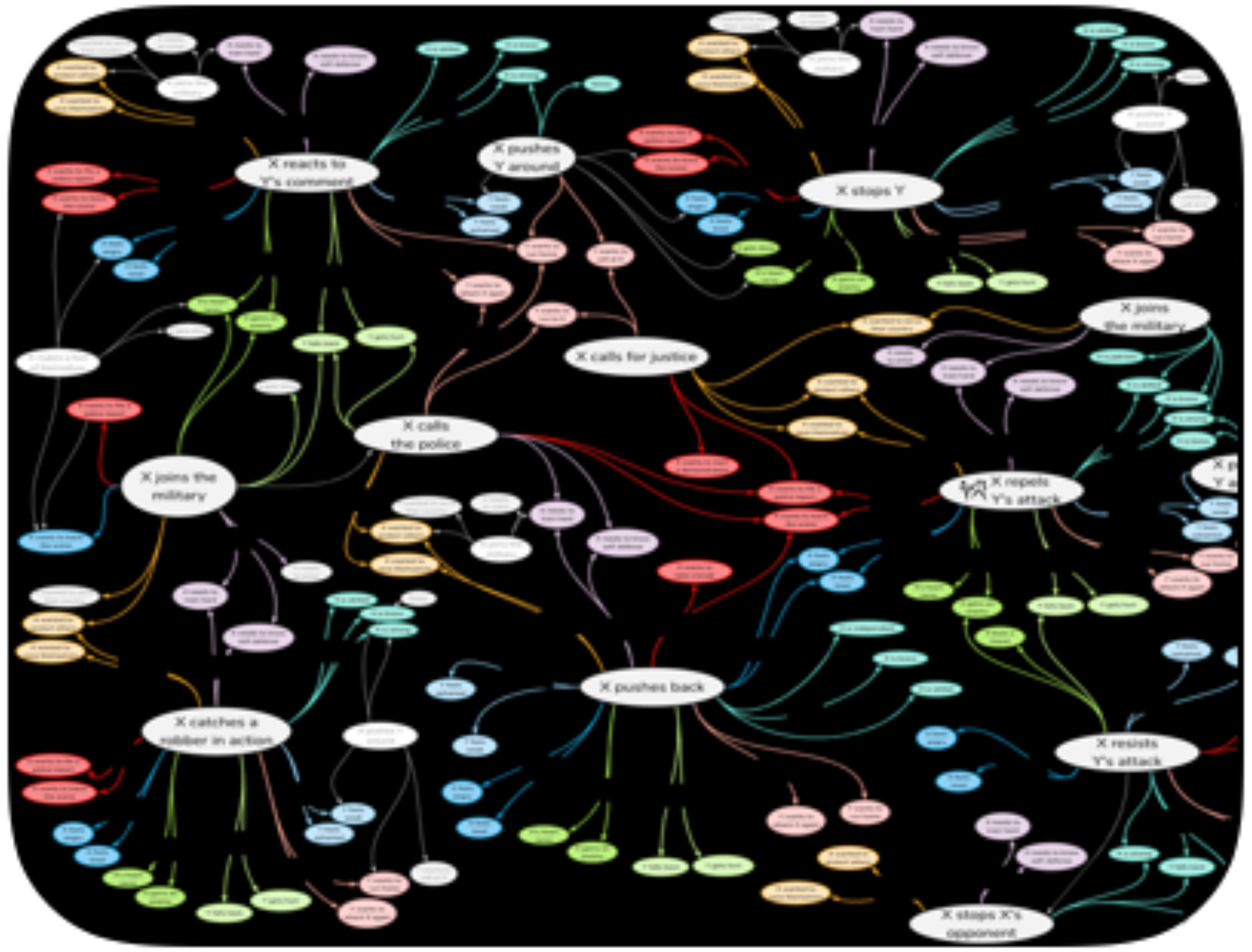
Switch Transformer ~1 T (Jan 2021)

GPT3 175B (July 2020)



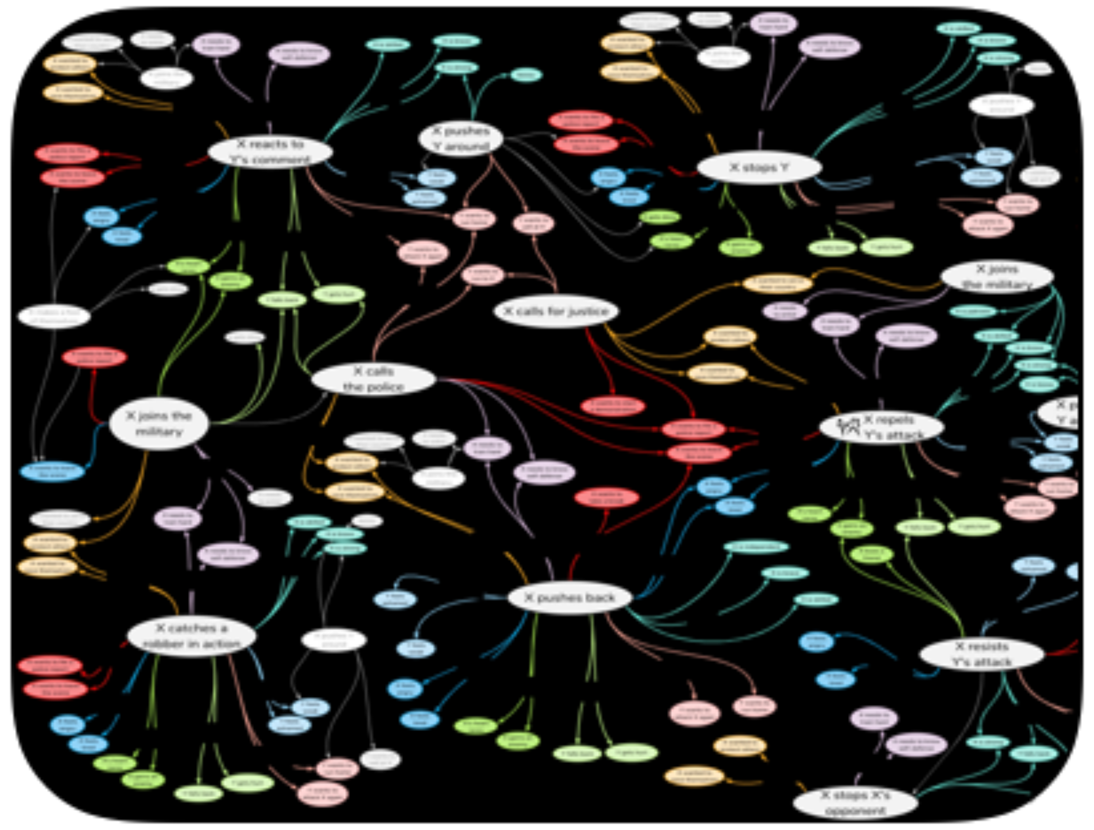
Limitations of Symbolic CSKGs

- Insufficient Coverage
- Limited expressivity
- No Contextualization



Limitations of Symbolic CSKGs

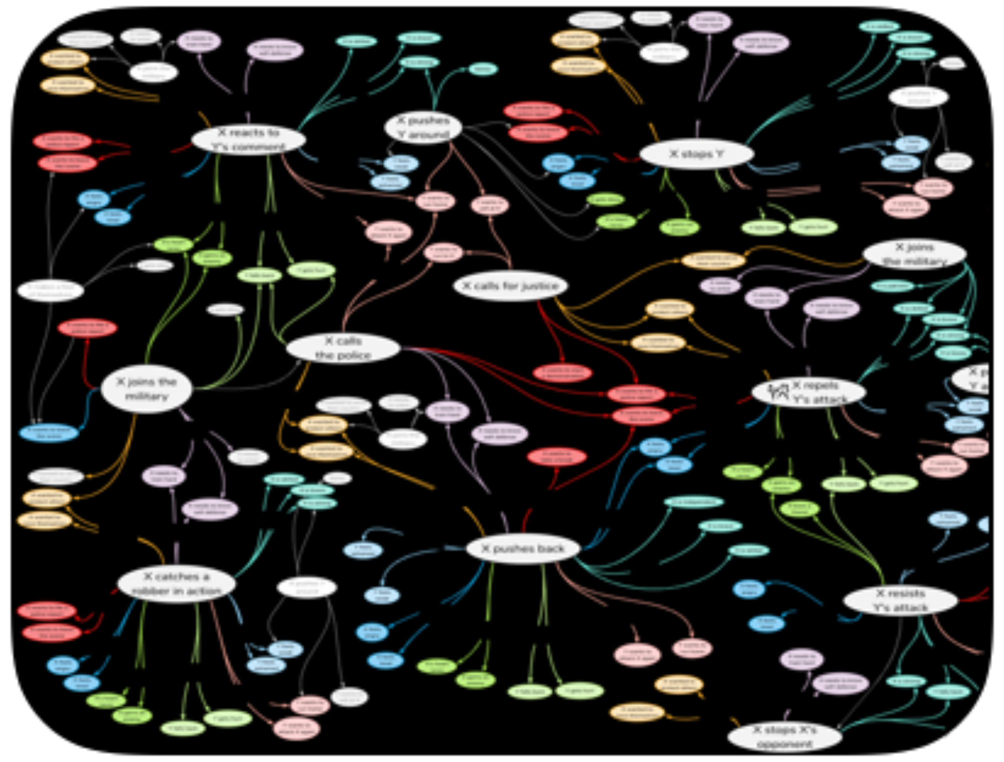
Kai knew that things were getting out of control and managed to keep his temper in check



Limitations of Symbolic CSKGs

- Situations rarely found **as-is** in commonsense knowledge graphs

ATOMIC



(Sap et al., 2019)

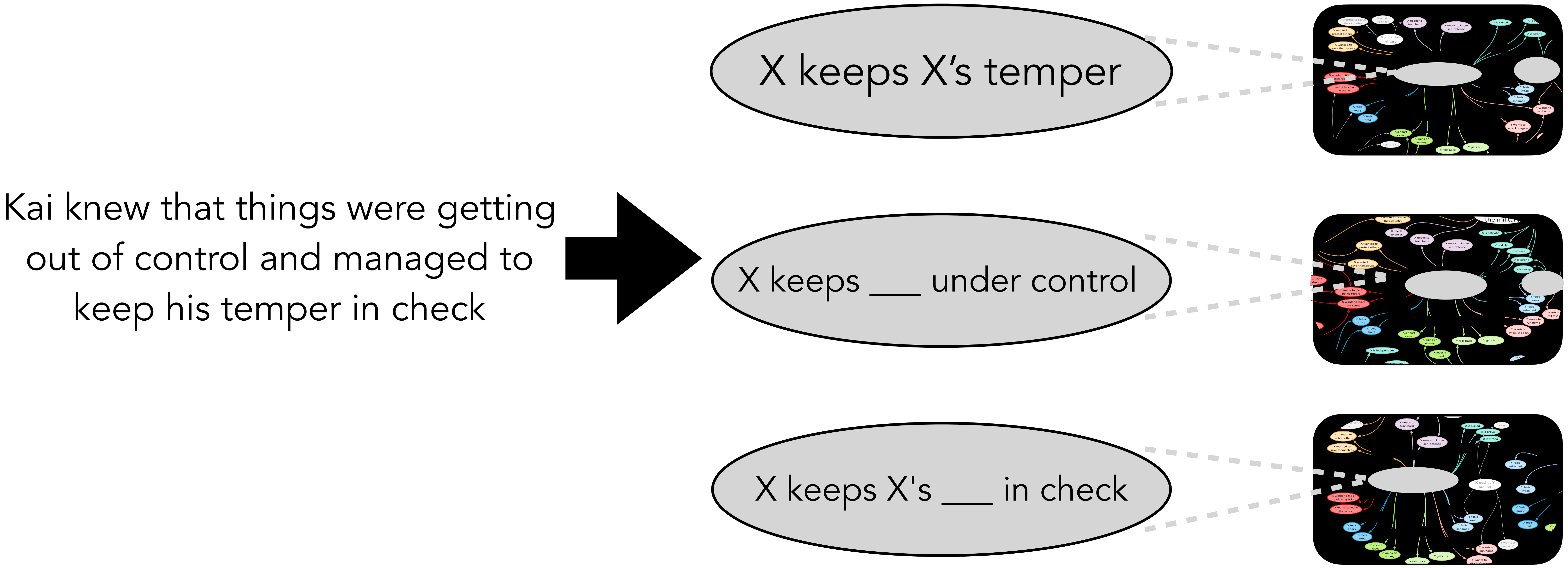
(X goes to the mall,
Effect on X, buys clothes)

(X goes the mall,
Perception of X, rich)

(X gives Y some money,
Reaction of Y, grateful)

Limitations of Symbolic CSKGs

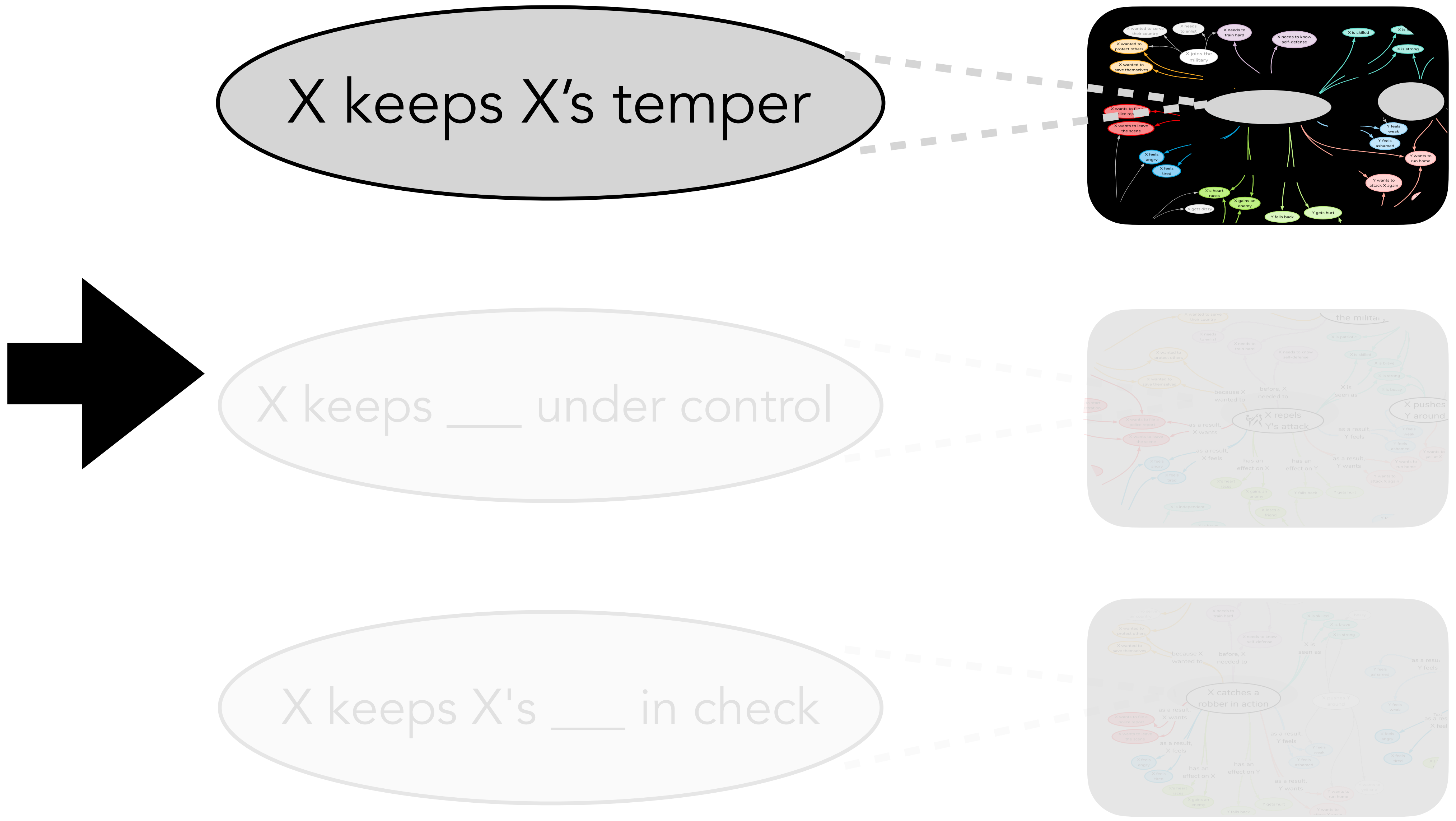
- Situations rarely found **as-is** in commonsense knowledge graphs
- Connecting to knowledge graphs can yield **incorrect** nodes



Limitations of Symbolic CSKGs

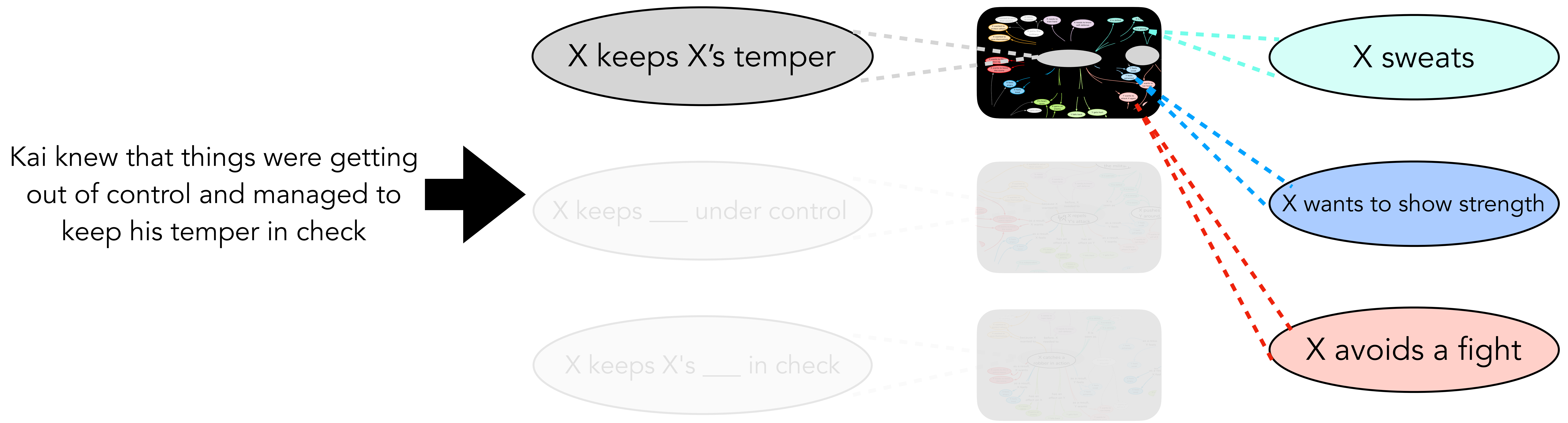
- Situations rarely found **as-is** in commonsense knowledge graphs
- Connecting to knowledge graphs can yield **incorrect** nodes

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



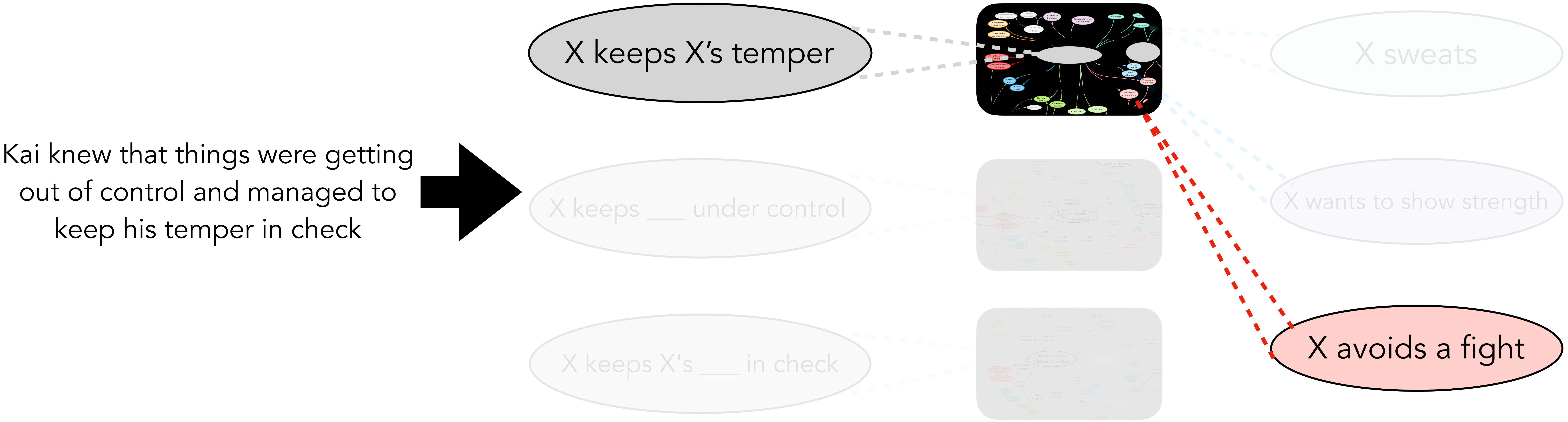
Limitations of Symbolic CSKGs

- Situations rarely found **as-is** in commonsense knowledge graphs
- Connecting to knowledge graphs can yield **incorrect** nodes
- Suitable nodes are often **uncontextualized**



Limitations of Symbolic CSKGs

- Situations rarely found **as-is** in commonsense knowledge graphs
- Connecting to knowledge graphs can yield **incorrect** nodes
- Suitable nodes are often **uncontextualized**



Challenge

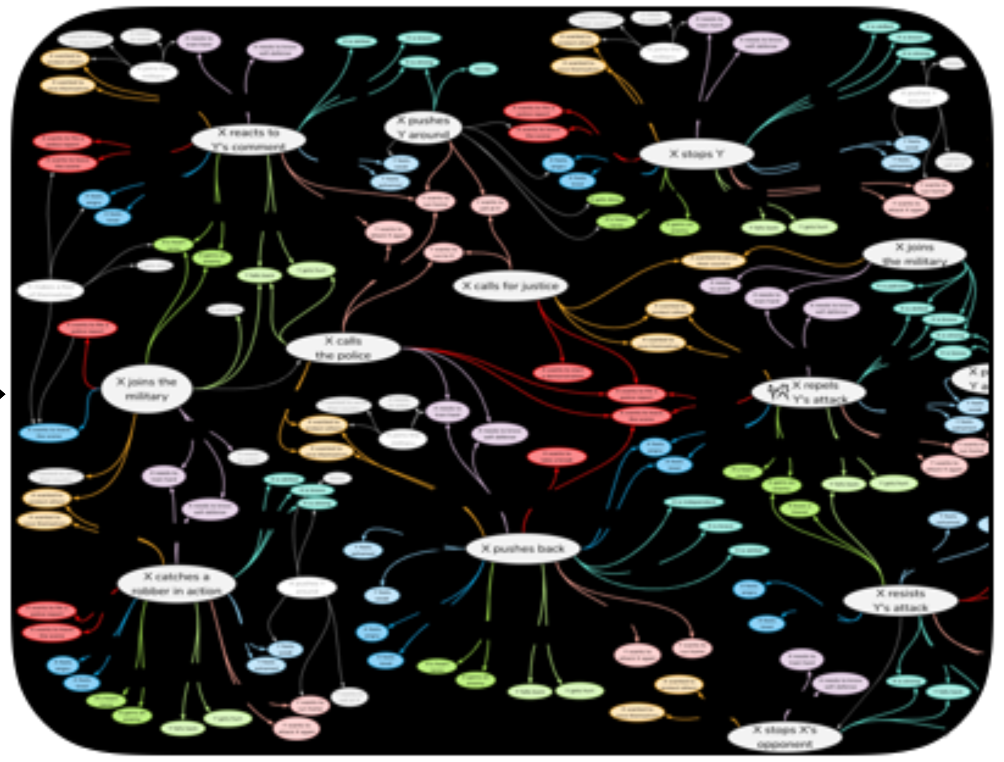
How do we provide machines with large-scale commonsense knowledge?

Constructing Knowledge Graphs

Observe world

Write commonsense knowledge facts

Store facts in knowledge graph



(Miller, 1995)

(Singh et al., 2002)

(Lenat, 1995)

(Sap et al., 2019)

Challenges

- Commonsense knowledge is **immeasurably vast**, making it **impossible to manually enumerate**

Extracting Knowledge Graphs from Text

Gather Textual Corpus

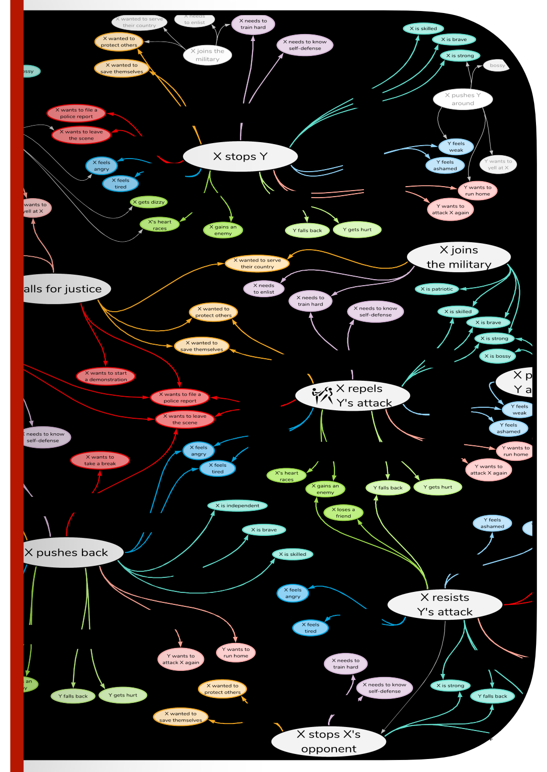
John went to the g
store to buy some
He was going to pr
dinner for his **daug**
birthday. She was
5 and would be sta
elementary **school**

Automatically extract knowledge

(Banko et al., 2007)

(Zhang et al., 2020)

Store in knowledge graph



ConceptNet
An open, multilingual knowledge graph

(Speer et al., 2017)

Webchild

(Tandon et al., 2019)

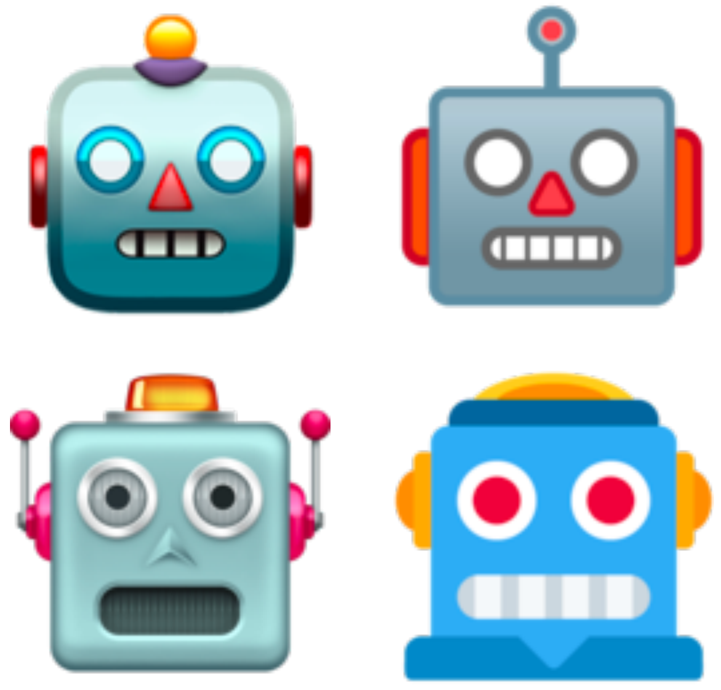
Underspecified Reporting Bias

Learning Relations from Existing KGs

Gather training set of knowledge tuples

(person, CapableOf, buy)

Learn relationships among entities

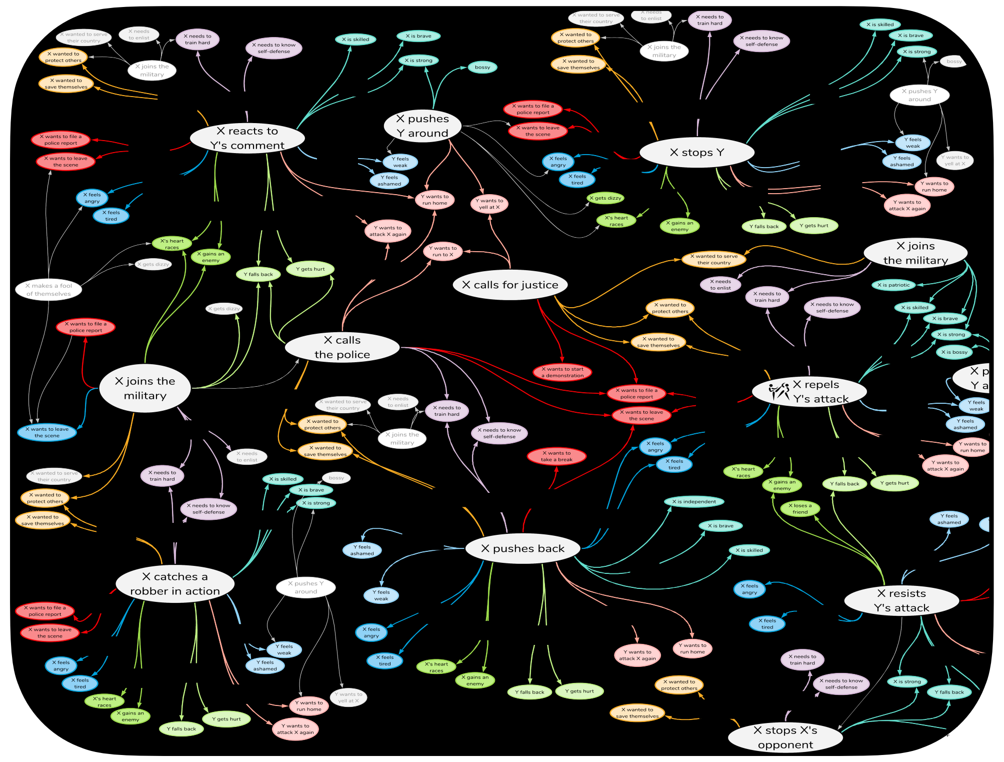


(Socher et al., 2013)
(Bordes et al., 2013)
(Riedel et al., 2013)
(Toutanova et al., 2015)
(Yang et al., 2015)
(Trouillon et al., 2016)
(Nguyen et al., 2016)
(Dettmers et al., 2018)

Predict new relationships

(person, CapableOf, ?)

