Commonsense Reasoning in Natural Language Processing

Vered Shwartz

Guest Lecture, Deep Learning for NLP

The Deep Learning Revolution

The Deep Learning Revolution

Translation

Google's AI translation system is approaching human-level accuracy

The Deep Learning Revolution

Translation

Google's AI translation system is approaching human-level accuracy

Reading Comprehension

ALIBABA AI BEATS HUMANS IN READING-COMPREHENSION TEST

CHRISTINE CHOU | JULY 9, 2019

The Deep Learning Revolution

Translation

Google's AI translation system is approaching human-level accuracy

Reading Comprehension

ALIBABA AI BEATS HUMANS COMPREHENSION TEST

CHRISTINE CHOU | JULY 9, 2019

Chatbots

Artificial intelligence / Voice assistants

Your next doctor's appointment might be with an AI

A new wave of chatbots are replacing physicians and providing frontline medical advice—but are they as good as the real thing?

by **Will Douglas Heaven** October 16, 2018

The Deep Learning Revolution

Translation

Google's AI translation system is approaching human-level accuracy

Reading Comprehension

ALIBABA AI BEATS HUMANS COMPREHENSION TEST

CHRISTINE CHOU | JULY 9, 2019

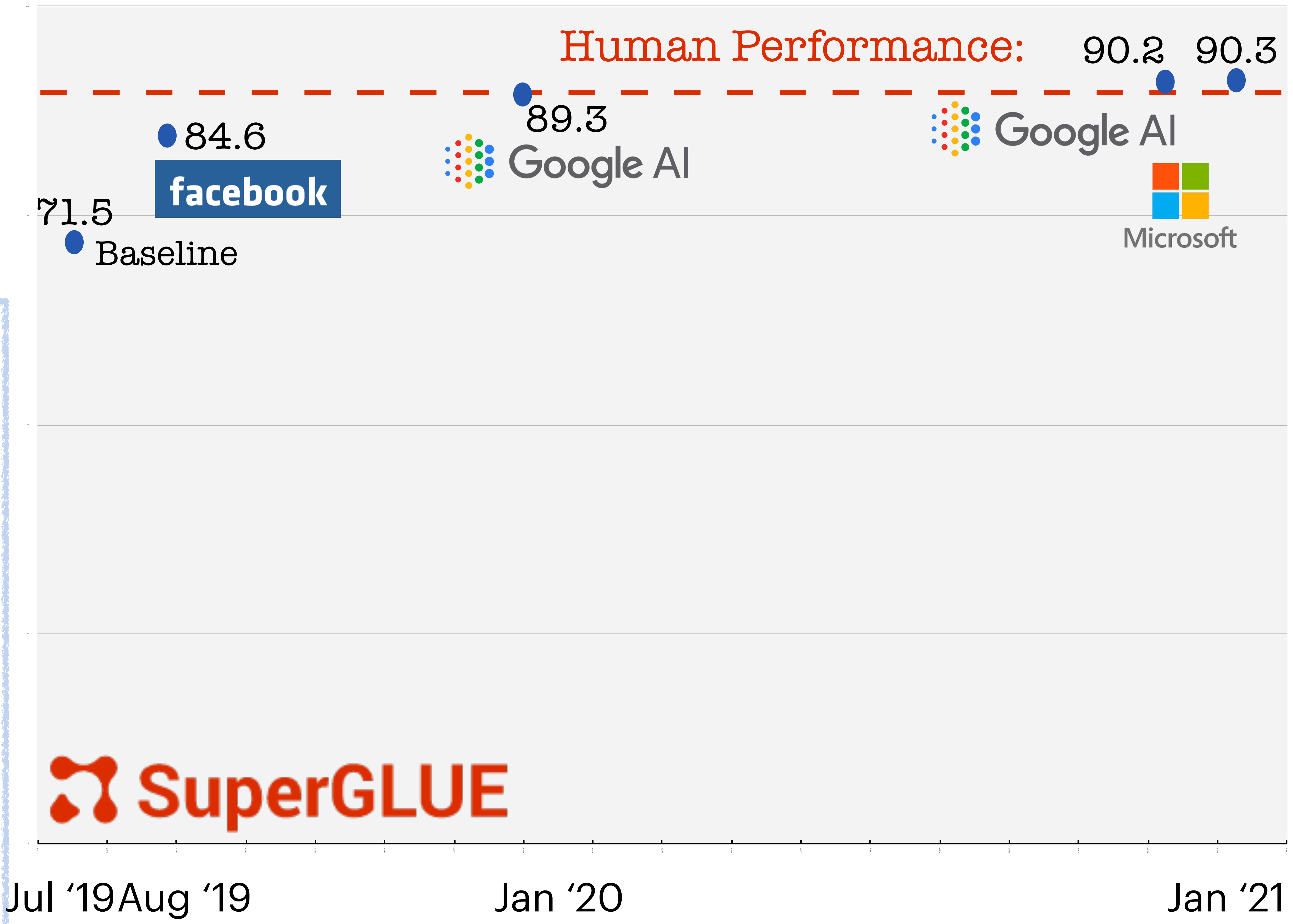
Chatbots

Artificial intelligence / Voice assistants

Your next doctor's appointment might be with an AI

A new wave of chatbots are replacing physicians and providing frontline medical advice—but are they as good as the real thing?

by Will Douglas Heaven   October 16, 2018



The Deep Learning Revolution

Does this mean language understanding is nearly solved?