Store in knowledge graph



Commonsense KGs are Different

Knowledge Graph	# Entities	# Edges	Average Fan-in
ConceptNet - 100k	78088	100000	**1.25**
ATOMIC	256570	610536	**2.25**
FB15k-237	14505	272115	**16.98**

Knowledge base completion assumes explicit connectivity

Malaviya et al., AAAI 2020

Commonsense KGs are Different

Knowledge Graph	# Entities	# Edges	Average Fan-in
ConceptNet - 100k	78088	100000	**1.25**
ATOMIC	256570	610536	**2.25**
FB15k-237	14505	272115	**16.98**

Knowledge base completion assumes explicit connectivity

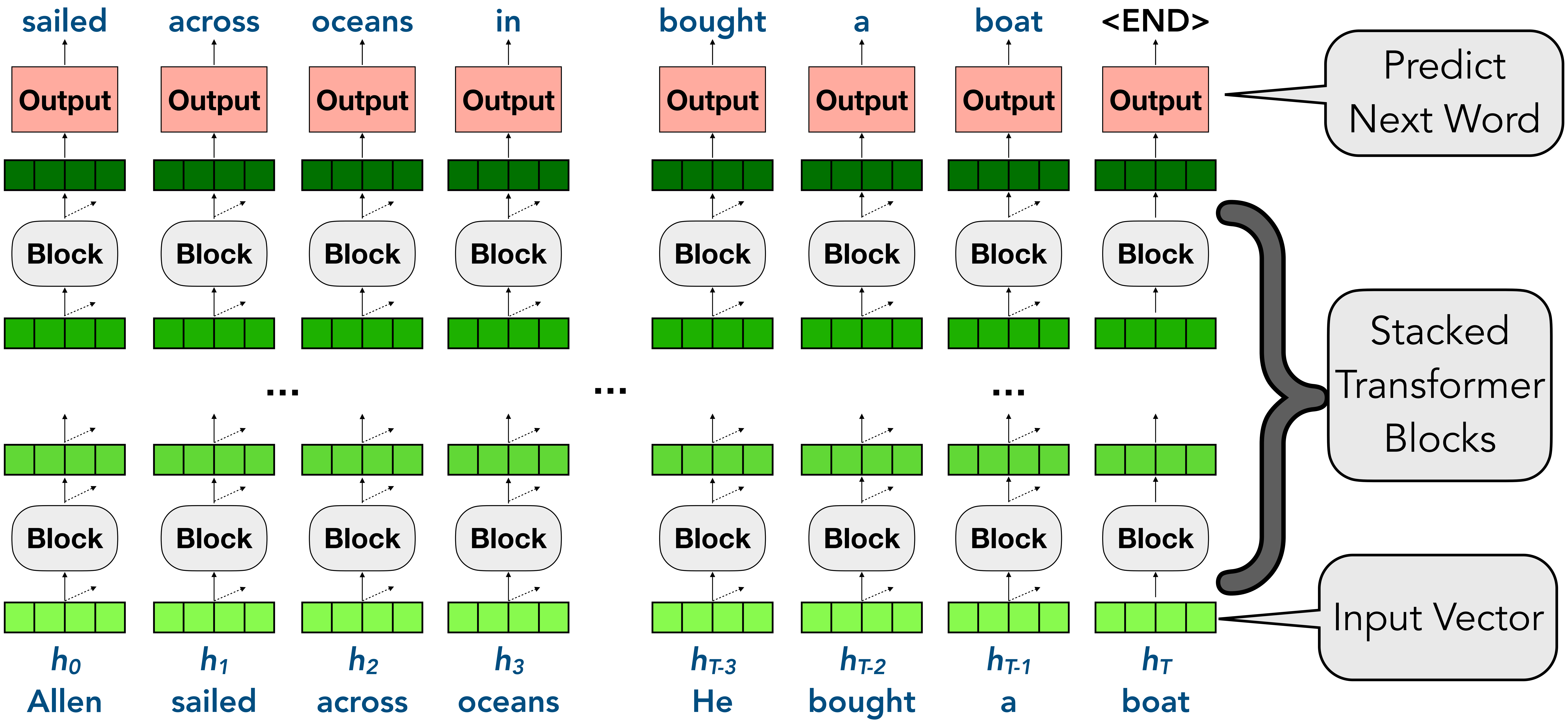
Malaviya et al., AAAI 2020

Challenges

- Commonsense knowledge is **immeasurably vast**, making it **impossible to manually enumerate**
- Commo … fied, and therefo …
- Commonsense knowledge resources are quite sparse, making them **difficult to extend by only learning from examples**

How else can we learn commonsense knowledge at scale?

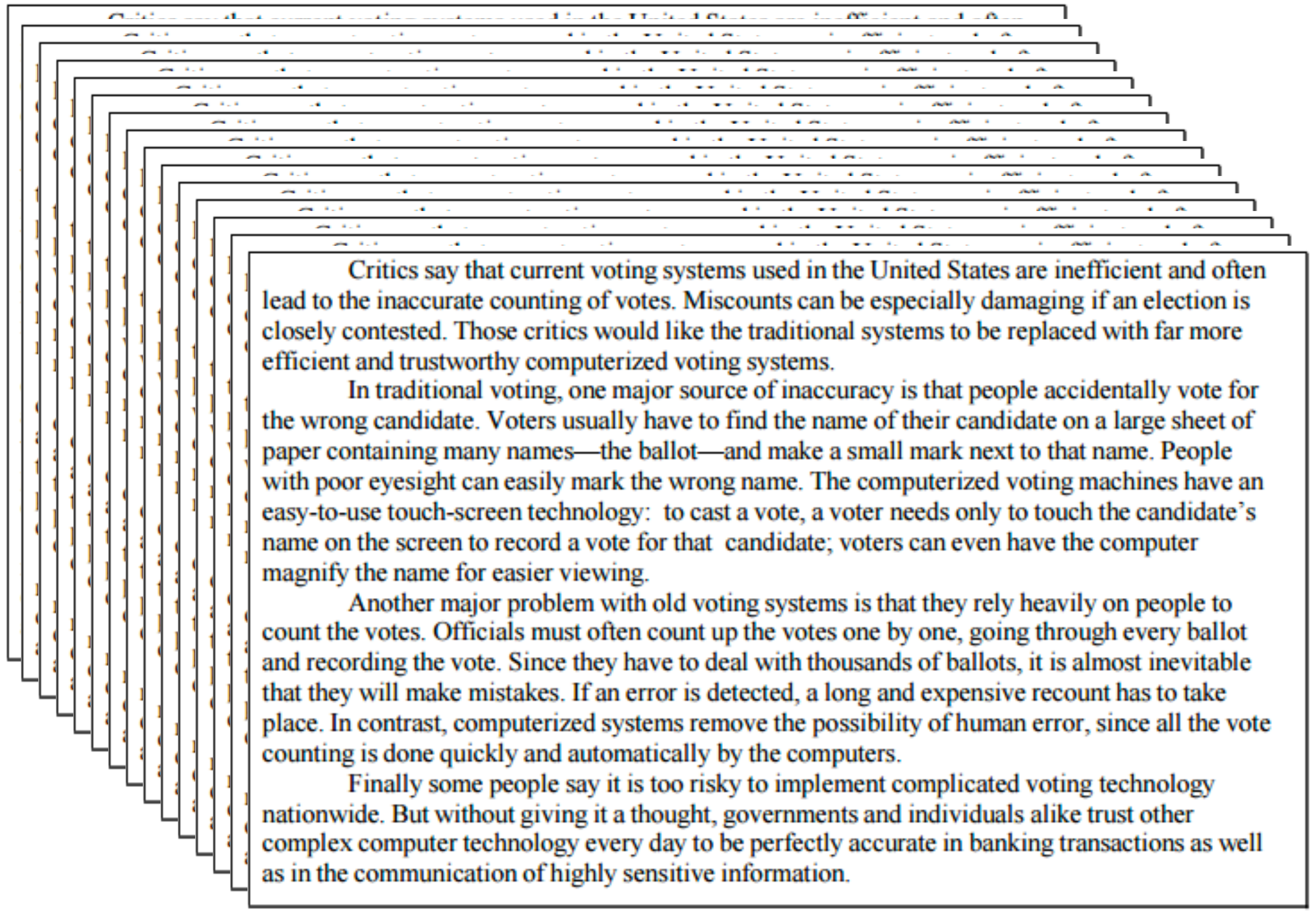
Deep Language Models



(Radford et al., 2018, 2019, many others)

Learning in Language Models

Text Corpus


Critics say that current voting systems used in the United States are inefficient and often lead to the inaccurate counting of votes. Miscounts can be especially damaging if an election is closely contested. Those critics would like the traditional systems to be replaced with far more efficient and trustworthy computerized voting systems.

In traditional voting, one major source of inaccuracy is that people accidentally vote for the wrong candidate. Voters usually have to find the name of their candidate on a large sheet of paper containing many names—the ballot—and make a small mark next to that name. People with poor eyesight can easily mark the wrong name. The computerized voting machines have an easy-to-use touch-screen technology: to cast a vote, a voter needs only to touch the candidate's name on the screen to record a vote for that candidate; voters can even have the computer magnify the name for easier viewing.

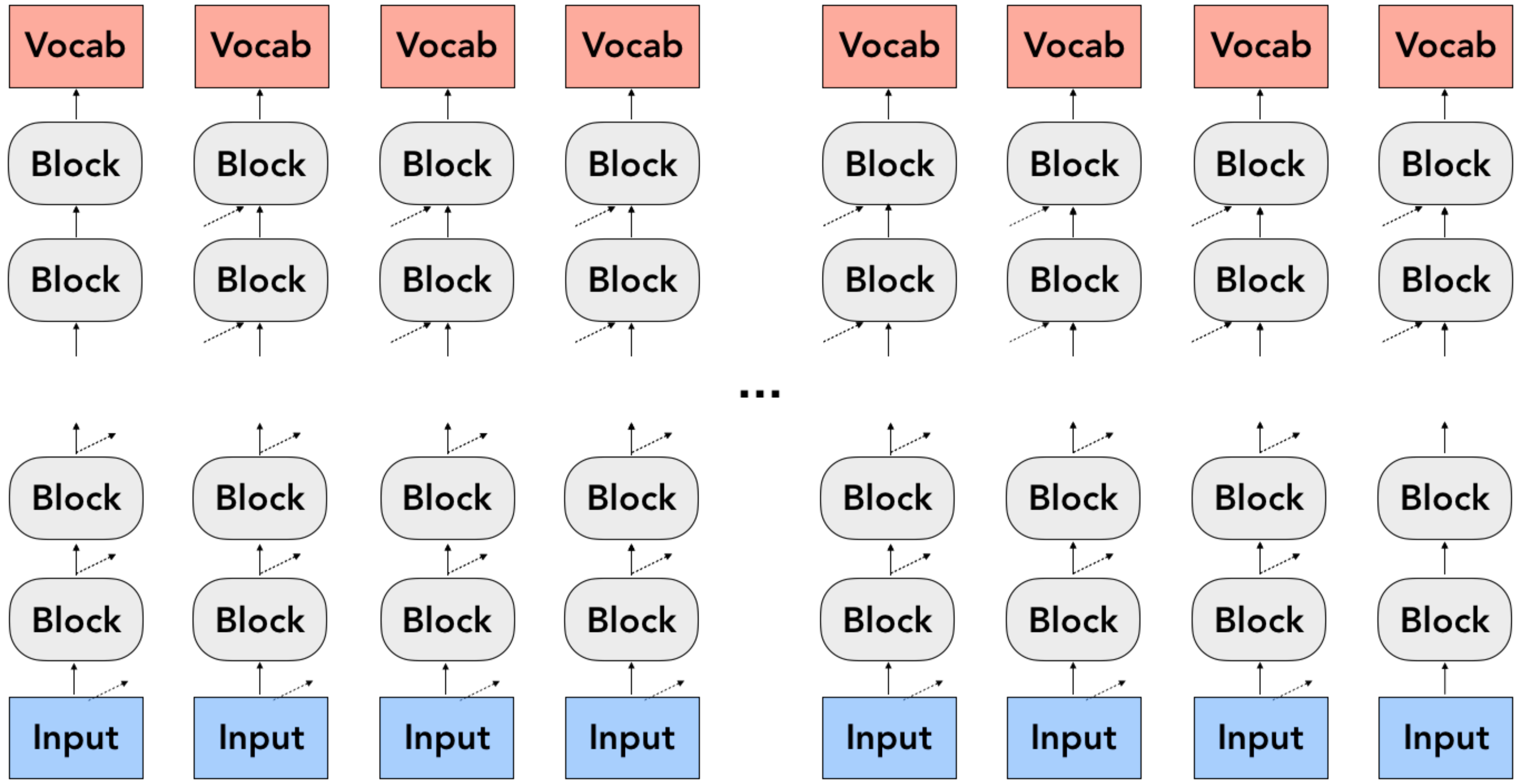
Another major problem with old voting systems is that they rely heavily on people to count the votes. Officials must often count up the votes one by one, going through every ballot and recording the vote. Since they have to deal with thousands of ballots, it is almost inevitable that they will make mistakes. If an error is detected, a long and expensive recount has to take place. In contrast, computerized systems remove the possibility of human error, since all the vote counting is done quickly and automatically by the computers.

Finally some people say it is too risky to implement complicated voting technology nationwide. But without giving it a thought, governments and individuals alike trust other complex computer technology every day to be perfectly accurate in banking transactions as well as in the communication of highly sensitive information.

Used to

Learn

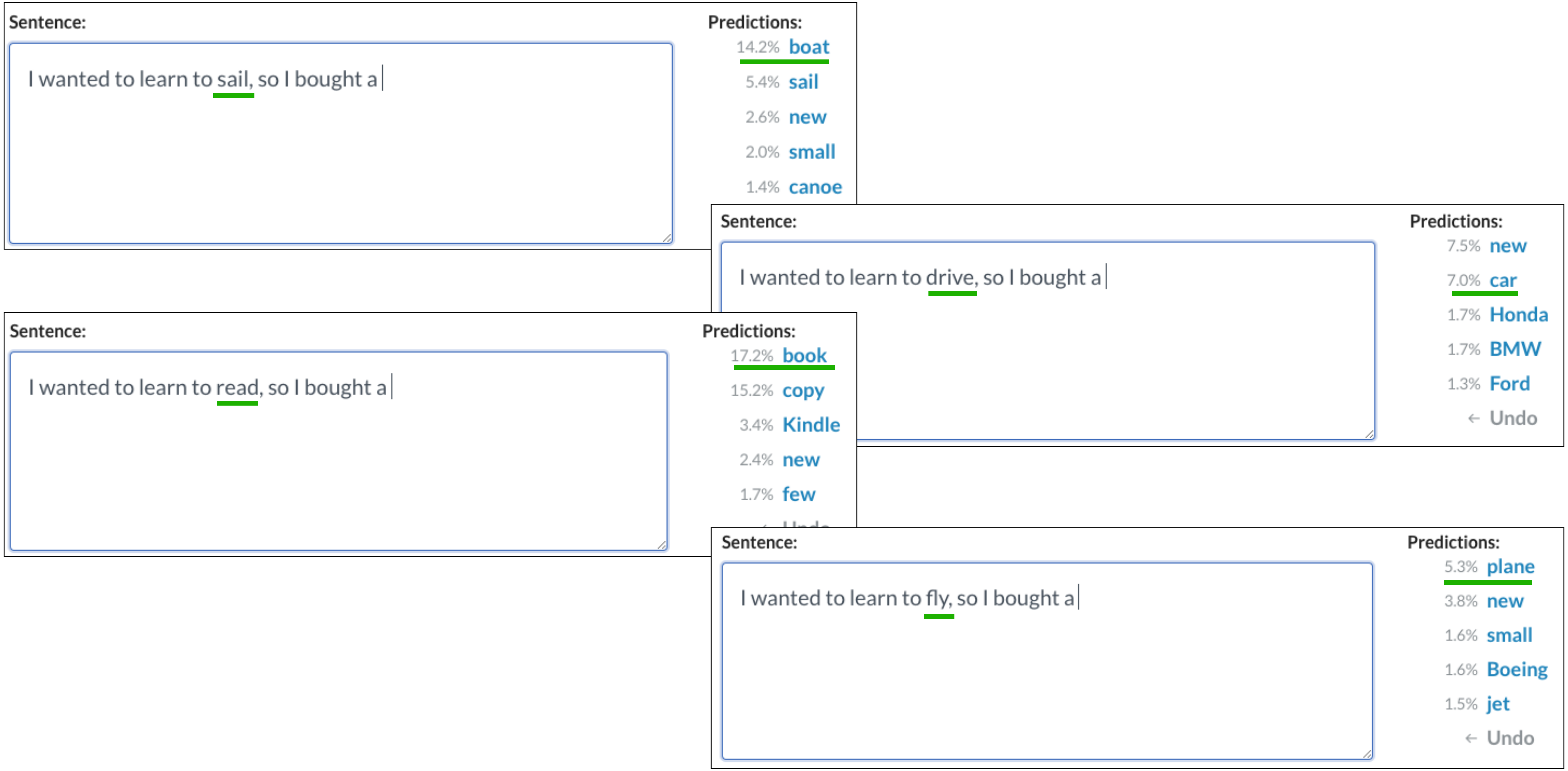
Transformer Language Model



(Radford et al., 2018, 2019, many others)

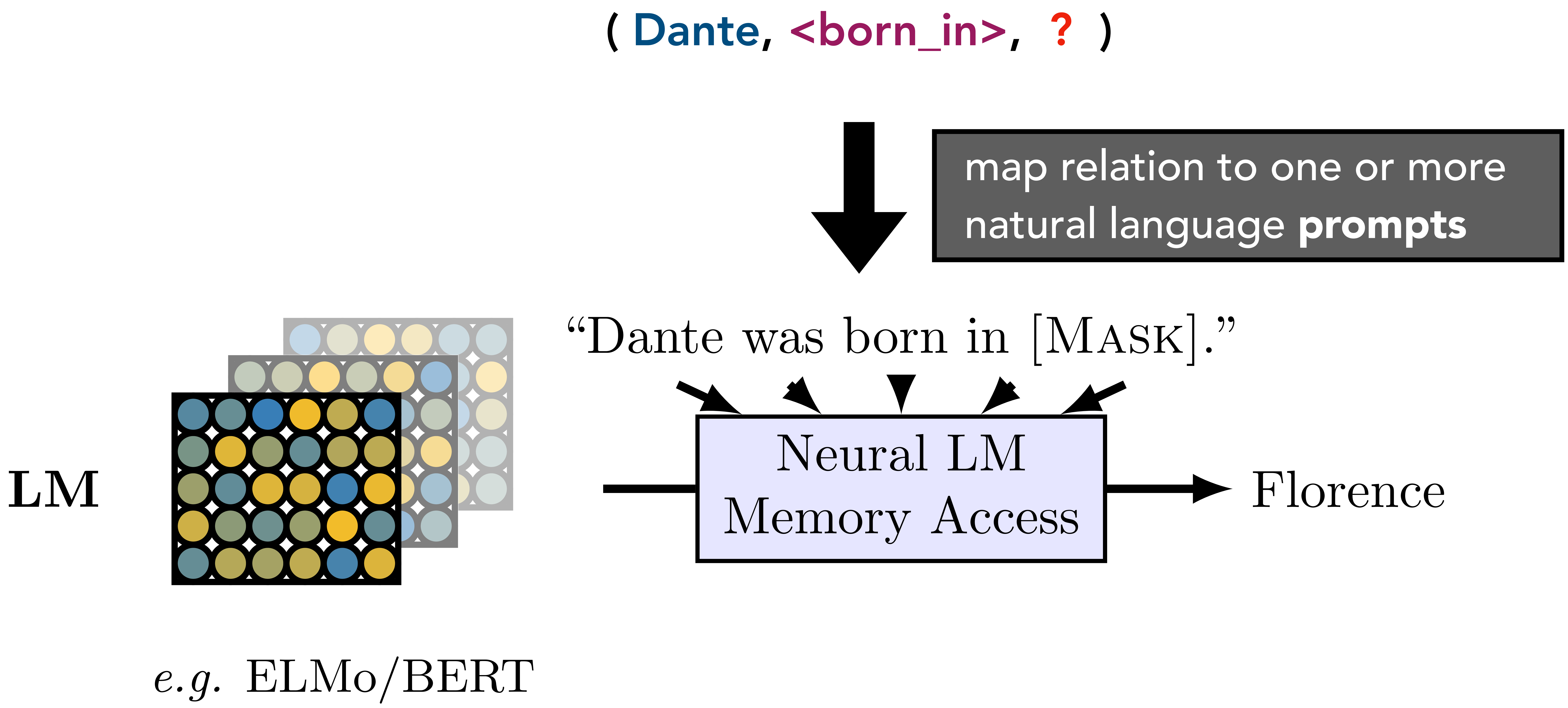
Do language models have commonsense knowledge?

Knowledge in Language Models



https://demo.allennlp.org/next-token-lm

Knowledge Prompting



Petroni et al., EMNLP 2019

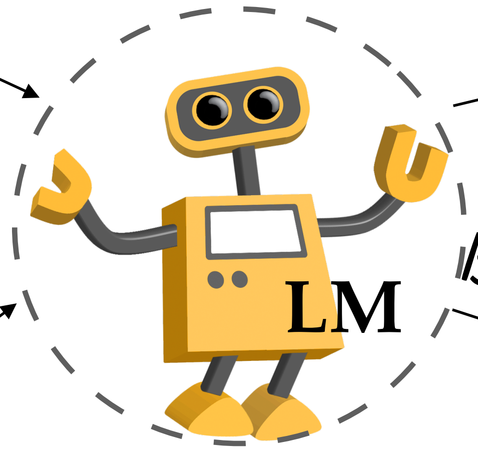
Knowledge Prompting

Instance:

Christmas was a special holiday to Eric but not Adam since ____ was a Jew.

Question Generation:

Christmas was a special holiday to Eric but not Adam since ____ was a Jew.



Christmas?

the celebration of the birth of Christ.

Clarification Question

What is the definition of Christmas?

Answer Generation:

Christmas

Clarification

The definition of Christmas is the

The purpose of Christmas is to celebrate

The definition of Christmas is the

The definition of Christmas is a Christian Holiday.

What is the definition of

The definition of ________ is

Question & Answer Prefixes

Shwartz et al., EMNLP 2021

Prompt Sensitivity

Prompts	
manual	`DirectX` *is developed by* y_{man}
mined	y_{mine} *released the* `DirectX`
paraphrased	`DirectX` *is created by* y_{para}

Top 5 predictions and log probabilities

	y_{man}		y_{mine}		y_{para}	
1	`Intel`	-1.06	<u>`Microsoft`</u>	-1.77	<u>`Microsoft`</u>	-2.23
2	<u>`Microsoft`</u>	-2.21	`They`	-2.43	`Intel`	-2.30
3	`IBM`	-2.76	`It`	-2.80	`default`	-2.96
4	`Google`	-3.40	`Sega`	-3.01	`Apple`	-3.44
5	`Nokia`	-3.58	`Sony`	-3.19	`Google`	-3.45

Jiang et al., TACL 2020

Prompt Sensitivity

	Prompts
manual	`DirectX` *is developed by* y_{man}
mined	y_{mine} *released the* `DirectX`
paraphrased	`DirectX` *is created by* y_{para}

Top 5 predictions and log probabilities

	y_{man}		y_{mine}		y_{para}	
1	`Intel`	-1.06	`Microsoft`	-1.77	`Microsoft`	-2.23
2	`Microsoft`	-2.21	`They`	-2.43	`Intel`	-2.30
3	`IBM`	-2.76	`It`	-2.80	`default`	-2.96
4	`Google`	-3.40	`Sega`	-3.01	`Apple`	-3.44
5	`Nokia`	-3.58	`Sony`	-3.19	`Google`	-3.45

Jiang et al., TACL 2020

Candidate Sentence S_i	$\log p(S_i)$
"musician can playing musical instrument"	-5.7
"musician can be play musical instrument"	-4.9
"musician often play musical instrument"	-5.5
"a musician can play a musical instrument"	$\mathbf{-2.9}$

Feldman et al., EMNLP 2019

Context Sensitivity

Prompts

manual	`DirectX` *is developed by* y_{man}
mined	y_{mine} *released the* `DirectX`
paraphrased	`DirectX` *is created by* y_{para}

Top 5 predictions and log probabilities

	y_{man}		y_{mine}		y_{para}	
1	`Intel`	-1.06	`Microsoft`	-1.77	`Microsoft`	-2.23
2	`Microsoft`	-2.21	`They`	-2.43	`Intel`	-2.30
3	`IBM`	-2.76	`It`	-2.80	`default`	-2.96
4	`Google`	-3.40	`Sega`	-3.01	`Apple`	-3.44
5	`Nokia`	-3.58	`Sony`	-3.19	`Google`	-3.45

Jiang et al., TACL 2020

Candidate Sentence S_i	$\log p(S_i)$
"musician can playing musical instrument"	-5.7
"musician can be play musical instrument"	-4.9
"musician often play musical instrument"	-5.5
"a musician can play a musical instrument"	$\mathbf{-2.9}$

Feldman et al., EMNLP 2019

Prompt	**Model Predictions**
A ___ has fur.	dog, cat, fox, ...
A ___ has fur, is big, and has claws.	cat, **bear**, lion, ...
A ___ has fur, is big, has claws, has teeth, is an animal, eats, is brown, and lives in woods.	**bear**, wolf, cat, ...

Weir et al., CogSci 2020

Do language models have commonsense?

- Distinction between **encoding** commonsense knowledge and **expressing** commonsense knowledge

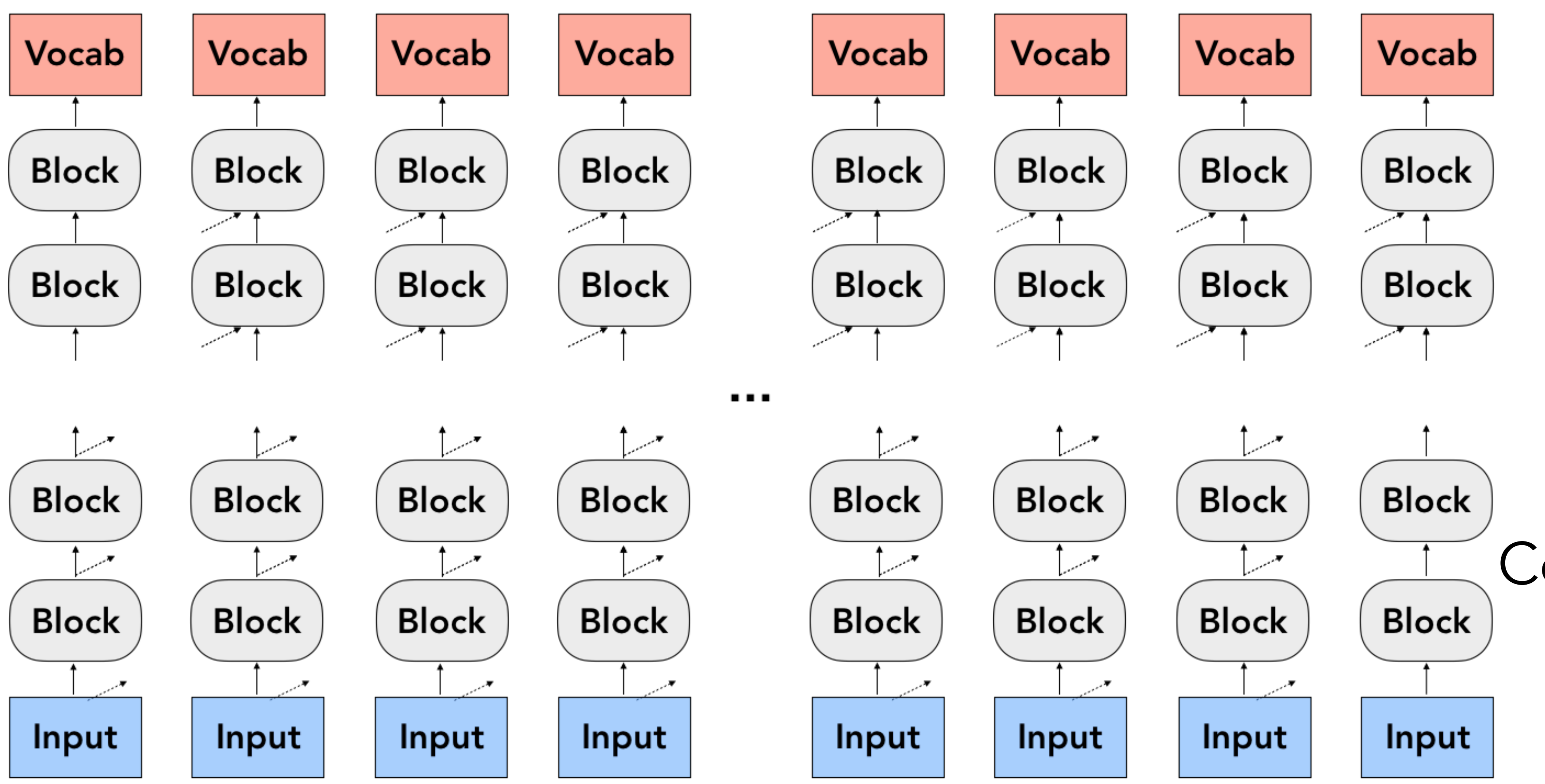
Do language models have commonsense?

- Distinction between **encoding** commonsense knowledge and **expressing** commonsense knowledge
- Probing with prompts measures whether LMs can **express** commonsense knowledge and the results are **mixed**

Do language models encode commonsense knowledge?

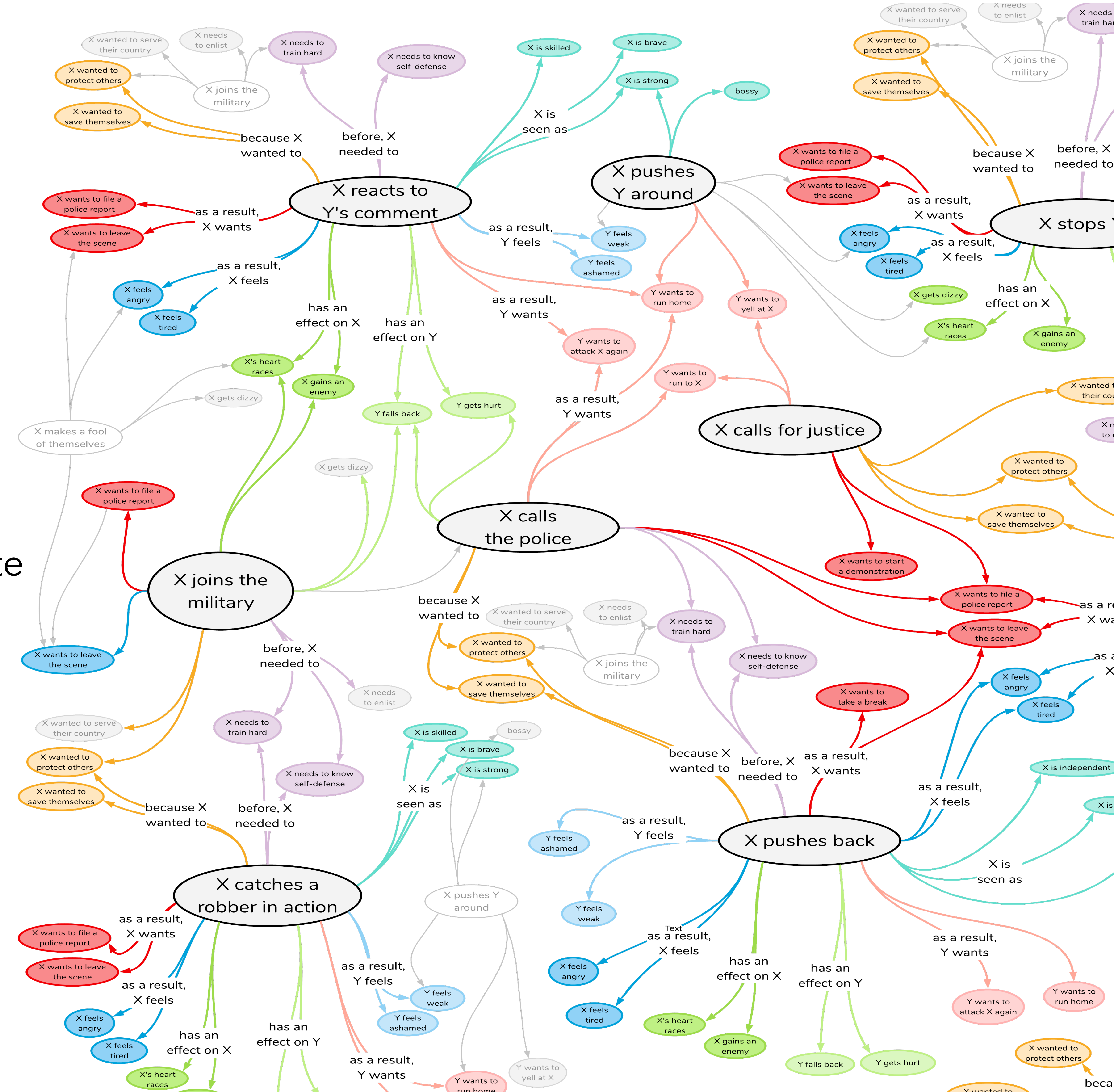
From Unstructured to Structured Knowledge

Transformer Language Model

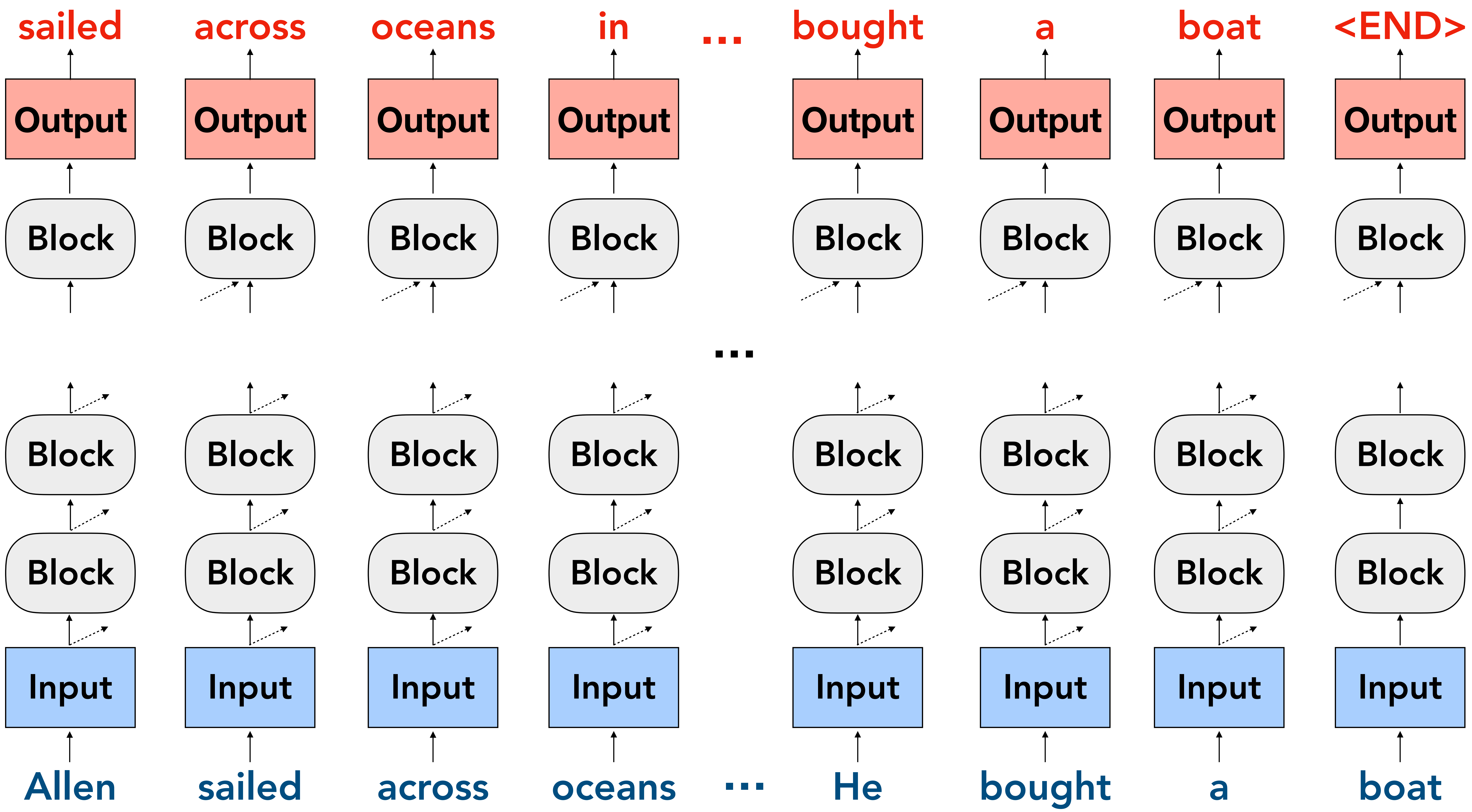


Used to

Communicate



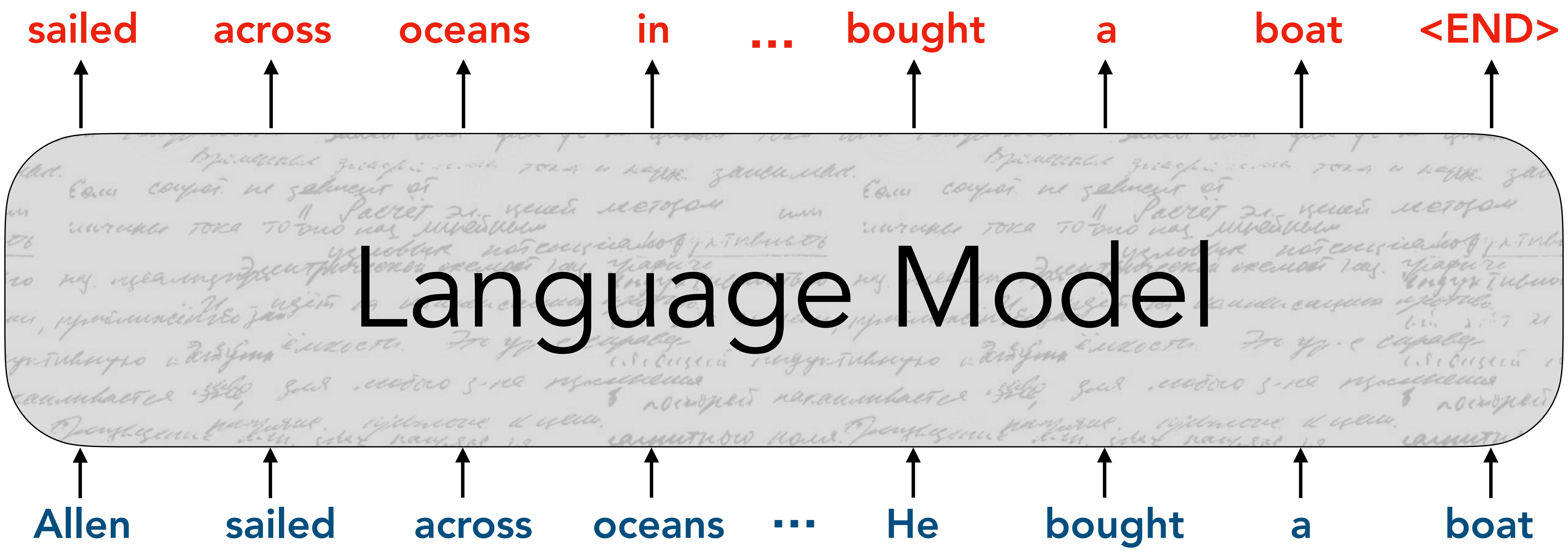
Transformer Language Models



(Radford et al., 2018, 2019, many others)

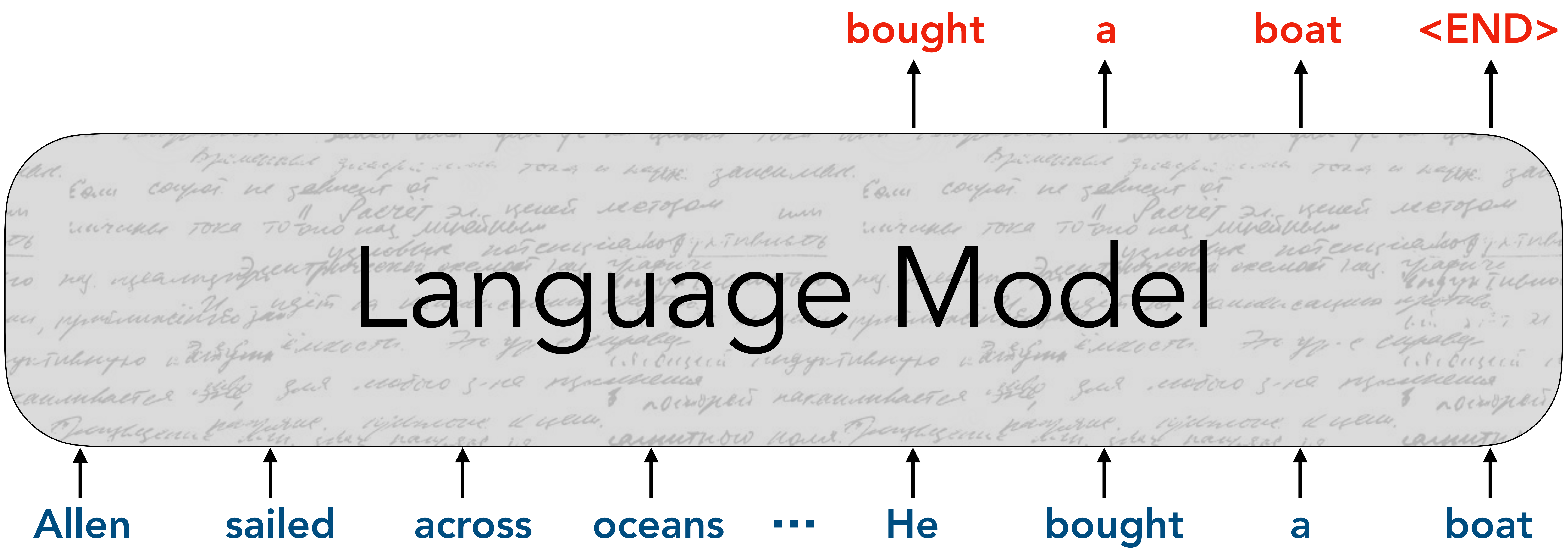
Transformer Language Models

- Trained to generate the next word given a set of preceding words



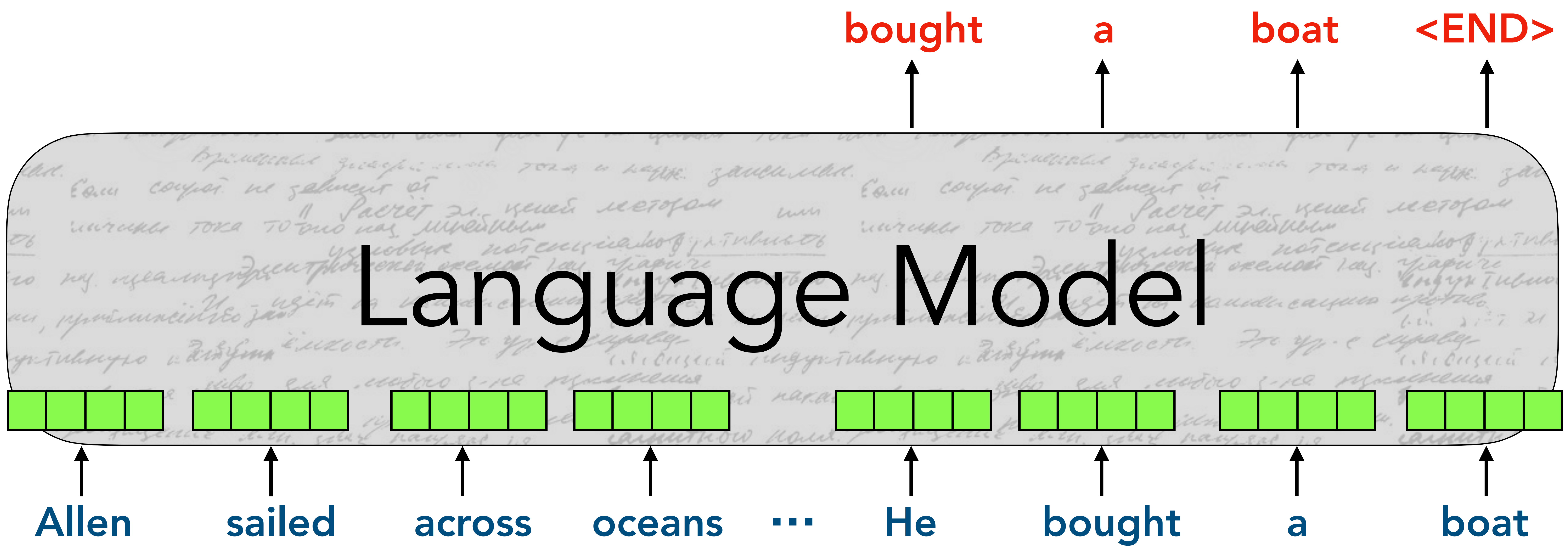
Transformer Language Models

- Trained to generate the next word given a set of preceding words
- Follow-up tokens can be generated using generated tokens as input



Transformer Language Models

- Trained to generate the next word given a set of preceding words
- Follow-up tokens can be generated using generated tokens as input

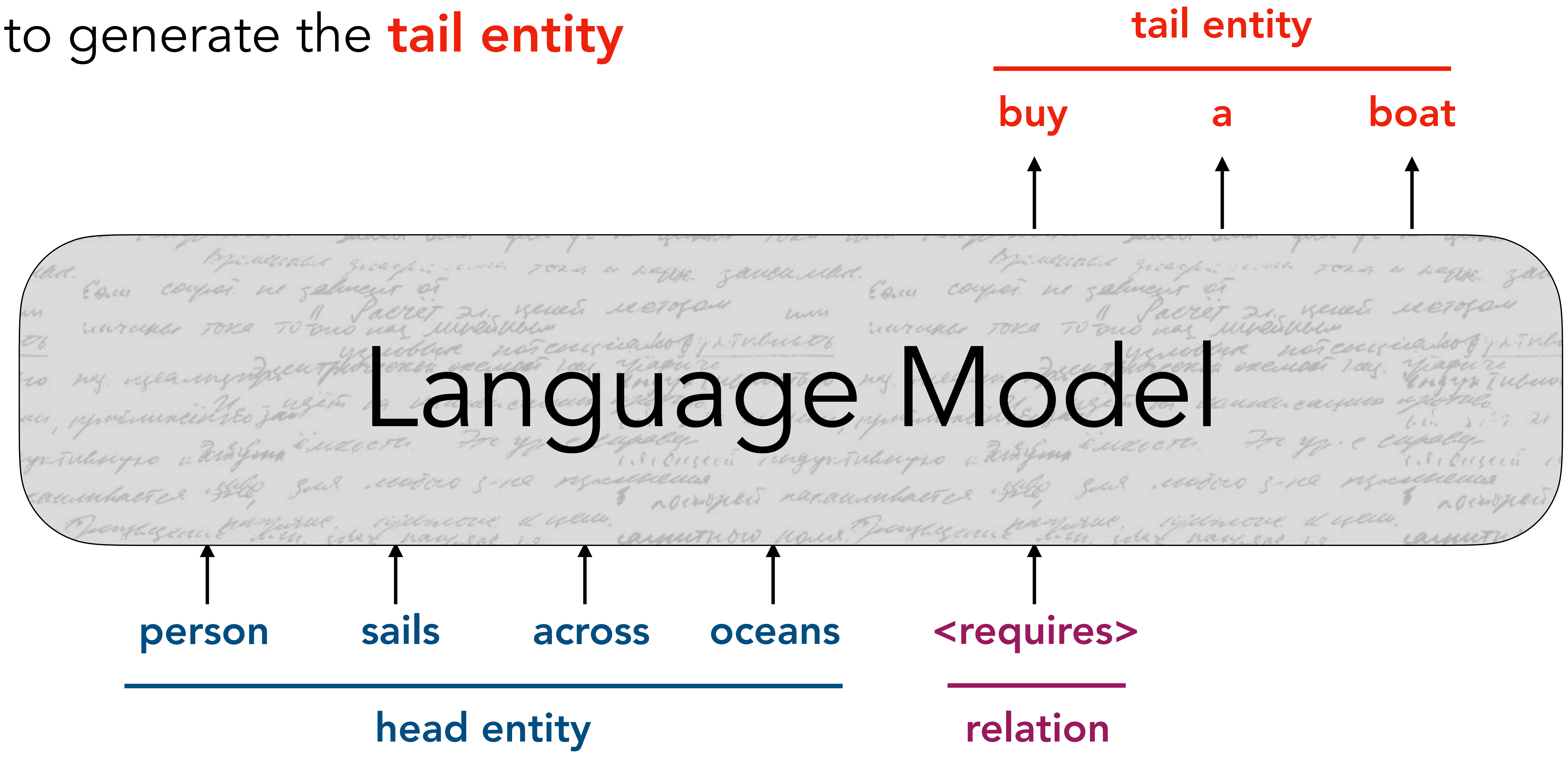


Structure of Knowledge Tuple

Learning Structure of Knowledge

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

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



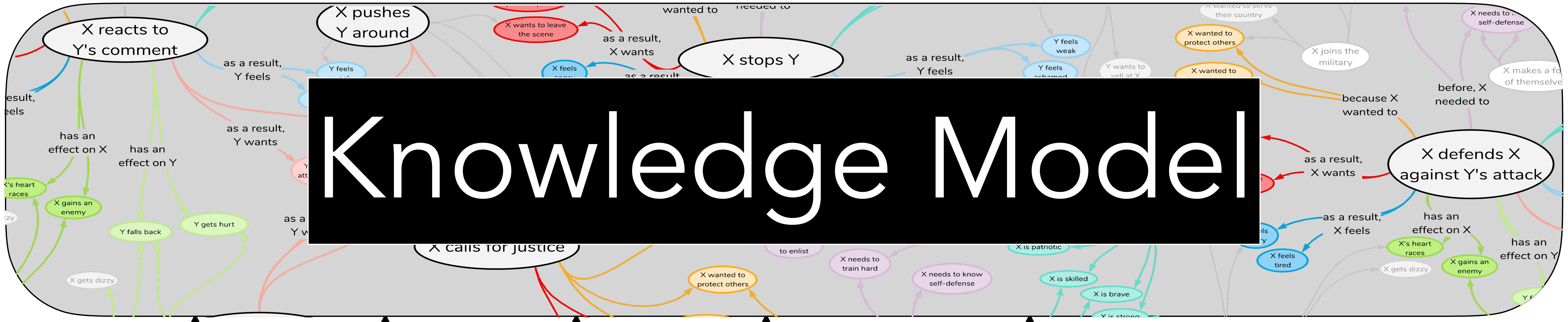
Bosselut et al., ACL 2019

Learning Structure of Knowledge

Language Model ⟶ **Knowledge Model:**
generates knowledge of the structure
of the examples used for training

tail entity

buy a boat



person sails across oceans <requires>

head entity relation

Bosselut et al., ACL 2019

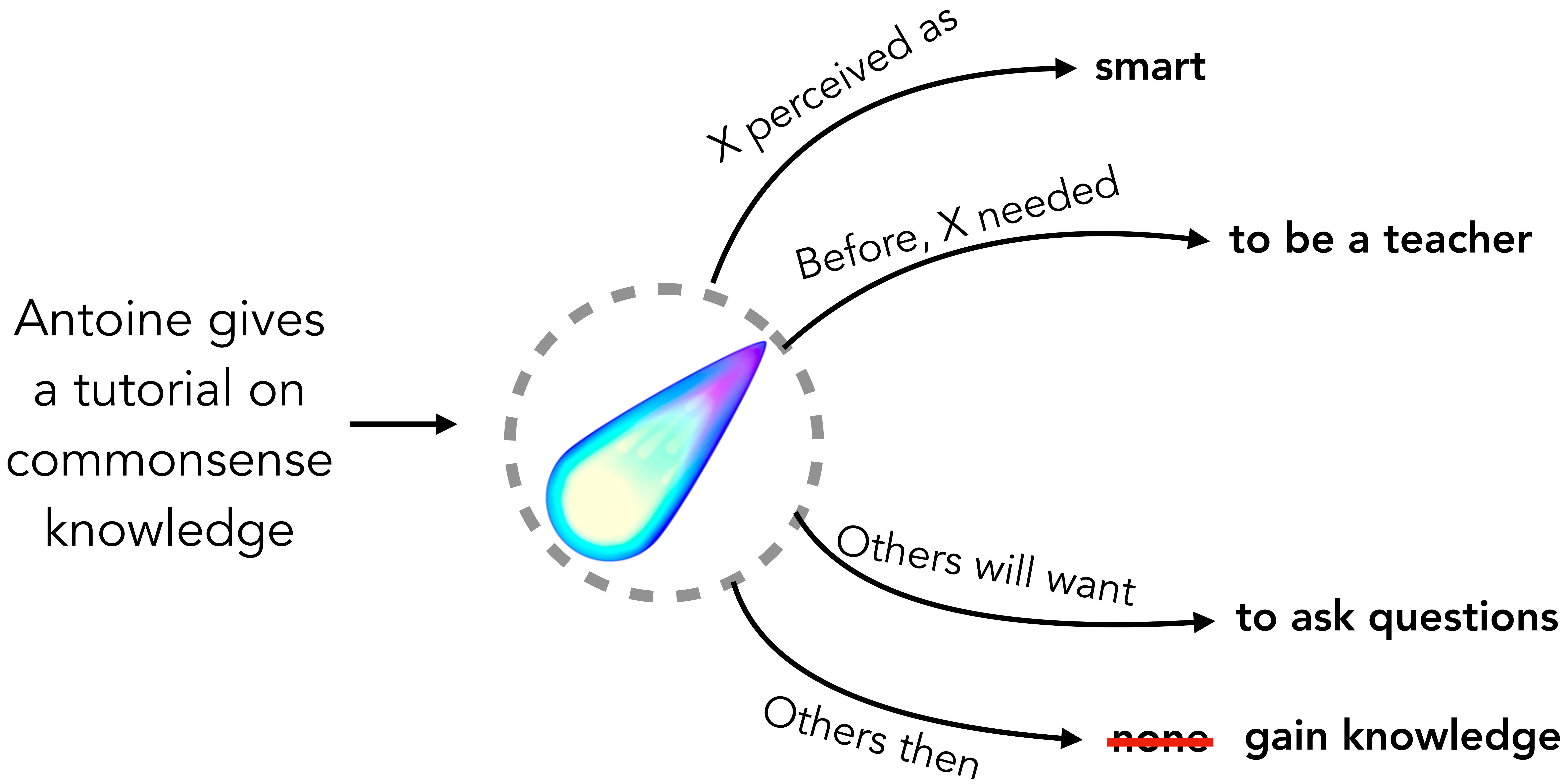
Generate commonsense knowledge for **any** input concept

Generate commonsense knowledge for **any** input concept

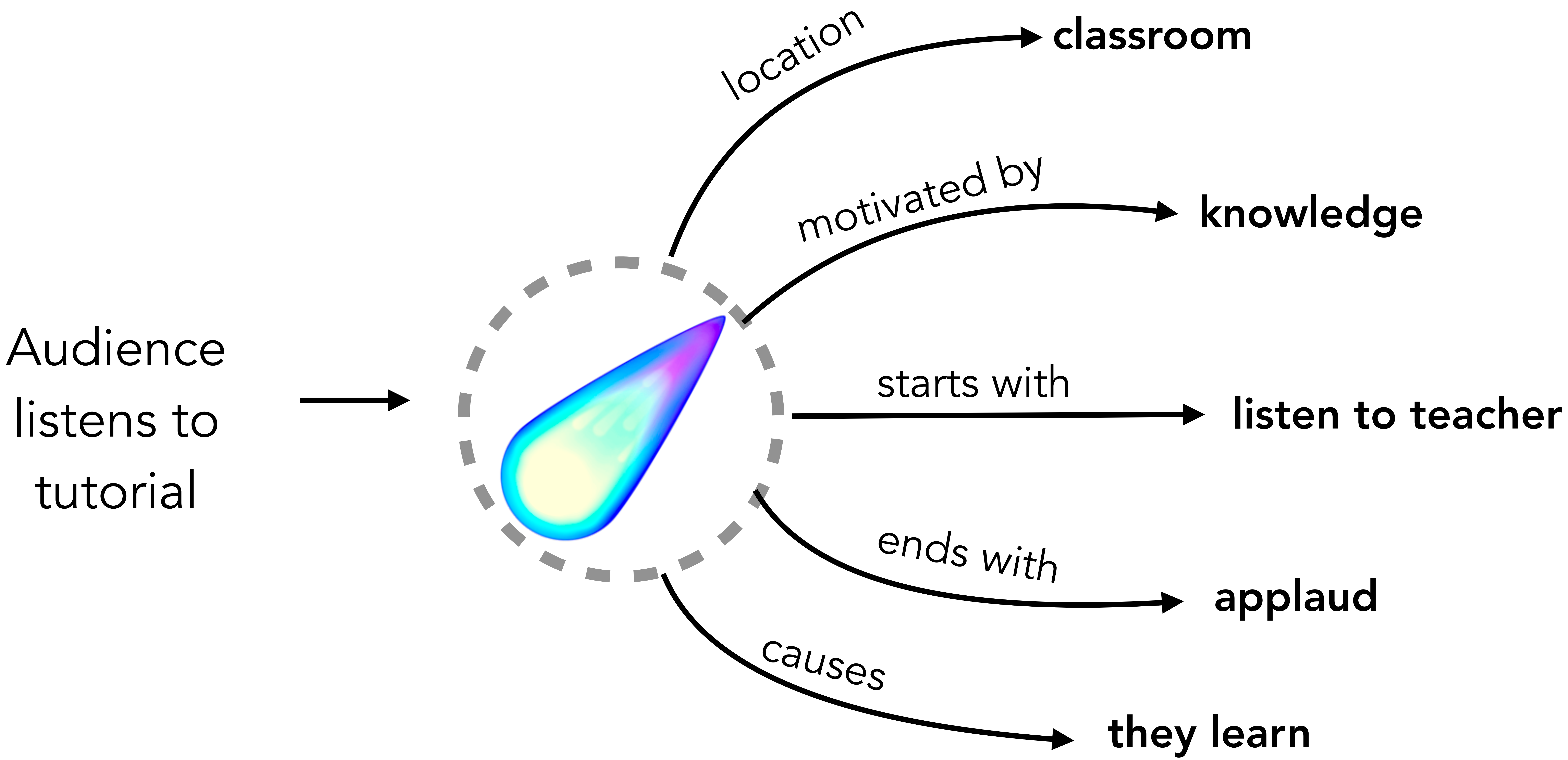
COMmons**E**nse **T**ransformers

Bosselut et al., ACL 2019

COMeT - ATOMIC

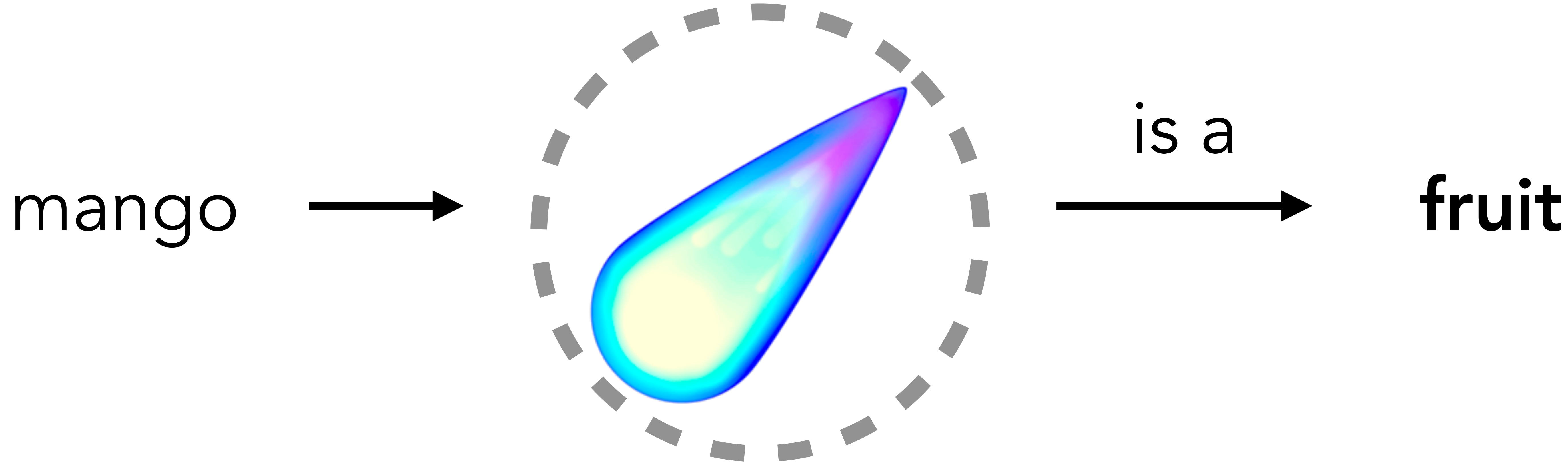


COMET - ConceptNet



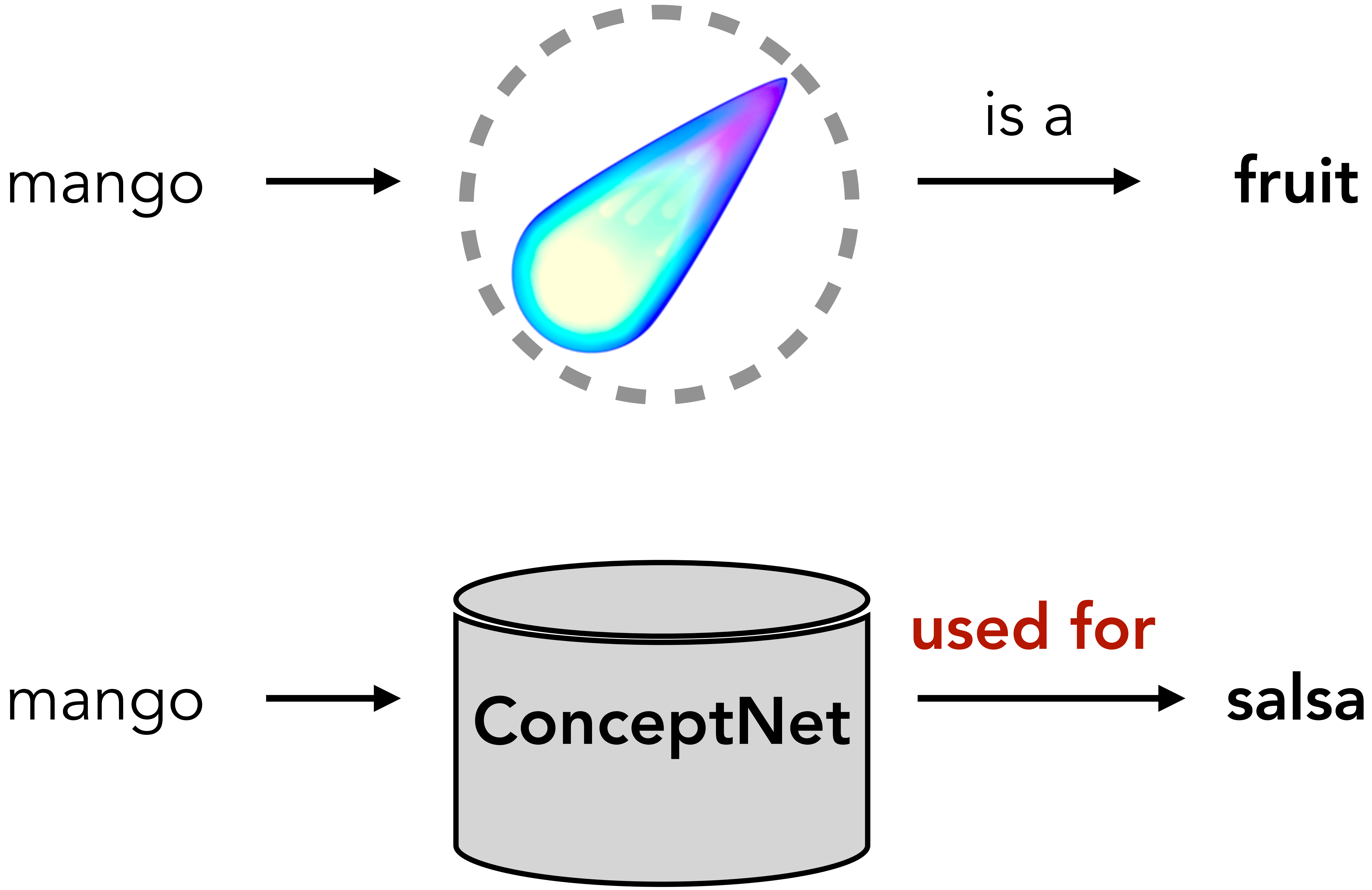
Why does this work?

Transfer Learning from Language



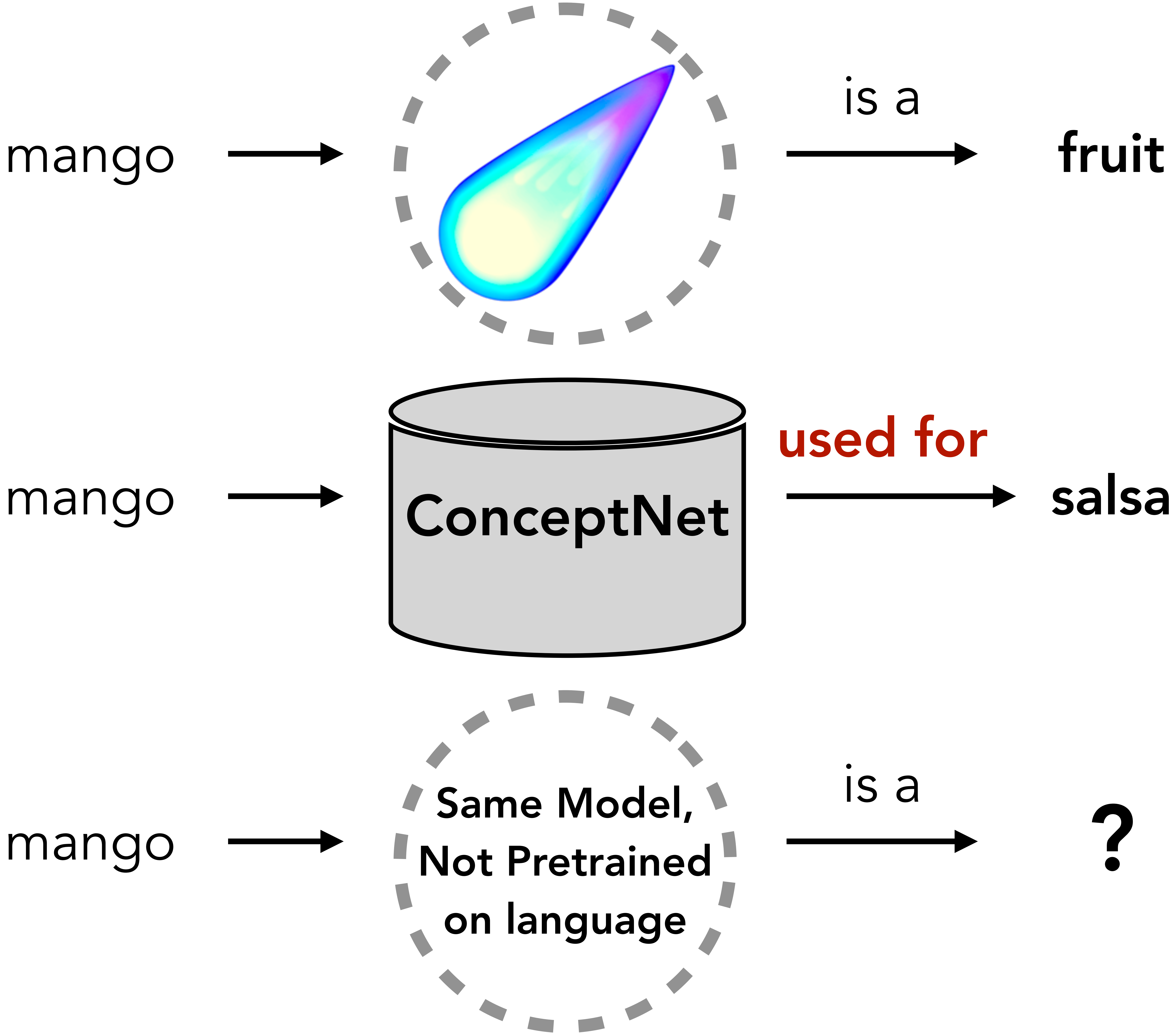
Bosselut et al., ACL 2019

Transfer Learning from Language



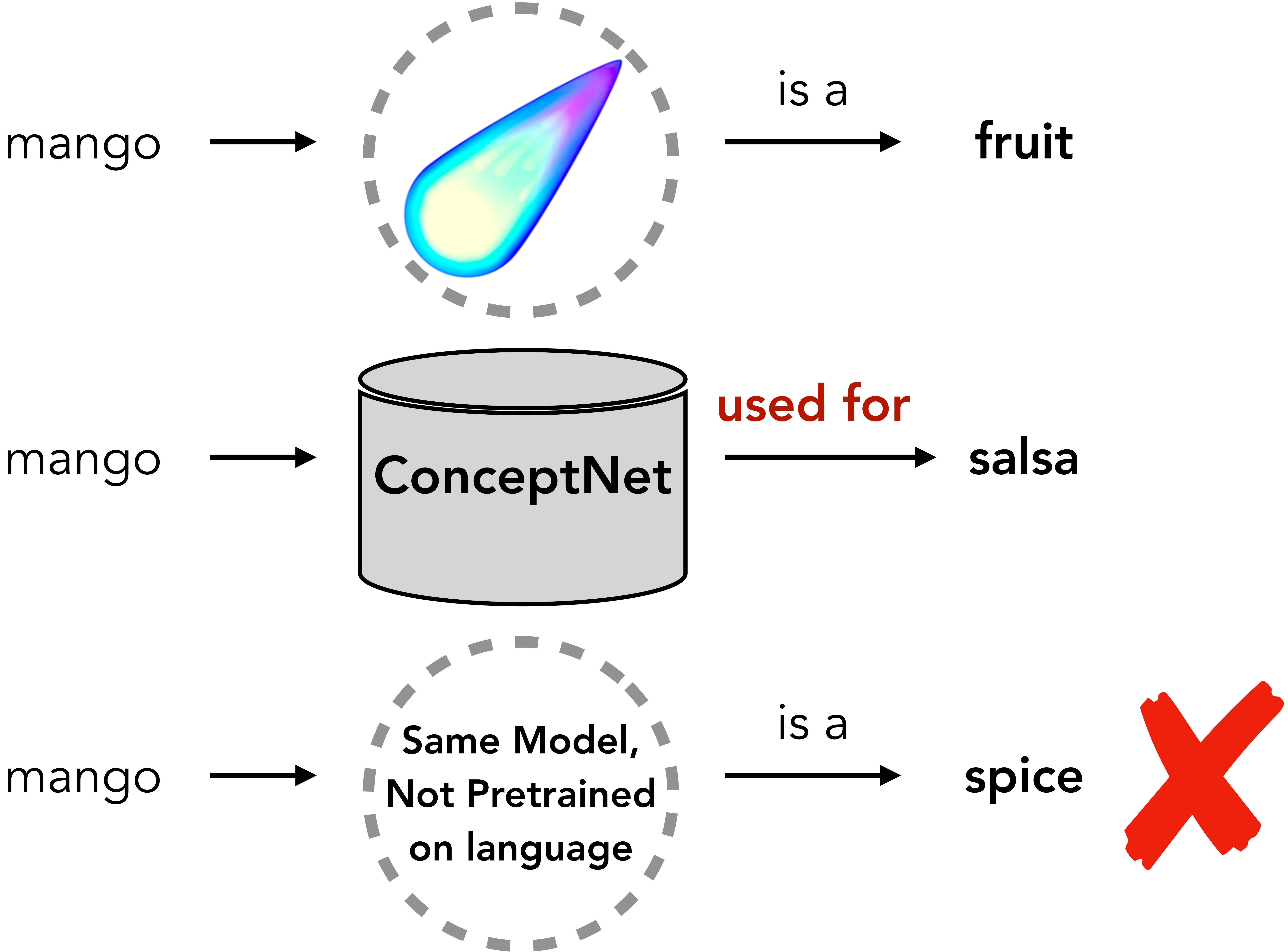
Bosselut et al., ACL 2019

Transfer Learning from Language



Bosselut et al., ACL 2019

Transfer Learning from Language



Bosselut et al., ACL 2019

Can’t language models just do this?

Do Language Models know this?

Sentence:

mango is a

Predictions:

2.1% great

1.9% very

1.2% new

1.0% good

1.0% small

← Undo

https://demo.allennlp.org/next-token-lm

Do Language Models know this?

Sentence:

mango is a

Predictions:

2.1% great
1.9% very
1.2% new
1.0% good
1.0% small
← Undo

a mango is a

4.2% good
4.0% very
2.5% great
2.4% delicious
1.8% sweet
← Undo

https://demo.allennlp.org/next-token-lm

Do Language Models know this?

Sentence:

mango is a

Predictions:

2.1% great
1.9% very
1.2% new
1.0% good
1.0% small
← Undo

a mango is a

4.2% good
4.0% very
2.5% great
2.4% delicious
1.8% sweet
← Undo

Sentence:

A mango is a

Predictions:

4.2% fruit
3.5% very
2.5% sweet
2.2% good
1.5% delicious
← Undo

https://demo.allennlp.org/next-token-lm

Do Masked Language Models know this?

Sentence:

mango is a [MASK]

Mask 1 Predictions:

69.7% **.**
9.3% **;**
1.7% **!**
0.8% **vegetable**
0.7% **?**

Sentence:

mango is a [MASK] .

Mask 1 Predictions:

7.6% **staple**
7.6% **vegetable**
4.6% **plant**
3.5% **tree**
3.5% **fruit**

Sentence:

A mango is a [MASK] .

Mask 1 Predictions:

16.0% **banana**
12.1% **fruit**
5.9% **plant**
5.5% **vegetable**
2.5% **candy**

https://demo.allennlp.org/masked-lm

Let’s talk about the elephant in the room.
What about GPT-3?

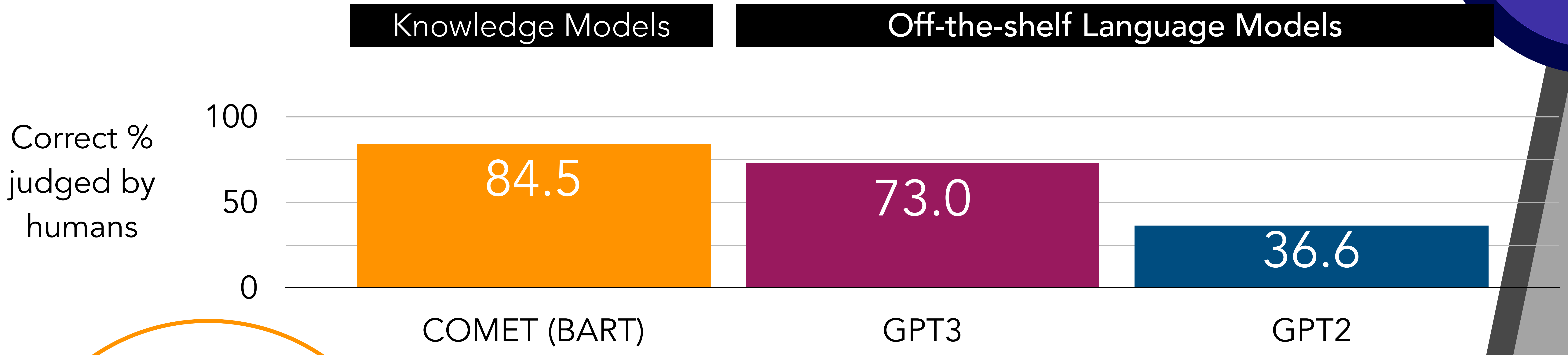
Does GPT-3 have commonsense knowledge?

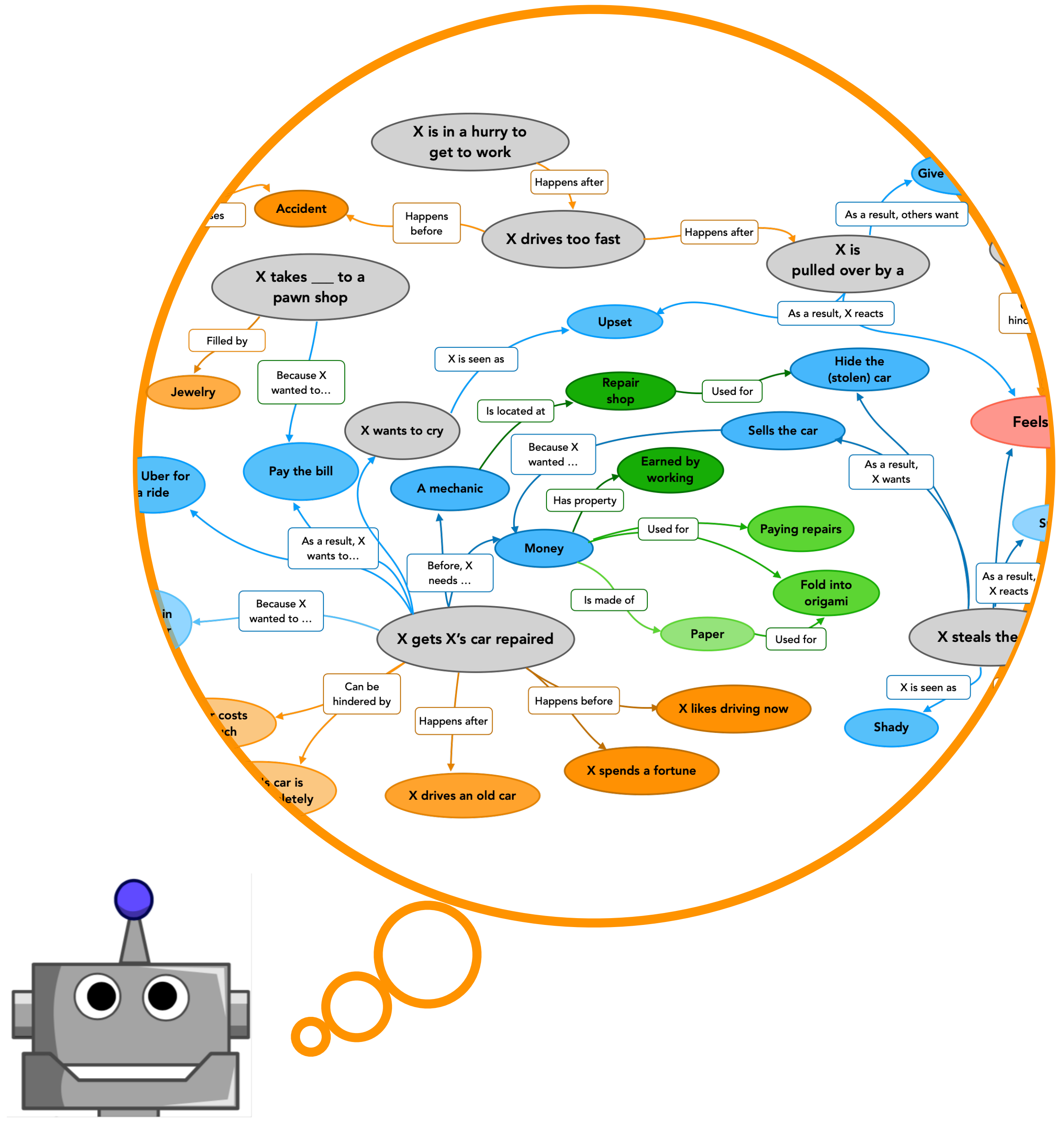
```
Q: What is your favorite animal?
A: My favorite animal is a dog.

Q: Wh
A: Be
```

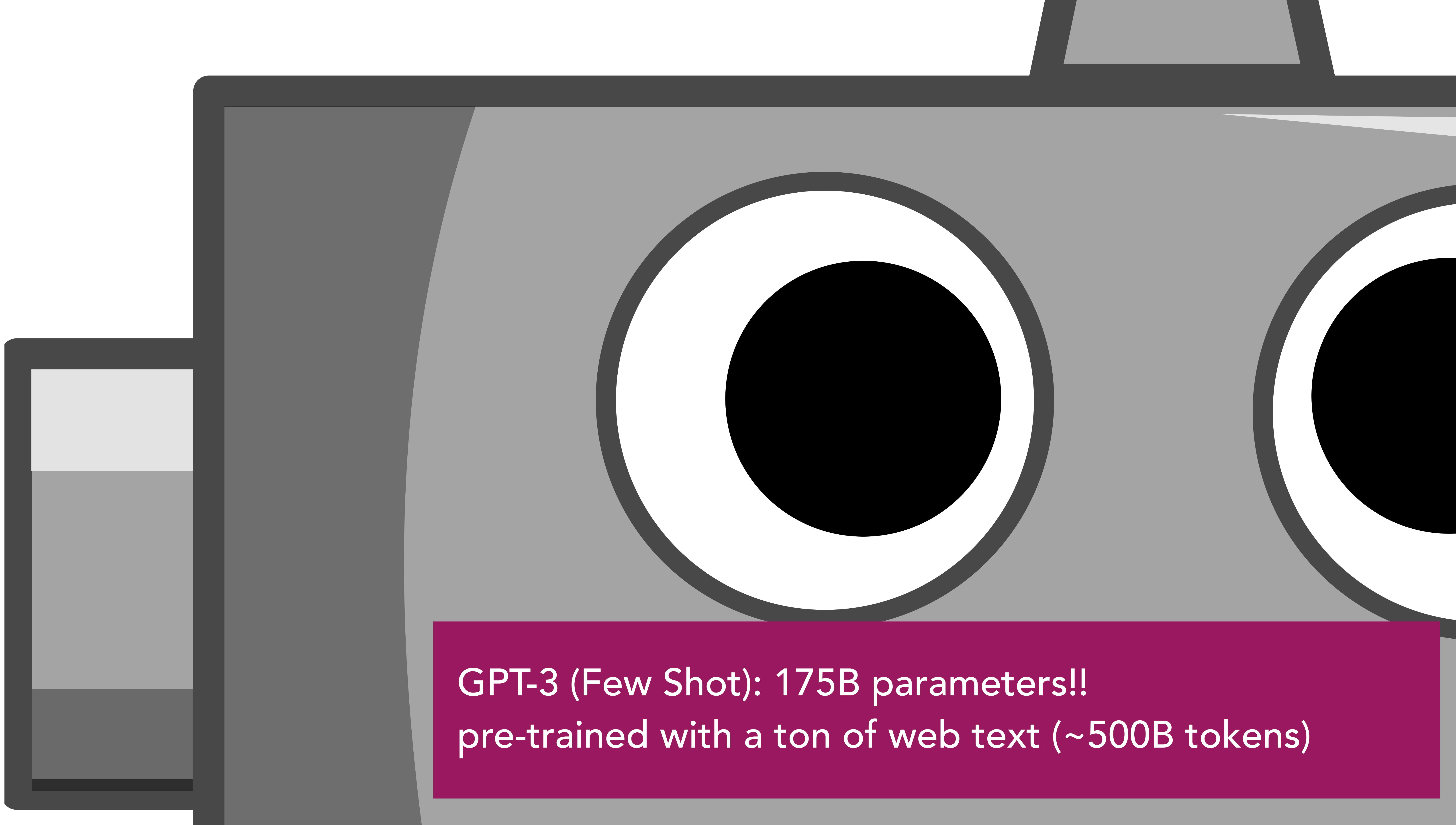
```
Q: How many eyes does a giraffe have?
A: A giraffe has two eyes.
```

```
Q: Why don't animals have three legs?
A: Animals don't have three legs because
they would fall over.
```

Kevin Lacker's Blog. July 6, 2020 (https://lacker.io/ai/2020/07/06/giving-gpt-3-a-turing-test.html)







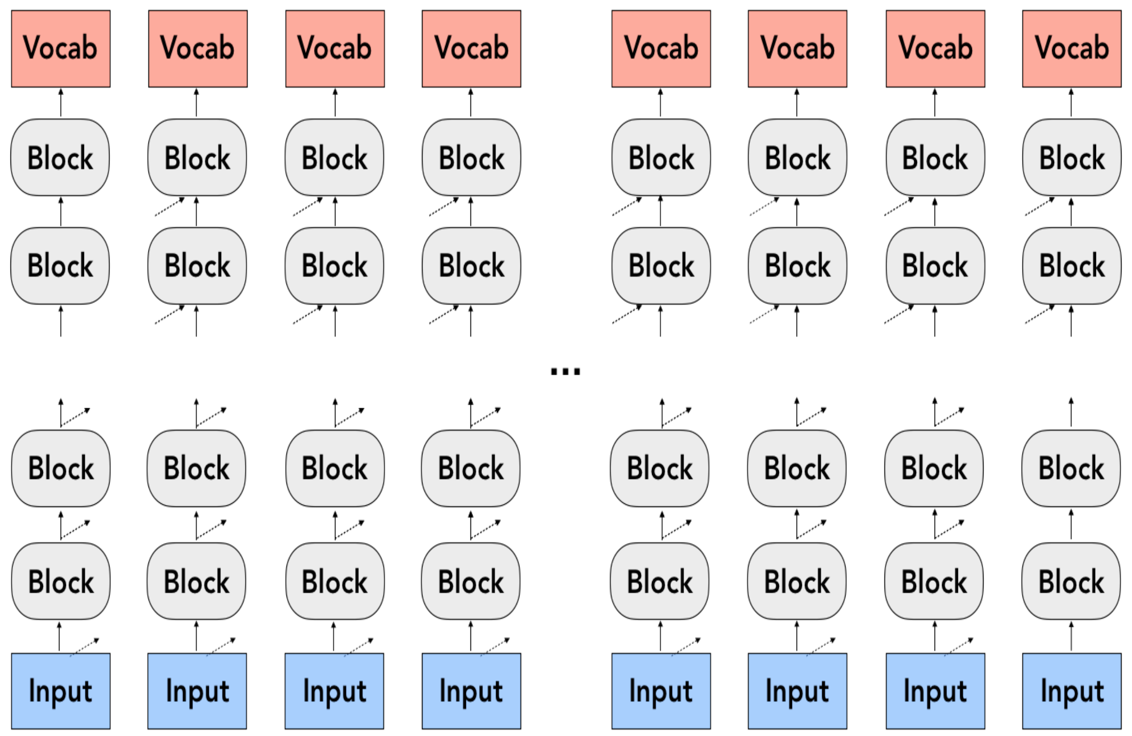
COMET (BART): 435x smaller model (~400M params), informed by **ATOMIC$^{20}_{20}$**

GPT-3 (Few Shot): 175B parameters!! pre-trained with a ton of web text (~500B tokens)

Hwang et al., AAAI 2021

Commonsense Transformers

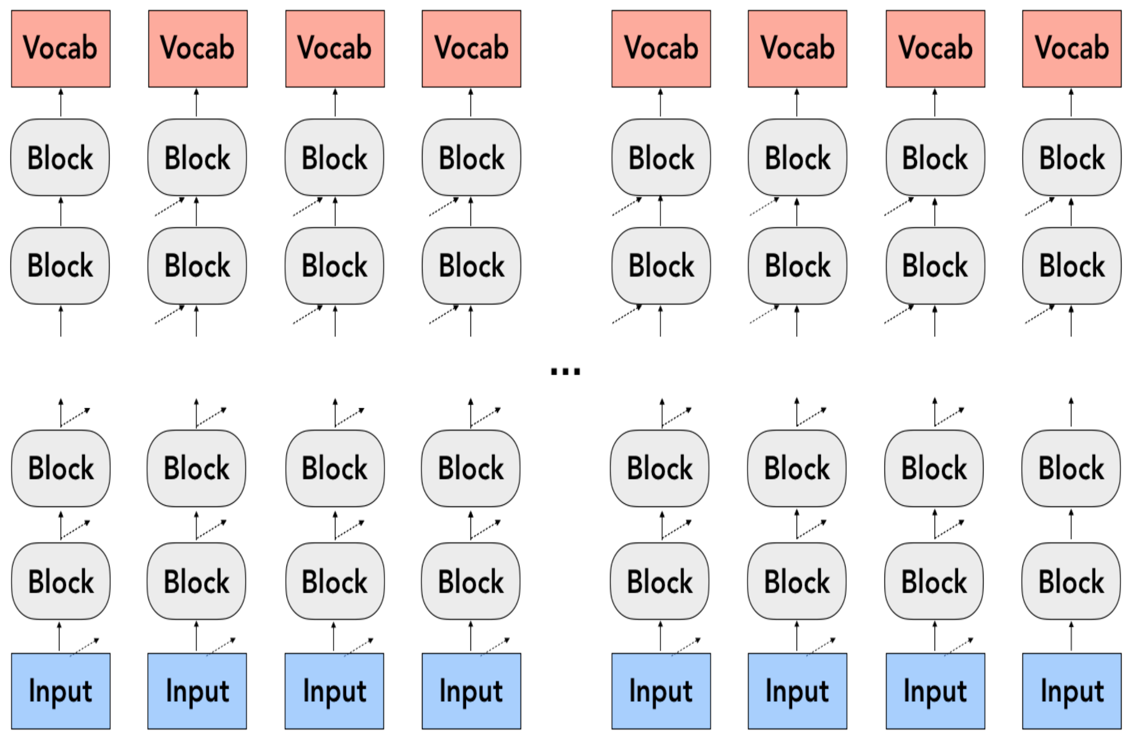
- **Learn implicit knowledge** at scale from language models and web-scale text



Pre-trained
Language Model

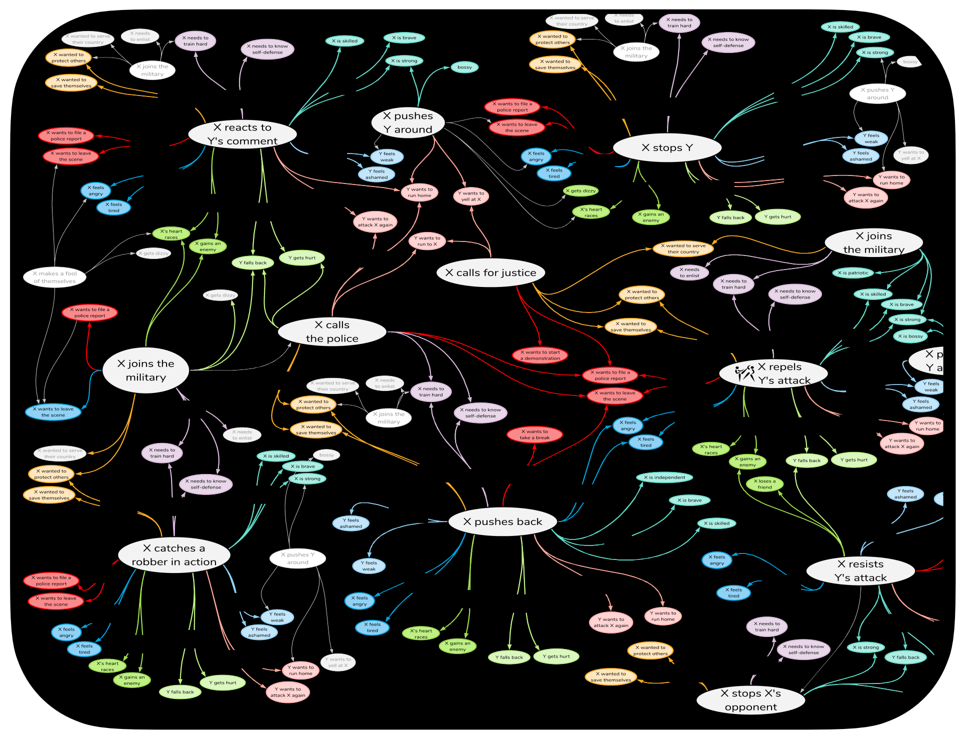
Commonsense Transformers

- **Learn implicit knowledge** at scale from language models and web-scale text
- **Learn explicit structure of knowledge** from symbolic knowledge graphs



Pre-trained
Language Model

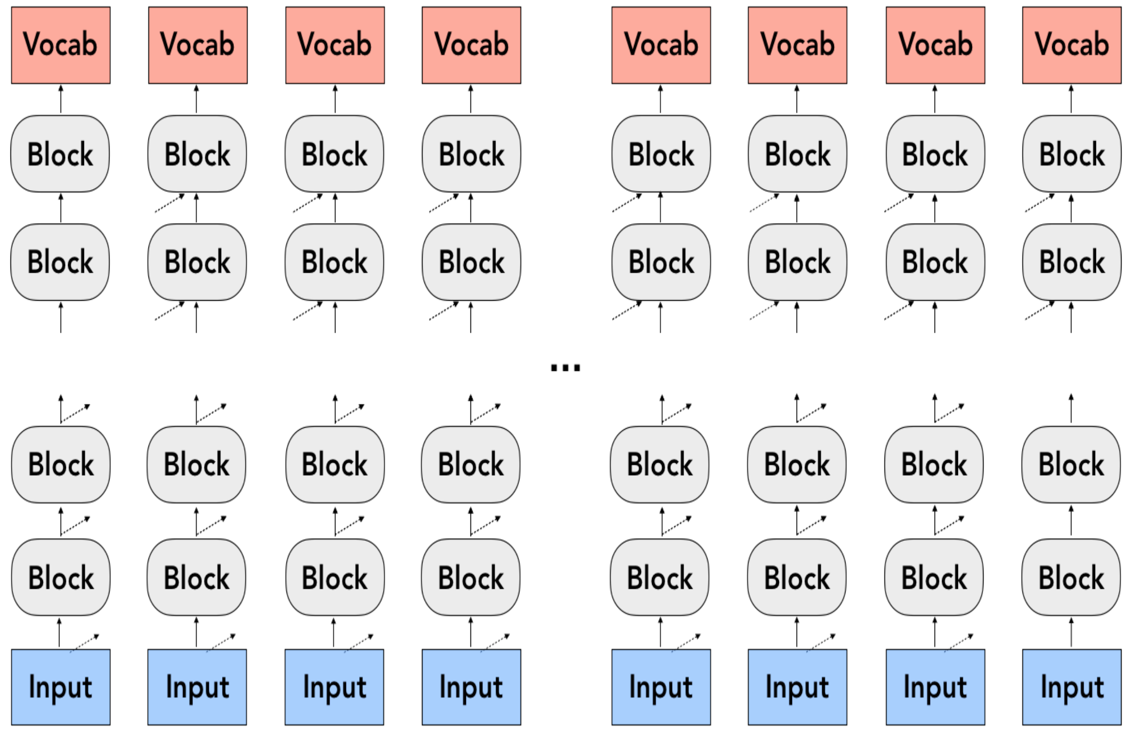
+



Seed Knowledge
Graph Training

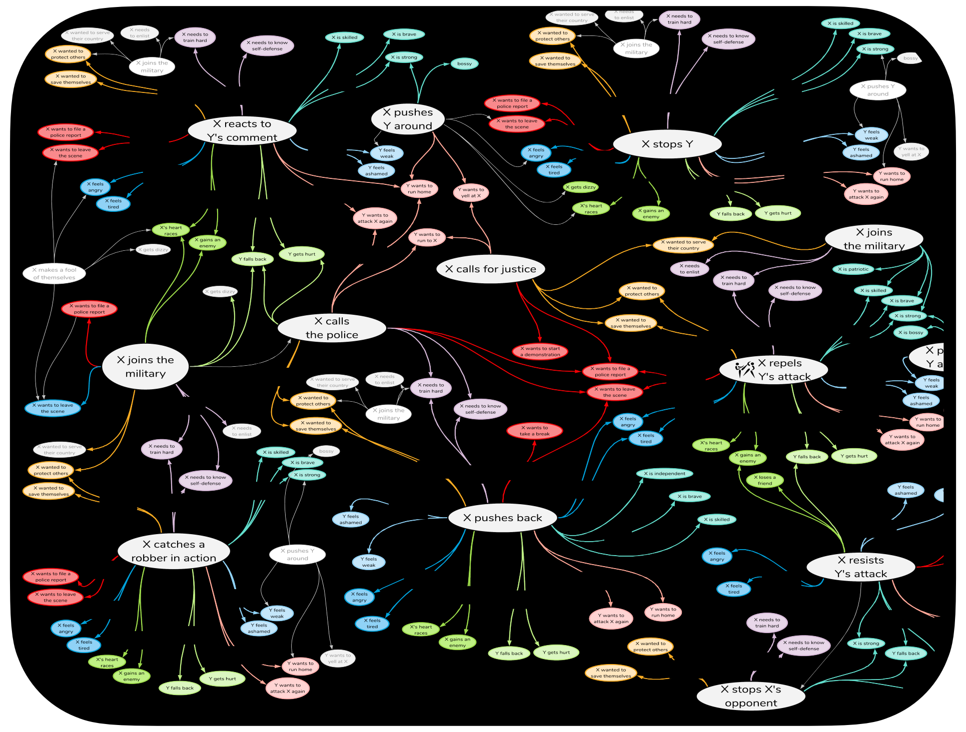
Commonsense Transformers

- **Learn implicit knowledge** at scale from language models and web-scale text
- **Learn explicit structure of knowledge** from symbolic knowledge graphs
- Resulting knowledge model **generalizes structure** to other concepts



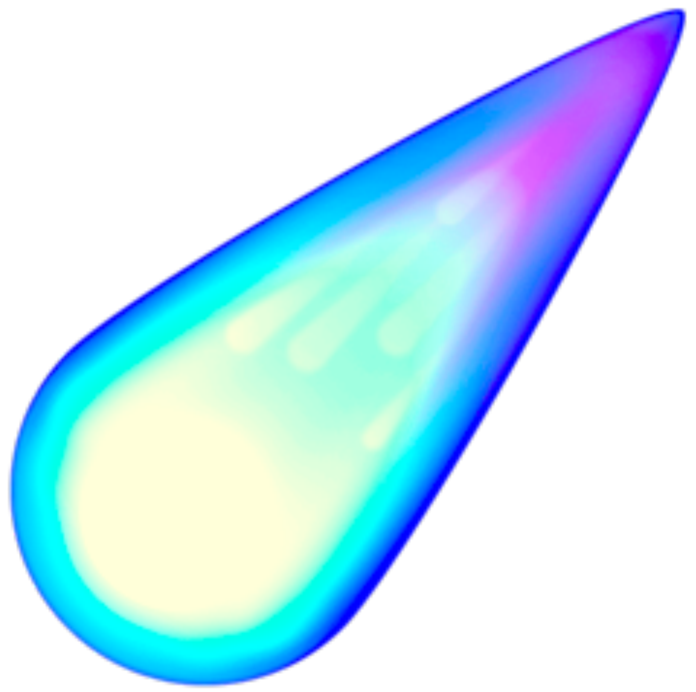
Pre-trained
Language Model

+



Seed Knowledge
Graph Training

=



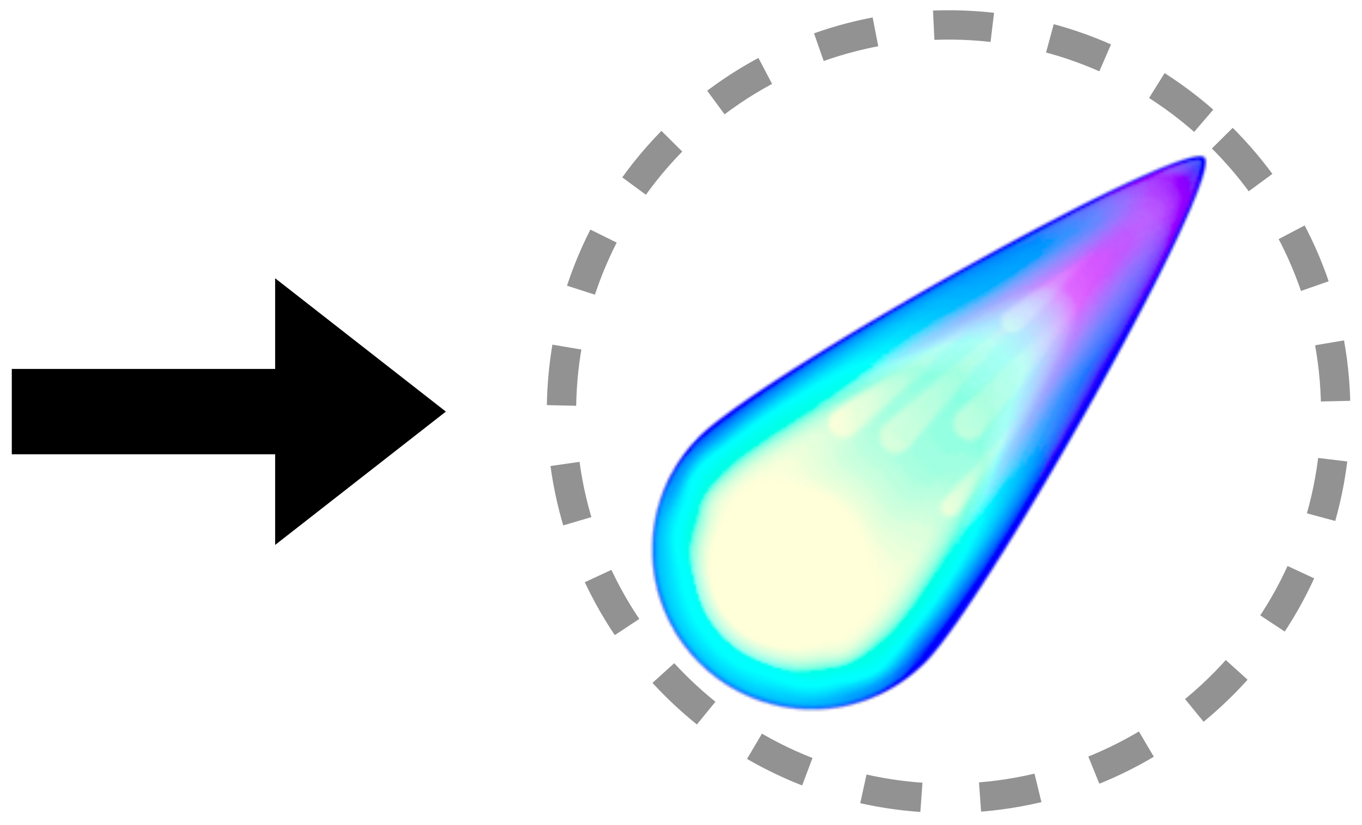
COMET

What are the implications of representing commonsense knowledge in this manner?

Commonsense Knowledge for any Situation

- transformer-style architecture — input format is natural language
  - event can be fully parsed

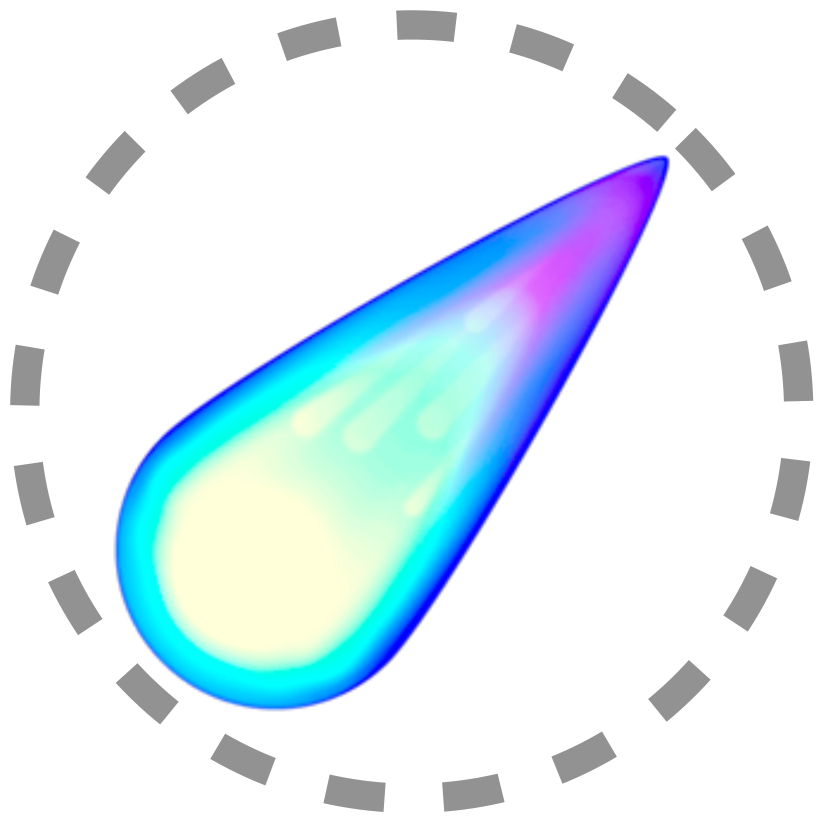
Kai knew that things were getting out of control and managed to keep his temper in check



Commonsense Knowledge for any Situation

- transformer-style architecture — input format is natural language
  - event can be fully parsed
  - knowledge generated **dynamically** from neural knowledge model

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



Kai wants to avoid trouble

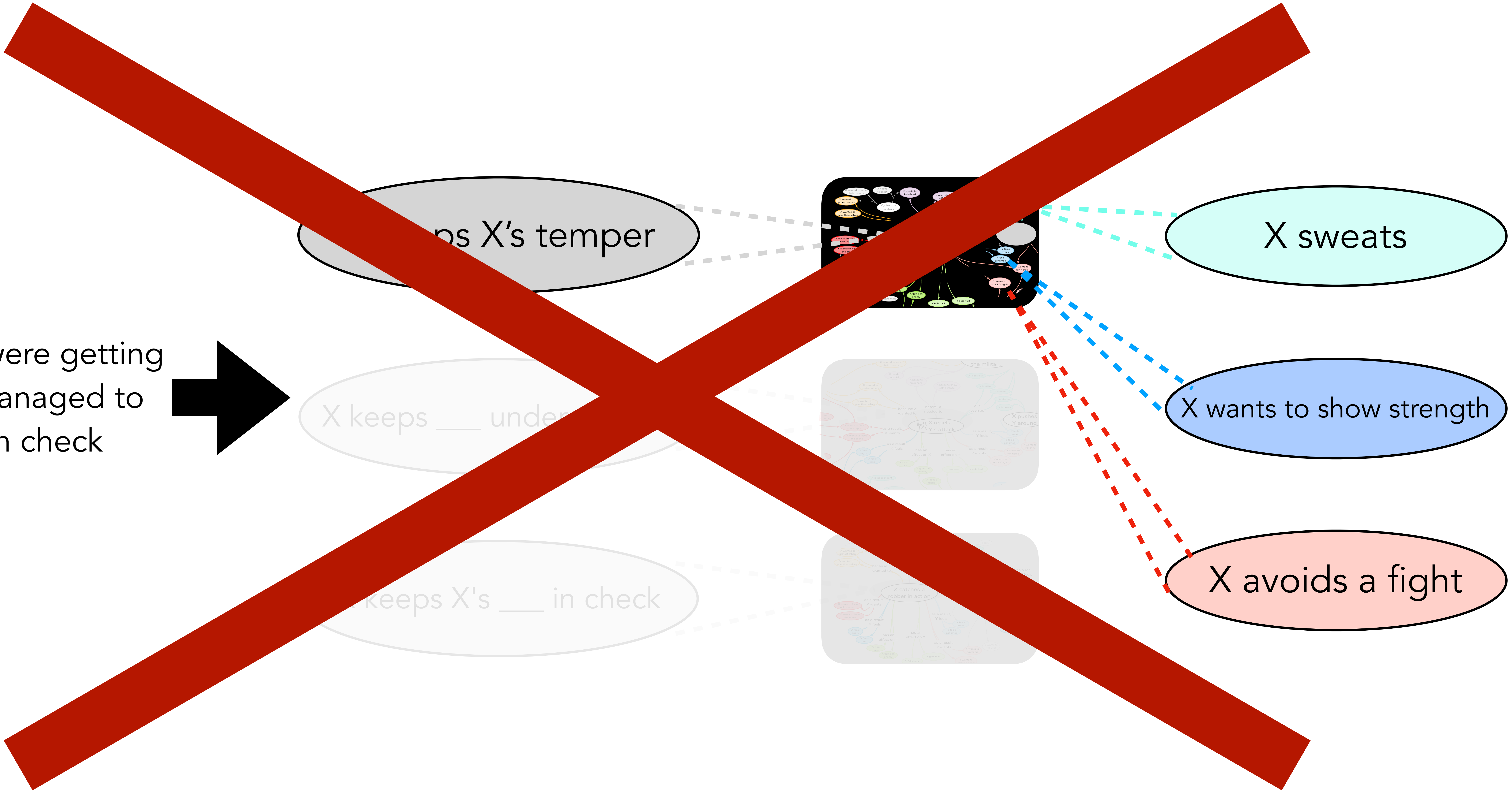
Kai intends to be calm

Kai stays calm

Kai is viewed as cautious

Reasoning with Knowledge Graphs

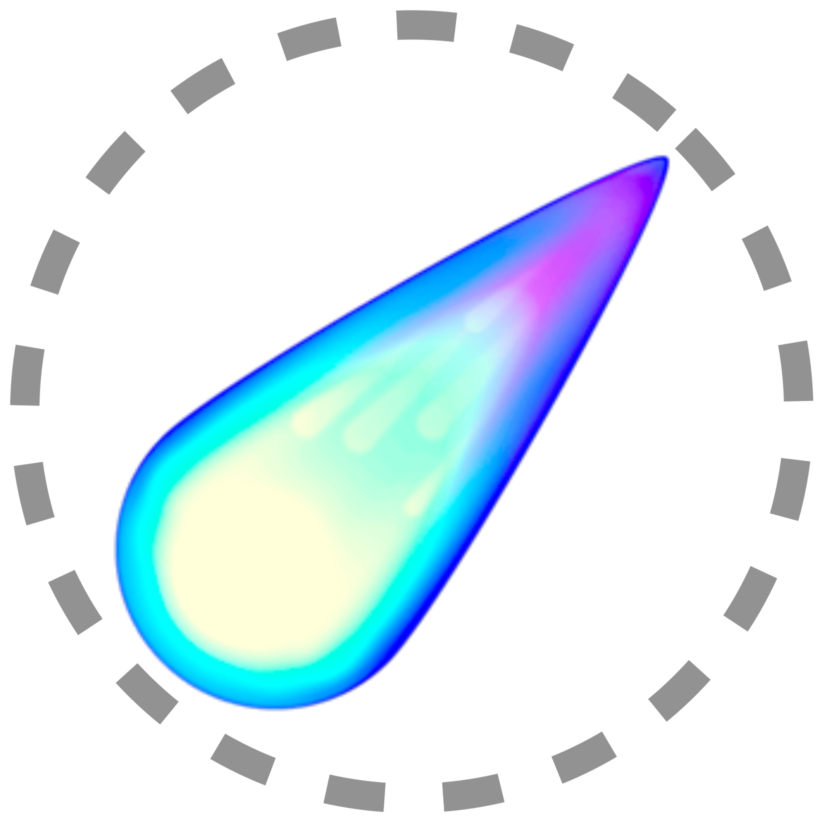
Kai knew that things were getting out of control and managed to keep his temper in check



Commonsense Knowledge for any Situation

- transformer-style architecture — input format is natural language
  - event can be fully parsed
  - knowledge generated **dynamically** from **neural** knowledge model

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



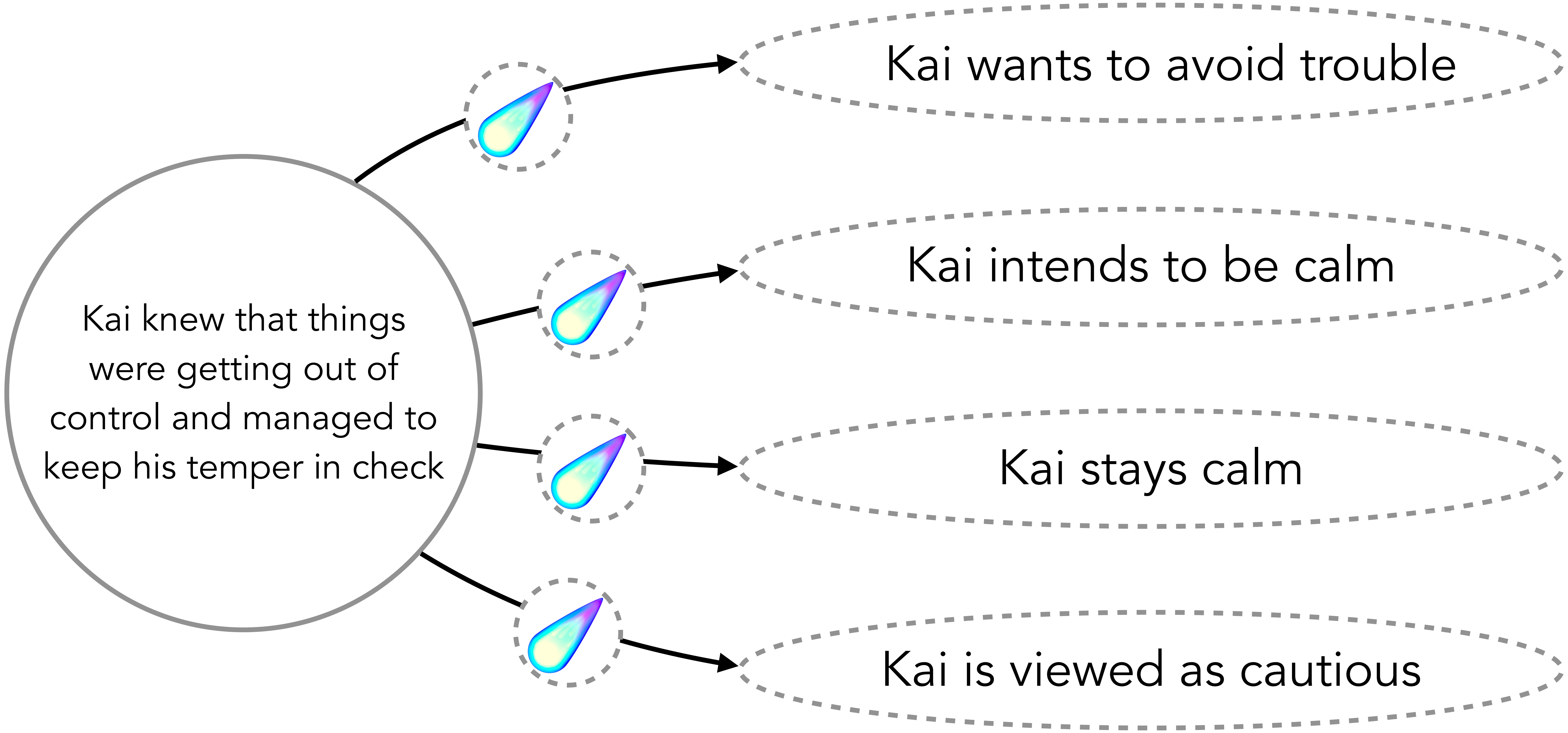
Kai wants to avoid trouble

Kai intends to be calm

Kai stays calm

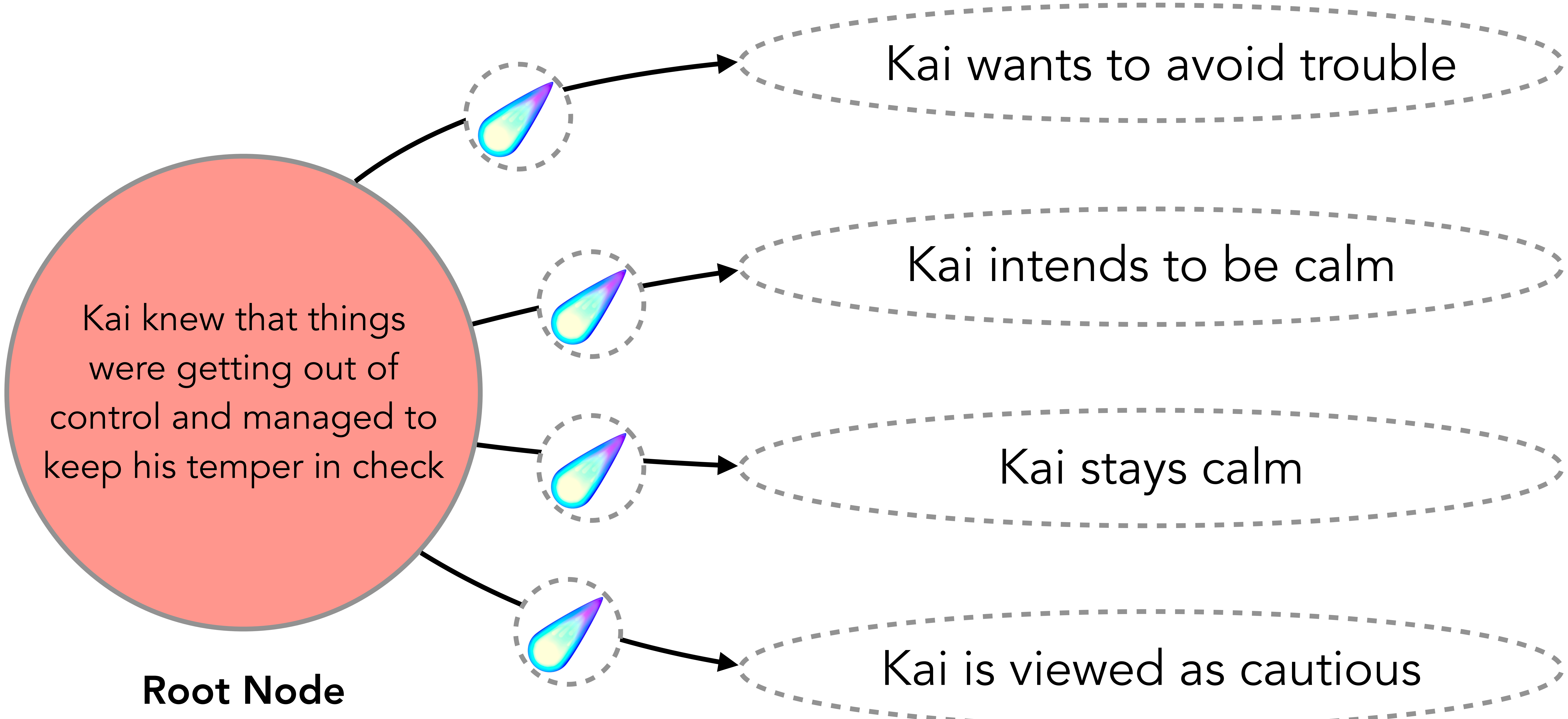
Kai is viewed as cautious

Dynamic Construction of Knowledge Graphs



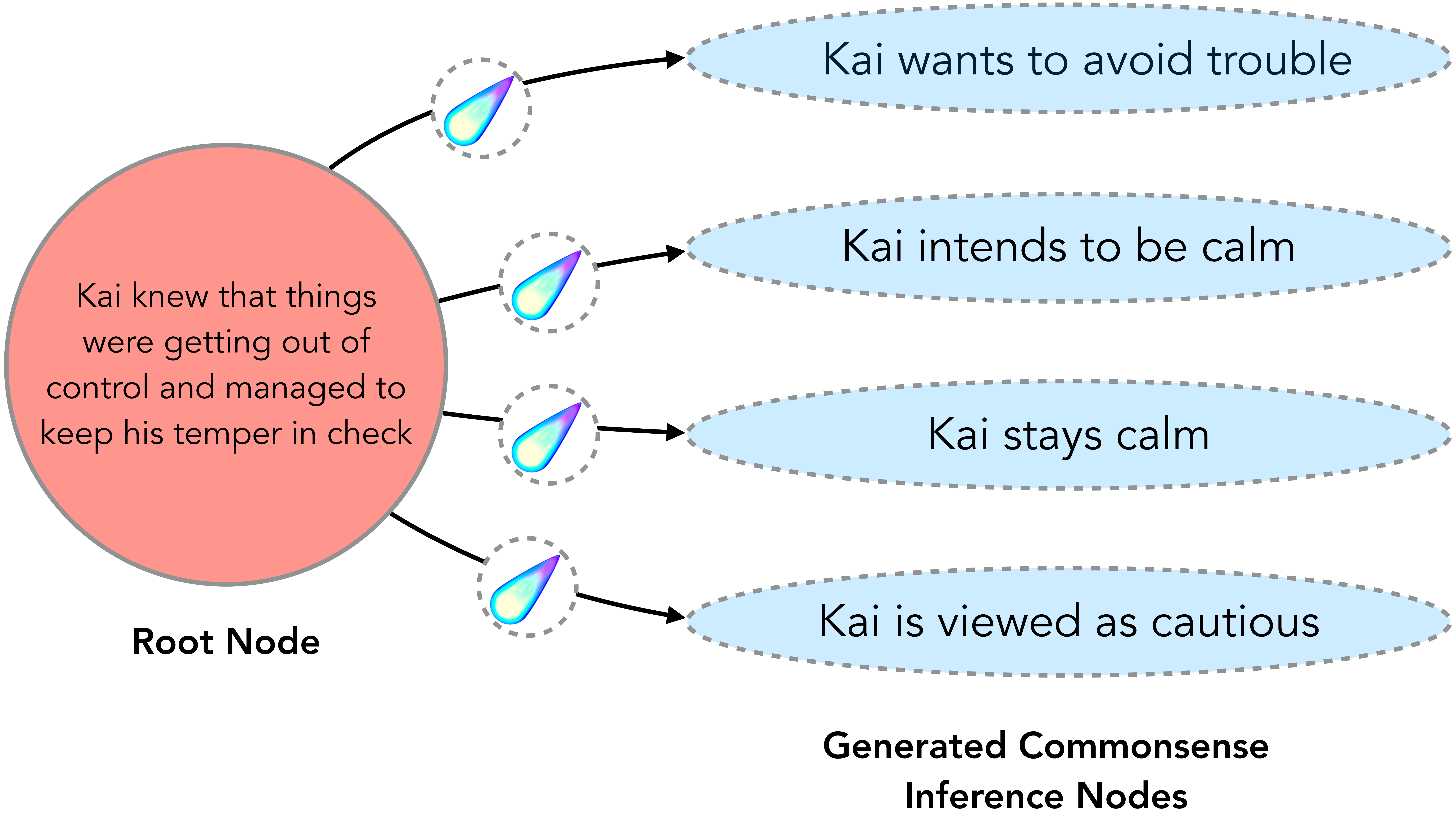
Bosselut et al., AAAI 2021

Dynamic Construction of Knowledge Graphs

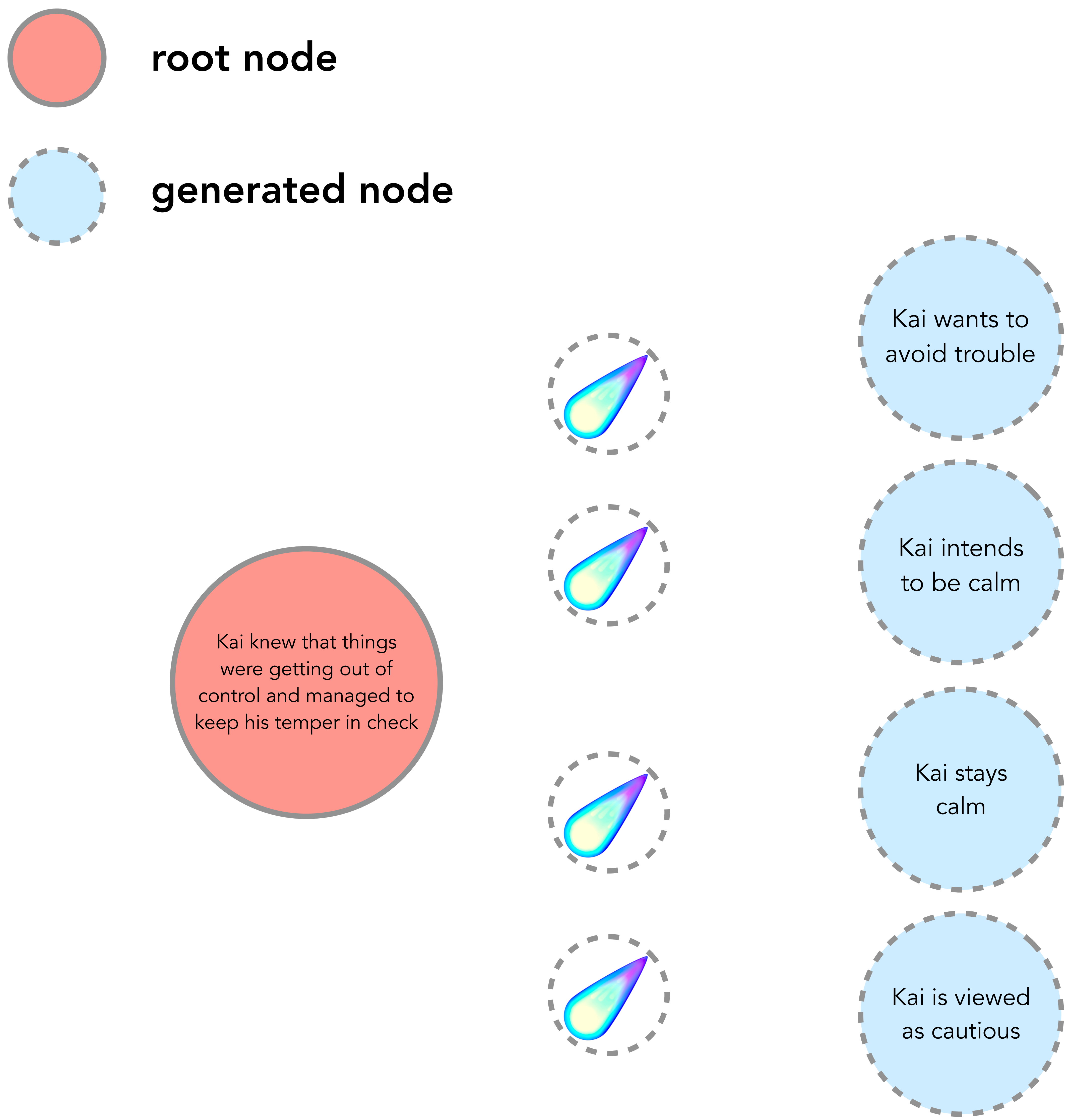


Bosselut et al., AAAI 2021

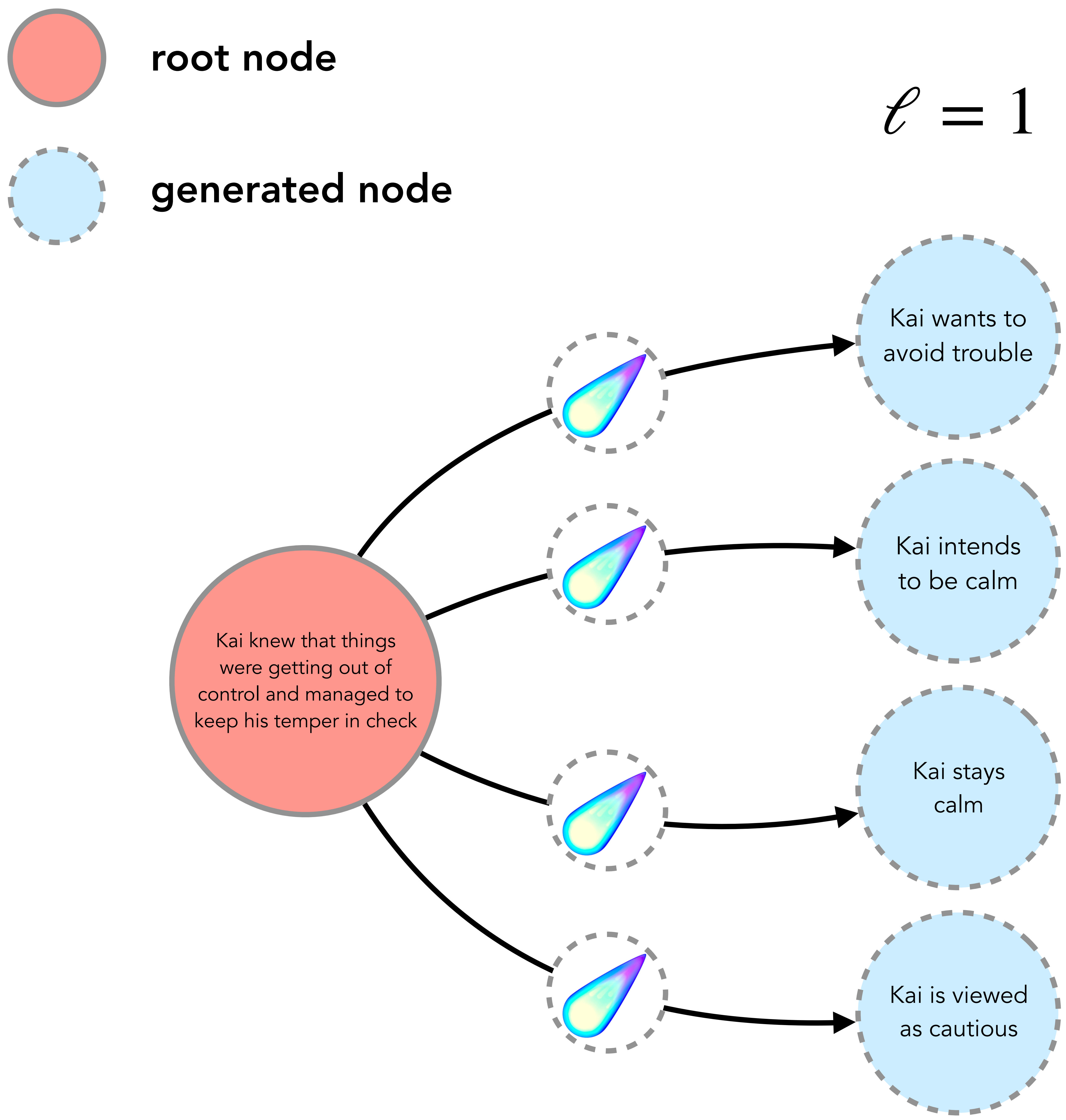
Dynamic Construction of Knowledge Graphs



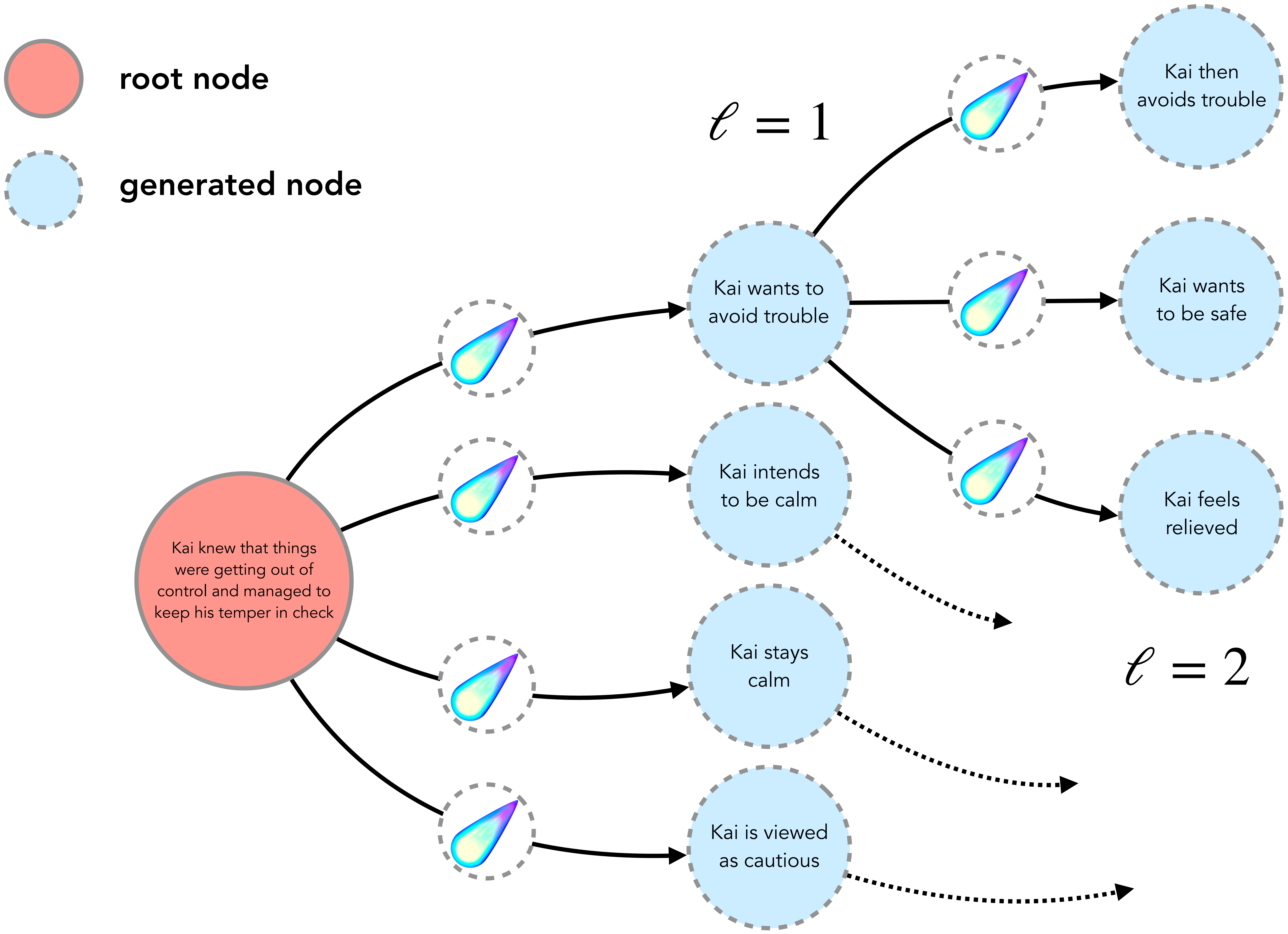
Bosselut et al., AAAI 2021



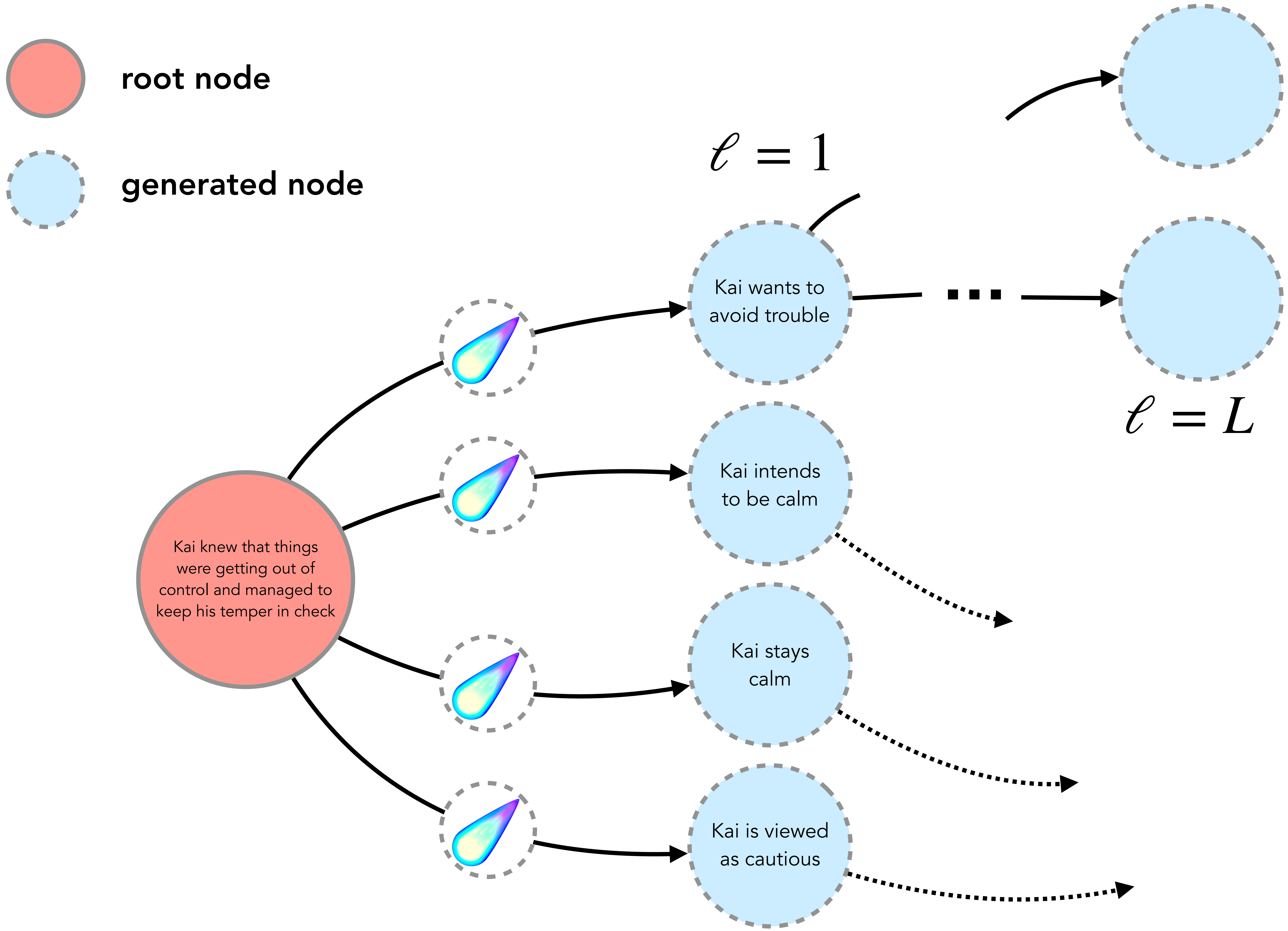
Bosselut et al., AAAI 2021



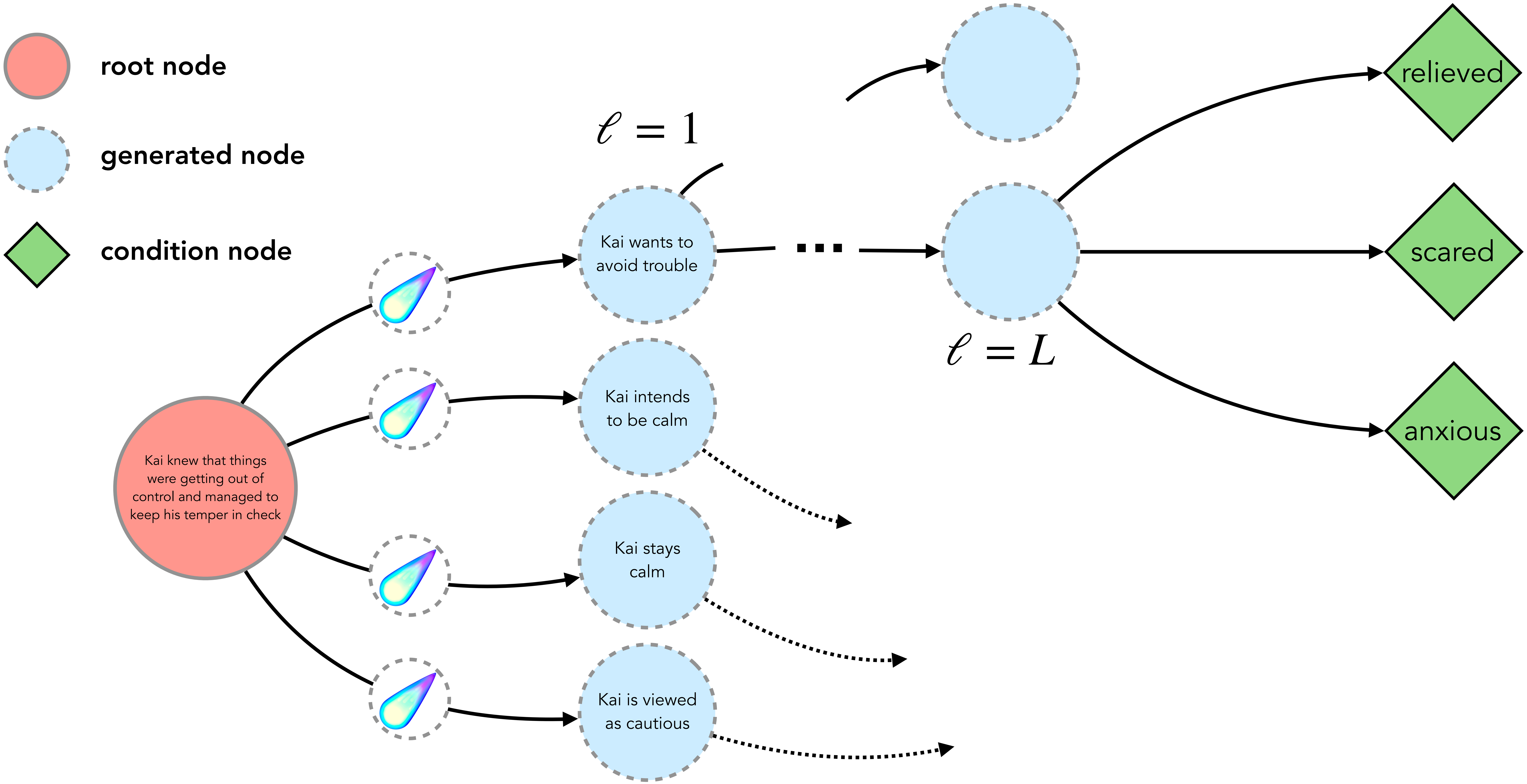
Bosselut et al., AAAI 2021



Bosselut et al., AAAI 2021



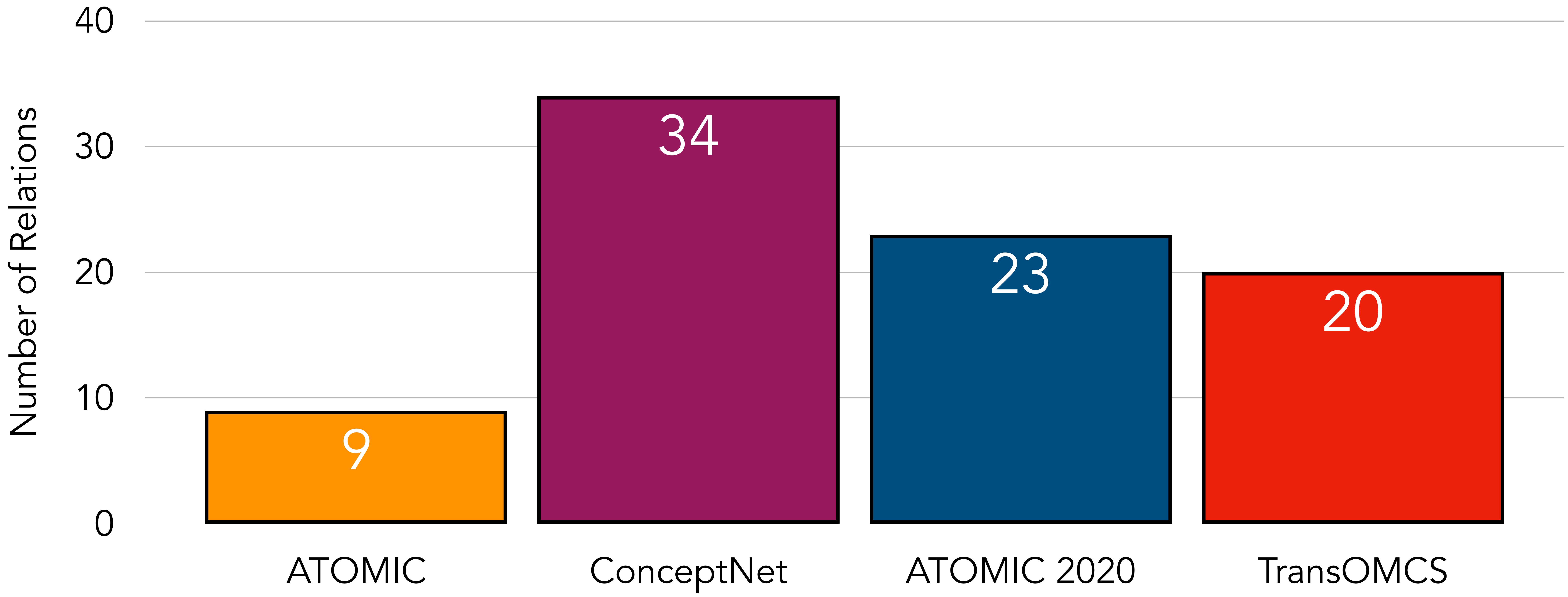
Bosselut et al., AAAI 2021



Bosselut et al., AAAI 2021

What about the relations? Those are fixed.

Fixed Relation Sets



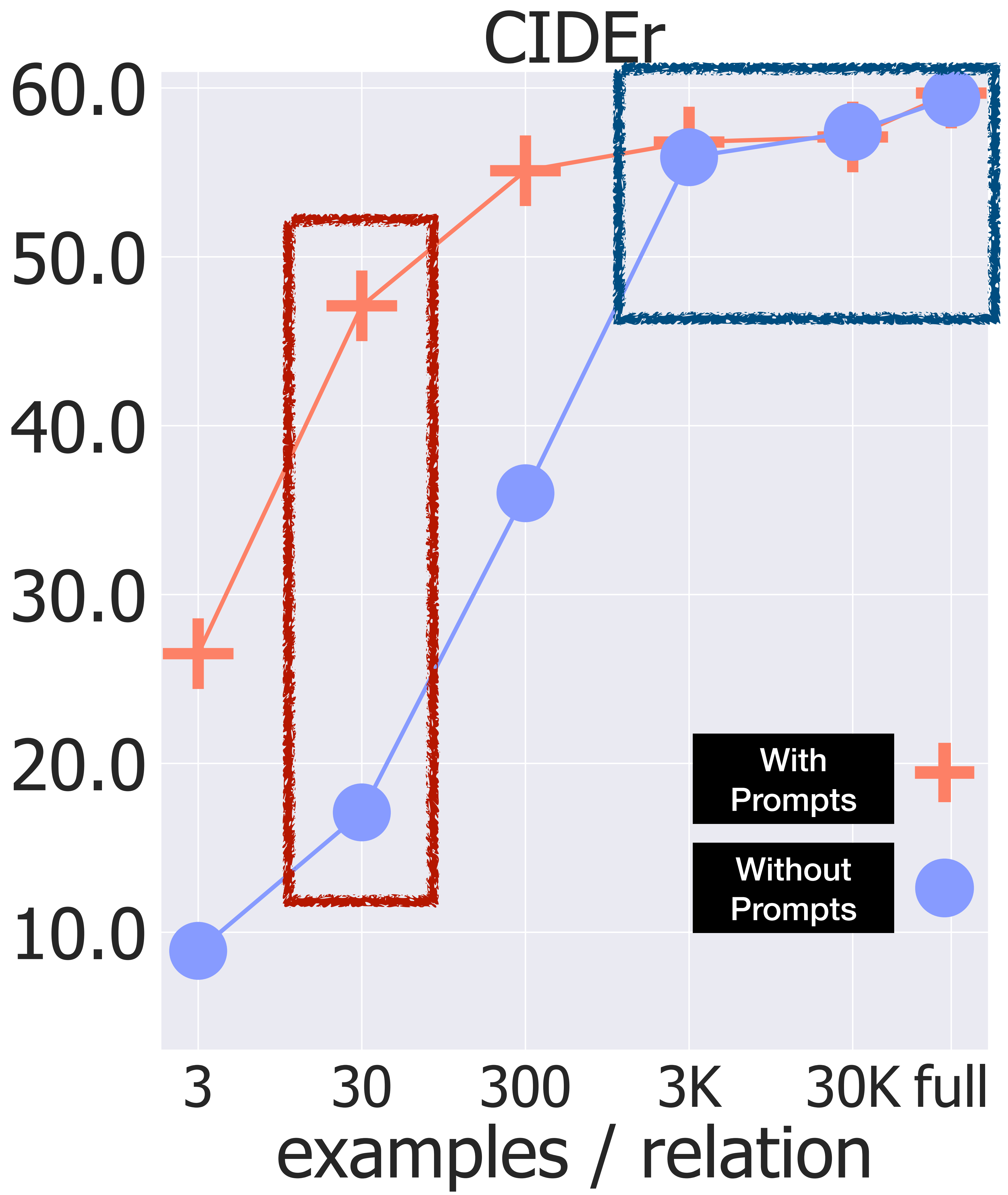
Da et al., 2021

Few-shot Knowledge Models

Using prompts induces rapid knowledge model adaptation in T5!

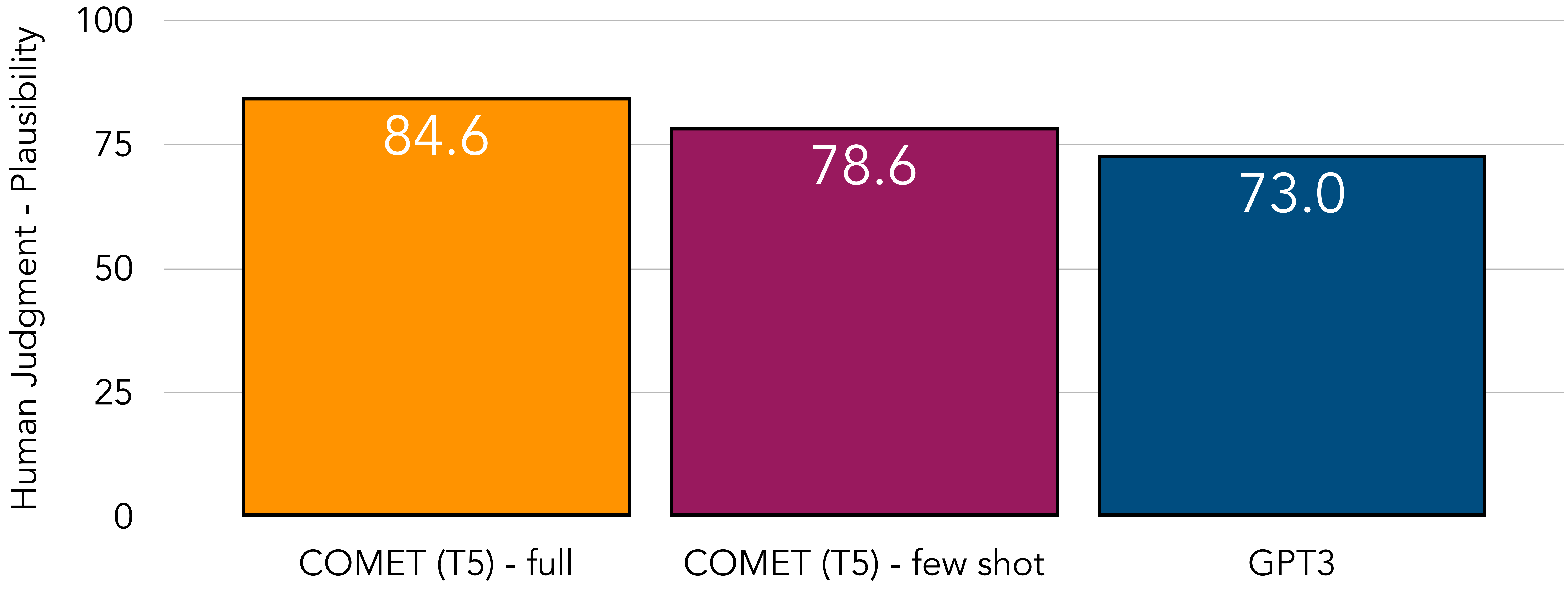
PersonX goes to the mall
<xIntent> to buy clothes

PersonX goes to the mall
because they wanted to
buy clothes



Da et al., 2021

Few-shot Knowledge Models

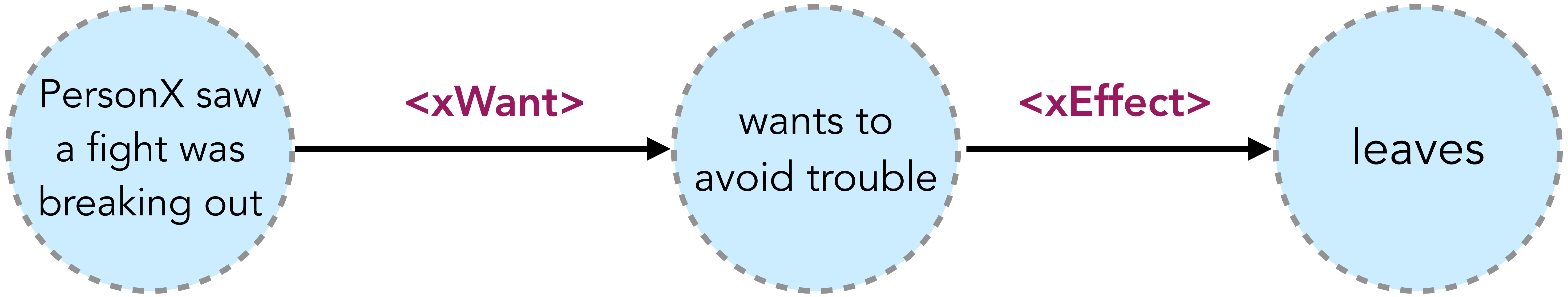


Just ~100 examples!
5 examples / relation

Da et al., 2021

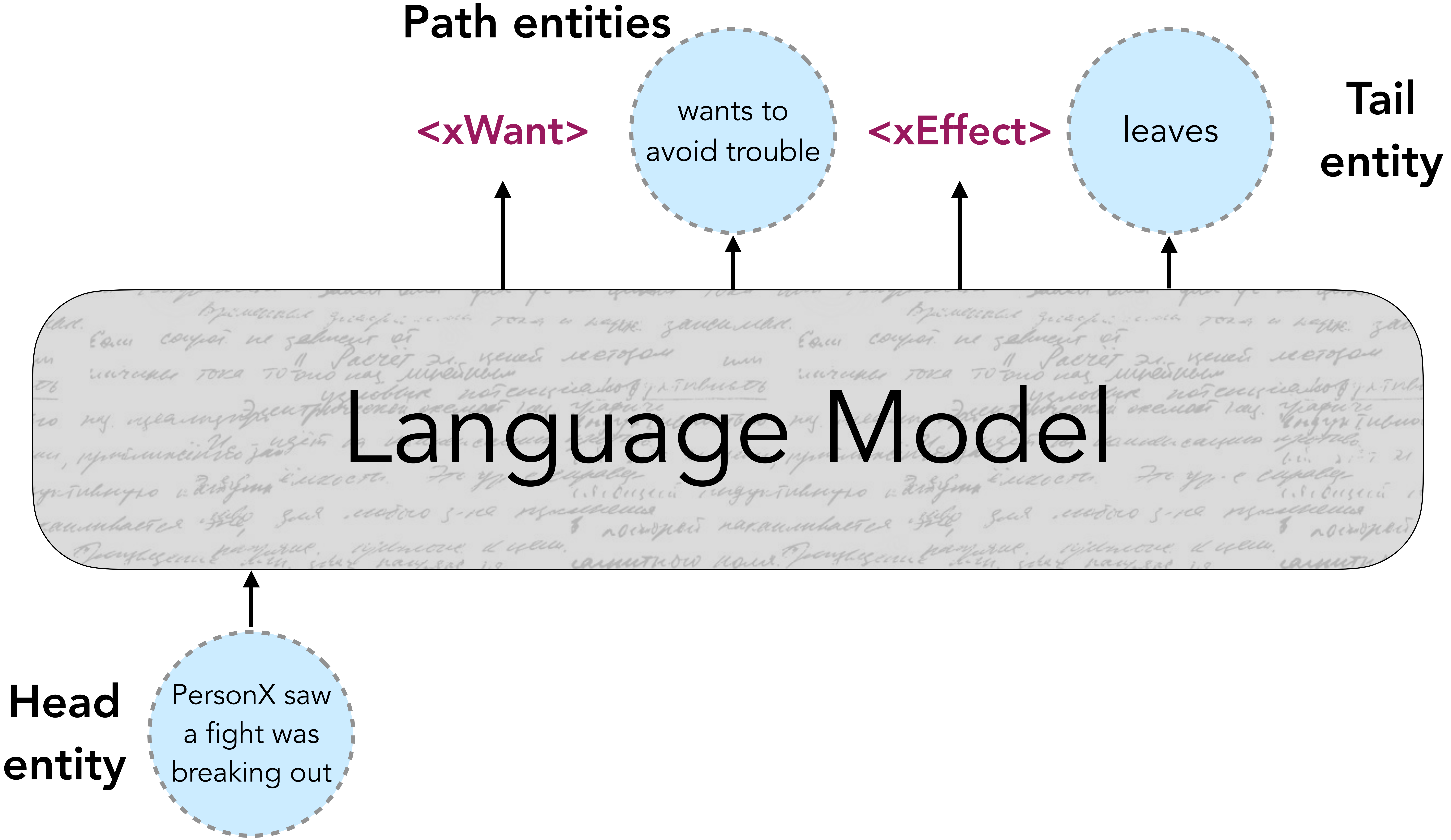
Can we model more complex commonsense knowledge?

Path Knowledge Models



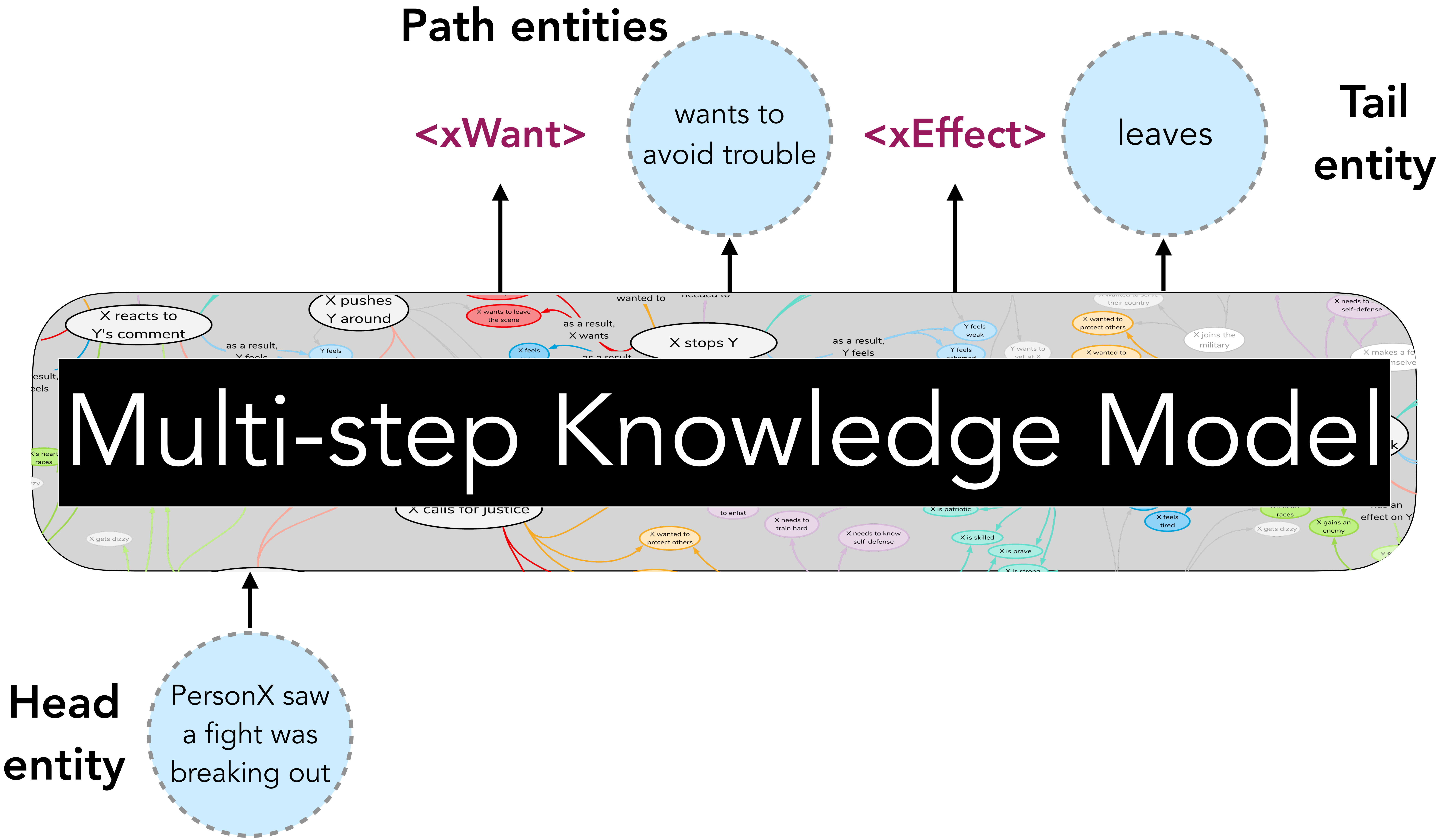
Wang et al., EMNLP Findings 2020

Path Knowledge Models



Wang et al., EMNLP Findings 2020

Path Knowledge Models

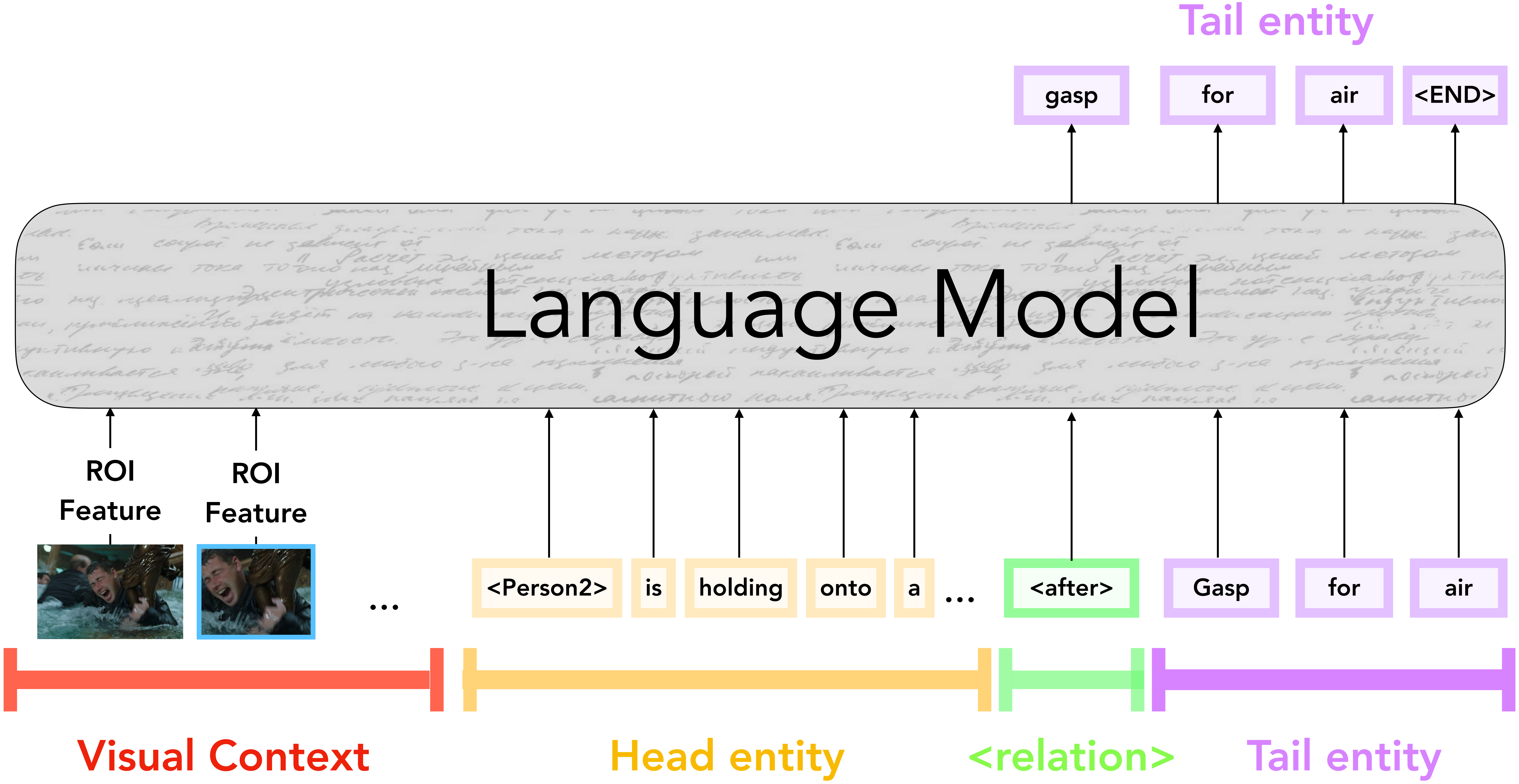


Wang et al., EMNLP Findings 2020

Visual Commonsense Knowledge Models

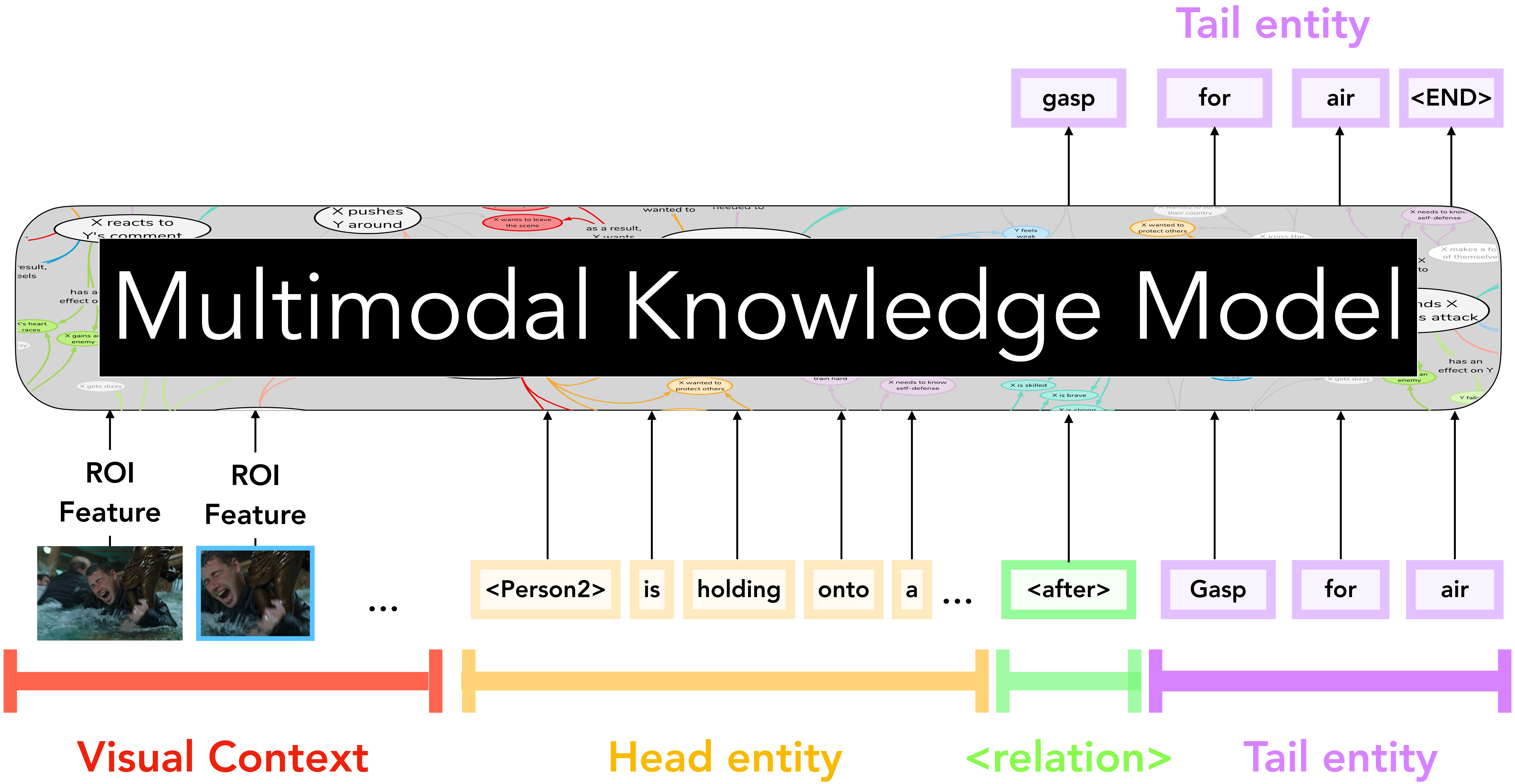
Park et al., ICCV 2020

VisualCOMET



Park et al., ICCV 2020

VisualCOMET

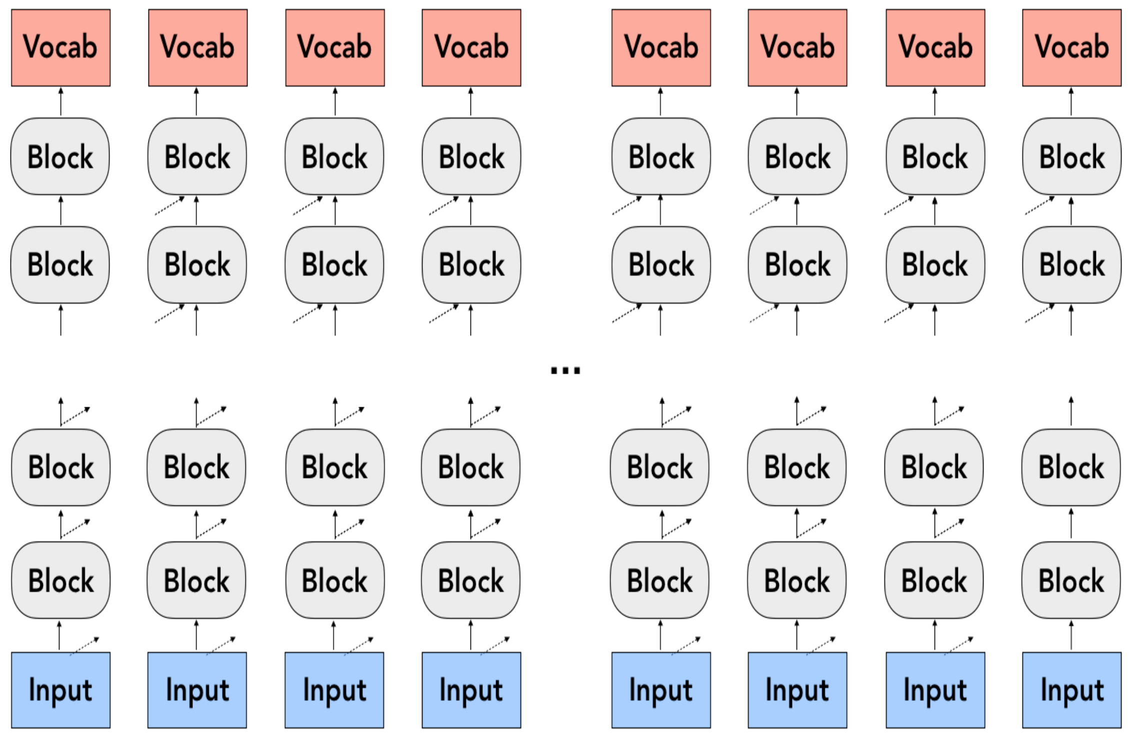


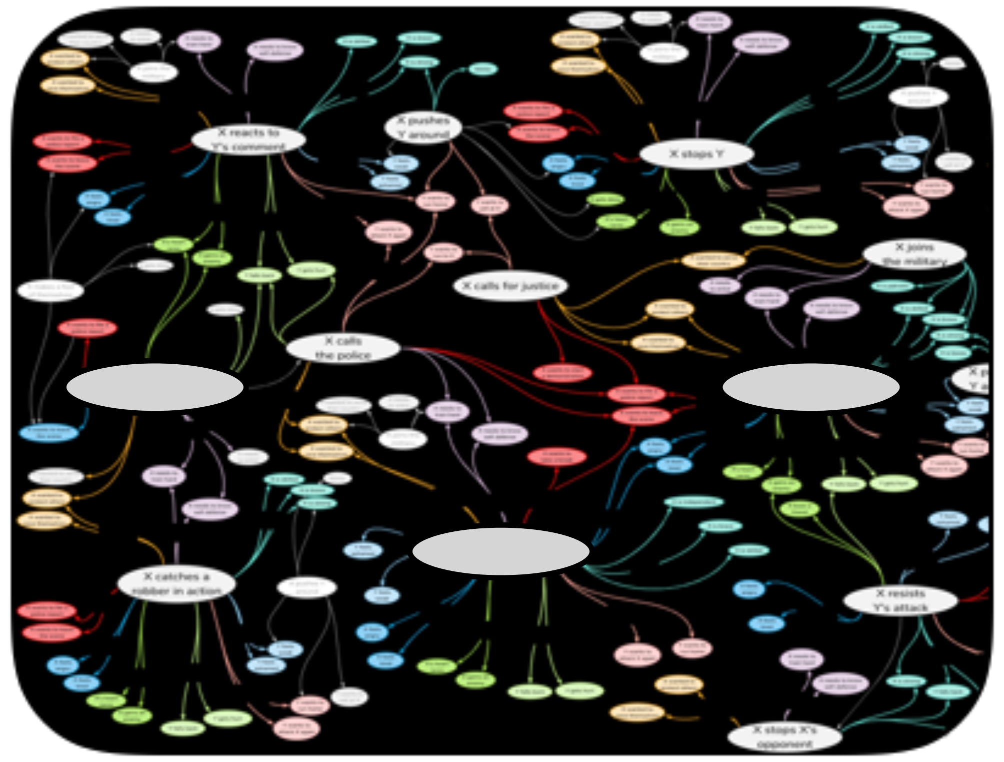
Park et al., ICCV 2020

Okay, what’s the catch?

Limitations of Knowledge Models

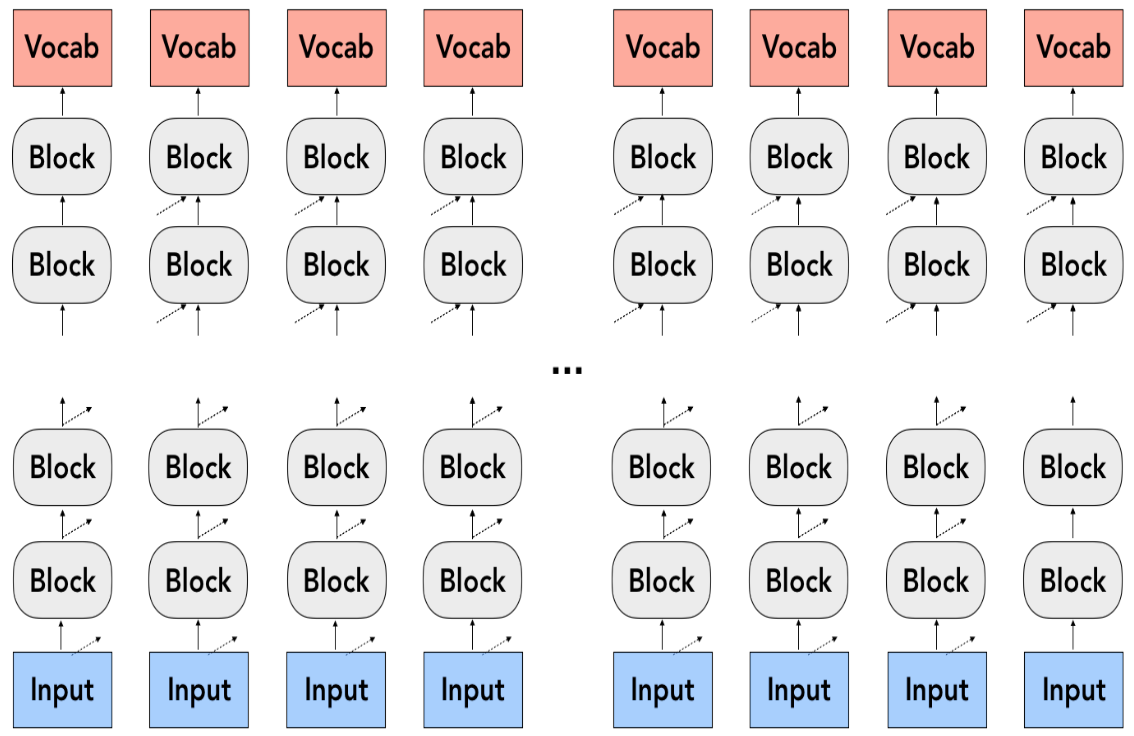
- Base Self-supervised Model
  - biases, history
- Seed Knowledge Graph
  - bias, language, relations, schema

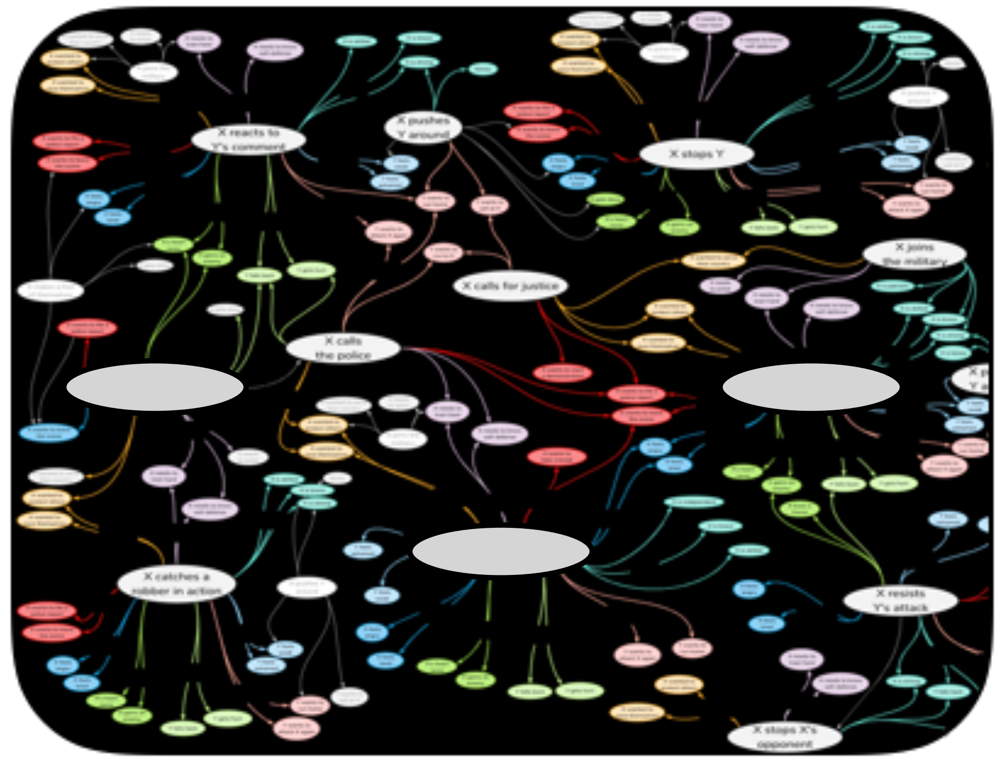




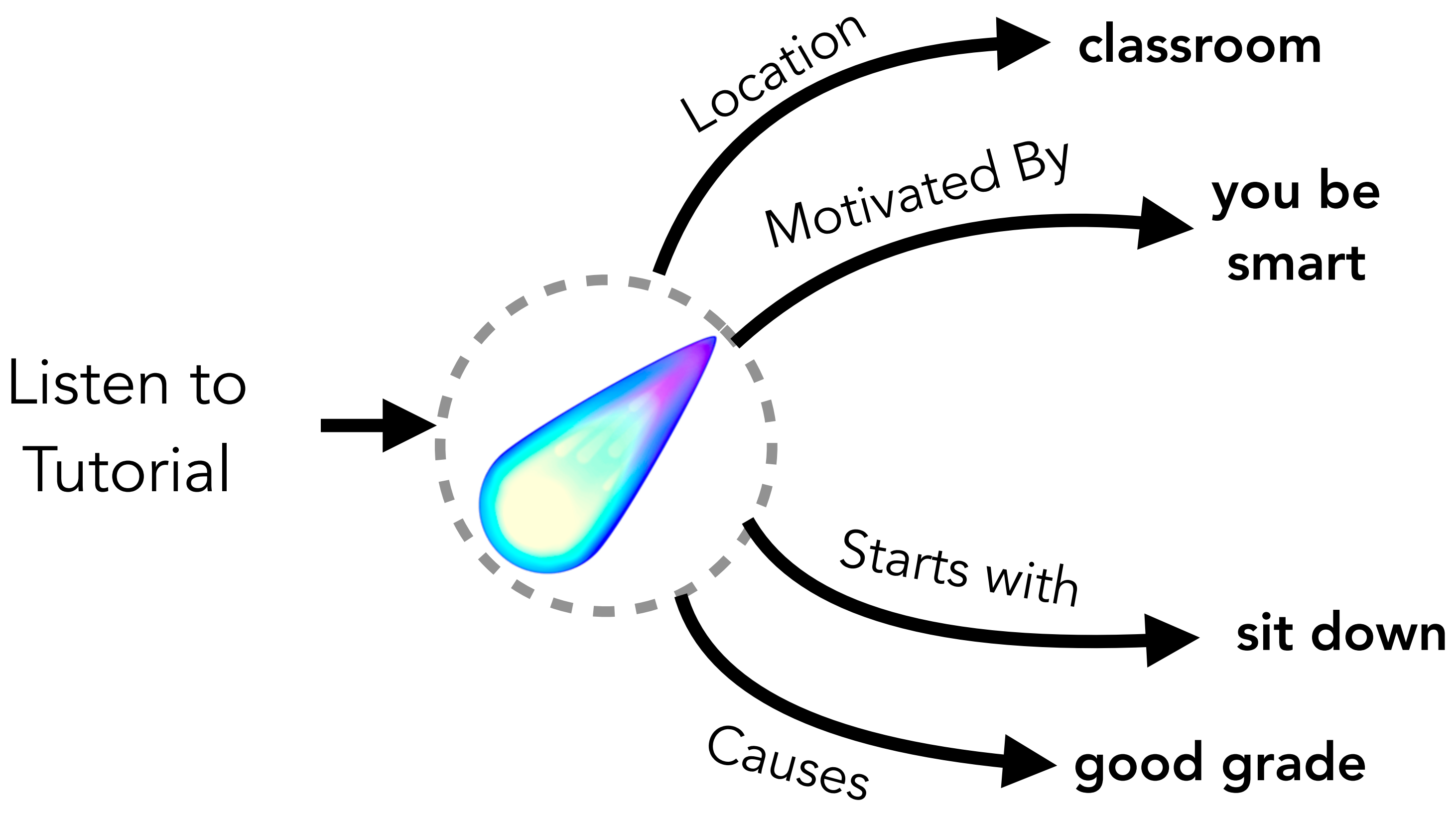
Limitations of Knowledge Models

- Base Self-supervised Model
  - biases, history





- Seed Knowledge Graph
  - bias, language, relations, schema
- Generation Algorithm
  - diversity, mode collapse



COMET

Bosselut et al. 2019 @ ACL 2019

Sarcasm generation

Chakrabarty et al. 2020 @ ACL 2020

Therapy Chabot

Kearns et al. 2020 @ CHI EA 2020

Personalized Dialogue

Majumder et al. 2020 @ EMNLP 2020

Simile generation

Chakrabarty et al. 2020 @ EMNLP 2020

Text-Based Games

Dambekodi et al. 2020 @ arXiv:2012.02757

Automated Storytelling

Ammanabrolu et al. 2021 @ AAAI 2021

Sarcasm generation
Chakrabarty et al. 2020 @ ACL 2020

Therapy Chabot
Kearns et al. 2020 @ CHI EA 2020

Personalized Dialogue
Majumder et al. 2020 @ EMNLP 2020

Simile generation
Chakrabarty et al. 2020 @ EMNLP 2020

Text-Based Games
Dambekodi et al. 2020 @ arXiv:2012.02757

Automated Storytelling
Ammanabrolu et al. 2021 @ AAAI 2021

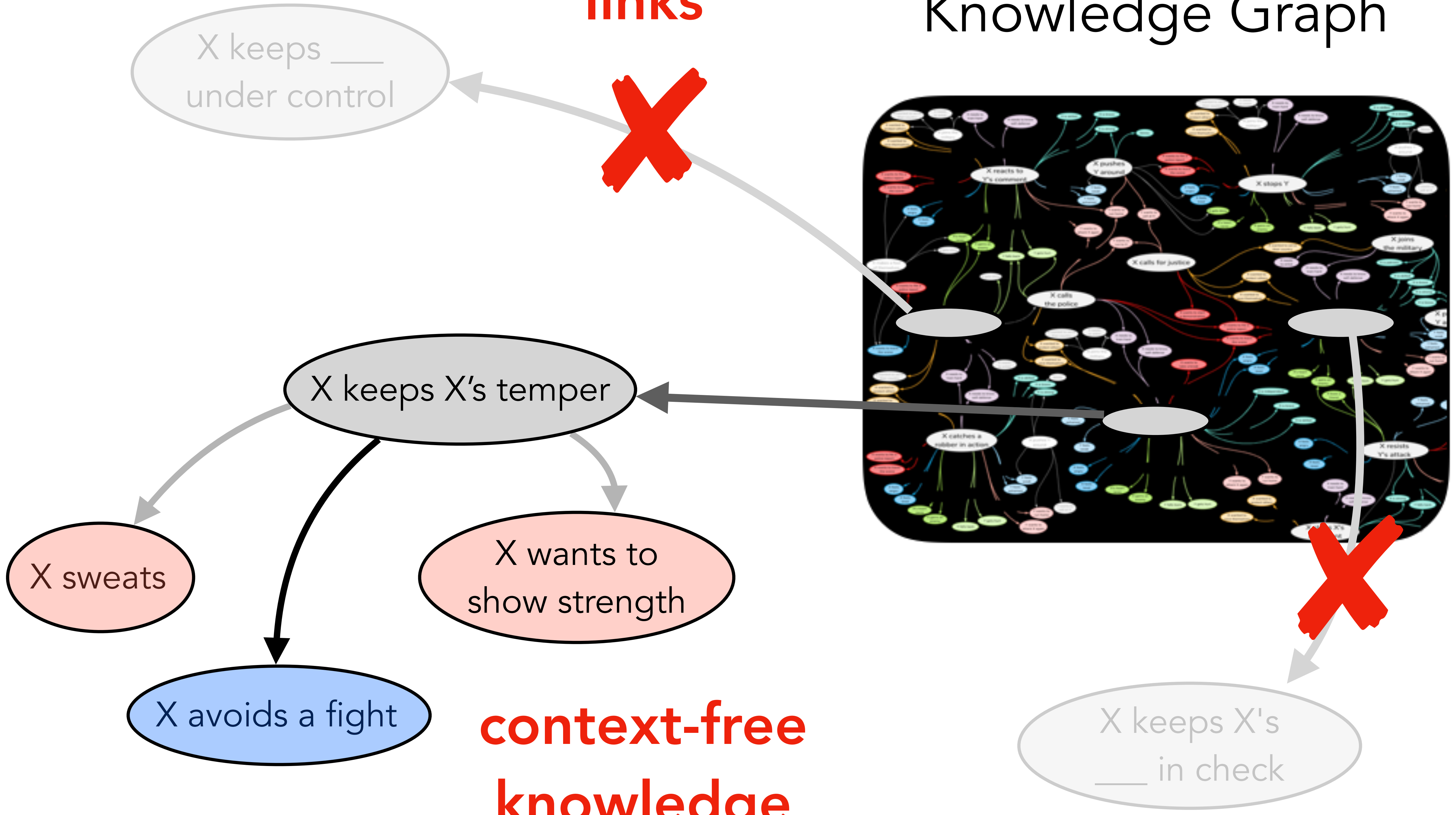
Where else can commonsense knowledge access improve our systems?

Static vs. Dynamic

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

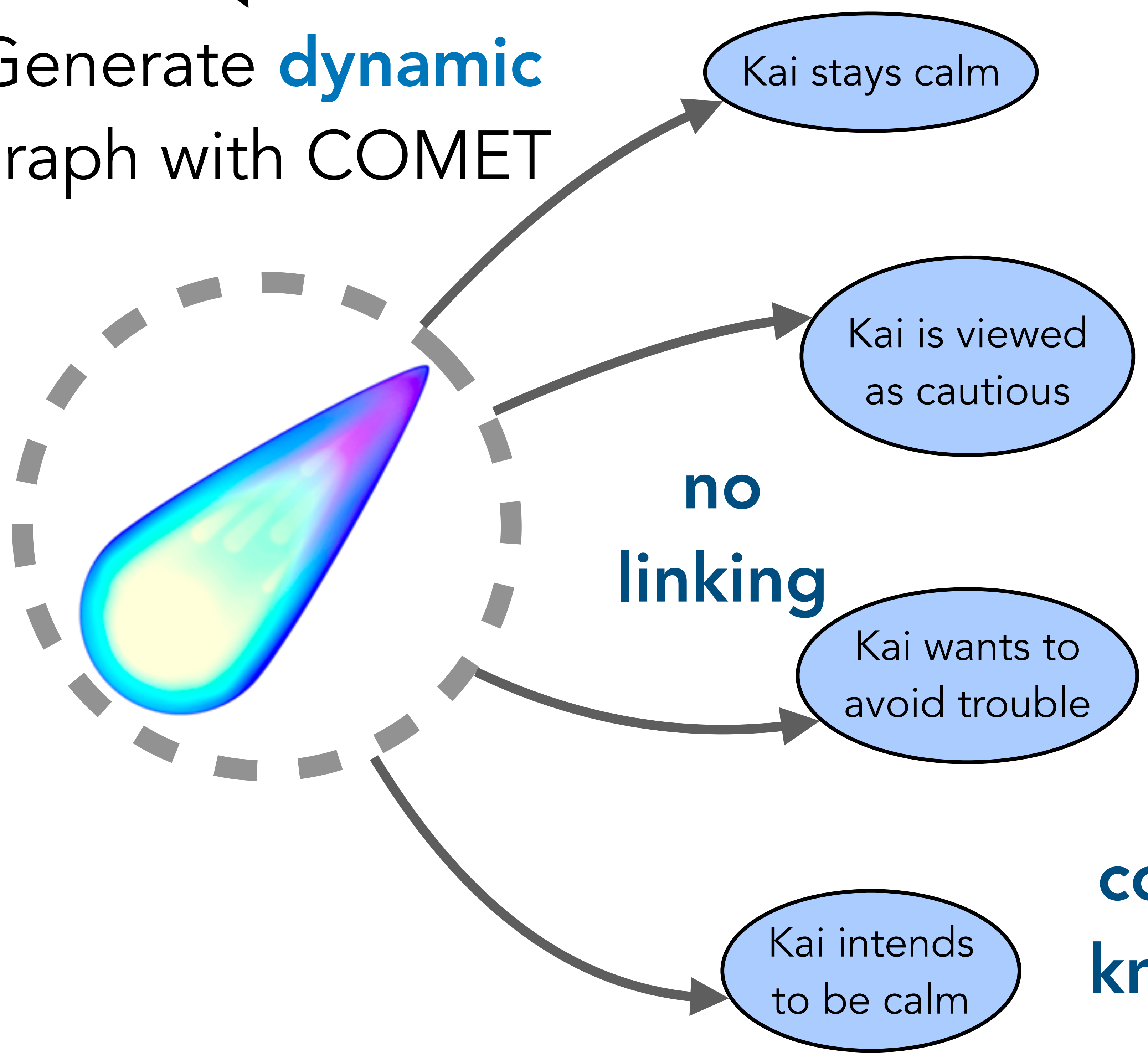
Link to **static** Knowledge Graph

bad links



context-free knowledge

Generate **dynamic** graph with COMET



no linking

contextual knowledge

Bosselut et al., AAAI 2021

Language Models as CSKBs?

- Large-scale language models encode a lot of commonsense knowledge **implicitly**, but it's not **not directly accessible**
- We can develop methods to extract it, but they need to be **adaptable, robust** and **efficient**
- Need to rethink how we design commonsense knowledge graphs if **transfer from language models** is a new use case

Building Commonsense Knowledge Bases

Quality of KB is important!

Careful validation is critical so that LMs can learn from precise examples.

More varieties of relations!

Represent a wide range of commonsense relationships (Use prompts for fast adaptation)

Focus on textually less explicit examples

These are less likely to be known by LMs, thus more impactful in knowledge transfer

Hwang et al., AAAI 2021

References & Additional Reading

[1] *Abductive commonsense reasoning*. Chandra Bhagavatula, Ronan Le Bras, Chaitanya Malaviya, Keisuke Sakaguchi, Ari Holtzman, Hannah Rashkin, Doug Downey, Scott Yih, Yejin Choi. ICLR 2020.

[2] *Dynamic Neuro-Symbolic Knowledge Graph Construction for Zero-shot Commonsense Question Answering*. Antoine Bosselut, Ronan Le Bras, Yejin Choi. AAAI 2021.

[3] *COMET: Commonsense Transformers for Automatic Knowledge Graph Construction*. Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, Yejin Choi. ACL 2019.

[4] *Commonsense knowledge mining from pretrained models*. Joshua Feldman, Joe Davison, Alexander Rush. EMNLP 2019.

[5] *On the Existence of Tacit Assumptions in Neural Language Models*. Nathaniel Weir, Adam Poliak, and Benjamin Van Durme. CogSci 2020.

[6] *(COMET-) ATOMIC$^{20}_{20}$: On Symbolic and Neural Commonsense Knowledge Graphs.* Jena Hwang*, Chandra Bhagavatula*, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, Yejin Choi. AAAI 2021.

[7] *Commonsense knowledge base completion*. Xiang Li, Aynaz Taheri, Lifu Tu, Kevin Gimpel. ACL 2016.

[8] *Understanding Few-shot Commonsense Knowledge Models*. Jeff Da, Ronan Le Bras, Ximing Lu, Yejin Choi, Antoine Bosselut. arXiv 2021.

[9] *Commonsense Knowledge Base Completion with Structural and Semantic Context.* Chaitanya Malaviya, Chandra Bhagavatula, Antoine Bosselut, Yejin Choi. AAAI 2020.

[10] *Language models as knowledge bases?* Fabio Petroni, Tim Rocktaschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander Miller. EMNLP 2019.

[11] *Commonsense knowledge base completion and generation*. Itsumi Saito, Kyosuke Nishida, Hisako Asano, Junji Tomita. CoNLL 2018.

[12] *oLMpics -- On what Language Model Pre-training Captures.* Alon Talmor, Yanai Elazar, Yoav Goldberg, and Jonathan Berant. TACL 2020.

References & Additional Reading

[13] *Unsupervised Commonsense Question Answering with Self-Talk.* Vered Shwartz, Peter West, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. EMNLP 2020.

[14] Do Neural Language Representations Learn Physical Commonsense? Maxwell Forbes, Ari Holtzman, and Yejin Choi. CogSci 2019.

[15] *Connecting the dots: A knowledgeable path generator for commonsense question answering*. Pei-Feng Wang, Nanyun Peng, Pedro A. Szekely, Xiang Ren. Findings of EMNLP 2020.

[16] *Translating embeddings for modeling multi-relational data.* Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, Oksana Yakhnenko. NeurIPS 2013.

[17] *Relation extraction with matrix factorization and universal schemas.* Sebastian Riedel, Limin Yao, Andrew McCallum, Benjamin M. Marlin. NAACL 2013.

[18] *Representing text for joint embedding of text and knowledge bases.* Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, Michael Gamon. EMNLP 2015.

[19] *Complex embeddings for simple link prediction.* Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, Guillaume Bouchard. ICML 2016.

[20] *Embedding entities and relations for learning and inference in knowledge bases.* Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, Li Deng. ICLR 2015.

[21] *TransE: a novel embedding model of entities and relationships in knowledge bases.* Dat Quoc Nguyen, Kairit Sirts, Lizhen Qu, Mark Johnson. NAACL 2016.

[22] *Convolutional 2d knowledge graph embeddings.* Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, Sebastian Riedel. AAAI 2018.

[23] *VisualCOMET: Reasoning about the Dynamic Context of a Still Image*. Jae Sung Park, Chandra Bhagavatula, Roozbeh Mottaghi, Ali Farhadi, Yejin Choi. ECCV 2020.

[24] *How Can We Know What Language Models Know?* Zhengbao Jiang, Frank F. Xu, Jun Araki, Graham Neubig. TACL 2020.