Translation

Google's AI translation system is approaching human-level accuracy

Reading Comprehension

ALIBABA AI BEATS HUMANS IN READING-COMPREHENSION TEST

CHRISTINE CHOU | JULY 9, 2019

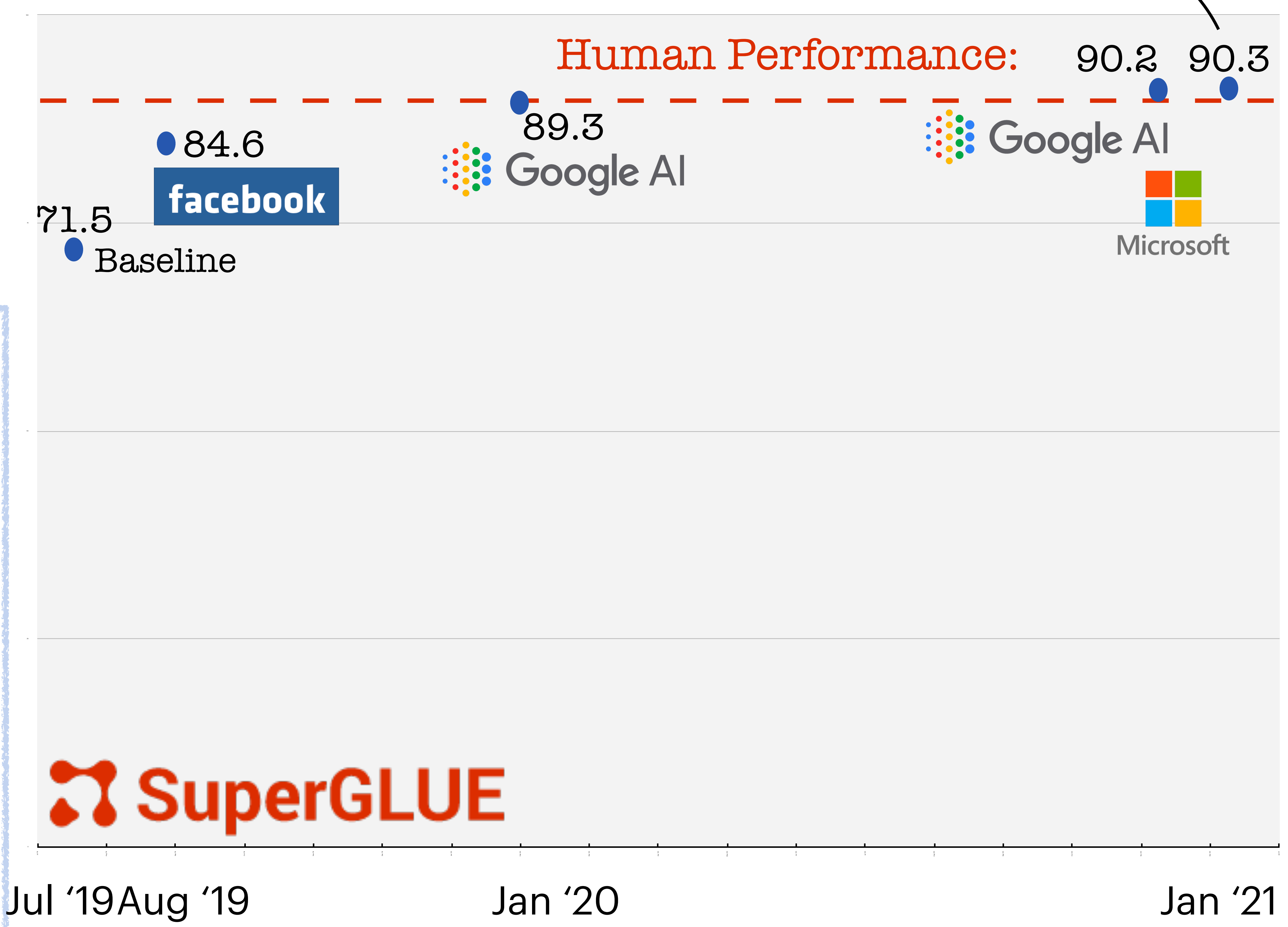
Chatbots

Artificial intelligence / Voice assistants

Your next doctor's appointment might be with an AI

A new wave of chatbots are replacing physicians and providing frontline medical advice—but are they as good as the real thing?

by Will Douglas Heaven October 16, 2018



The Deep Learning Revolution

Translation

Google's AI translation system is approaching human-level accuracy

Reading Comprehension

ALIBABA AI BEATS HUMANS IN READING-COMPREHENSION TEST

CHRISTINE CHOU | JULY 9, 2019

Chatbots

Artificial intelligence / Voice assistants

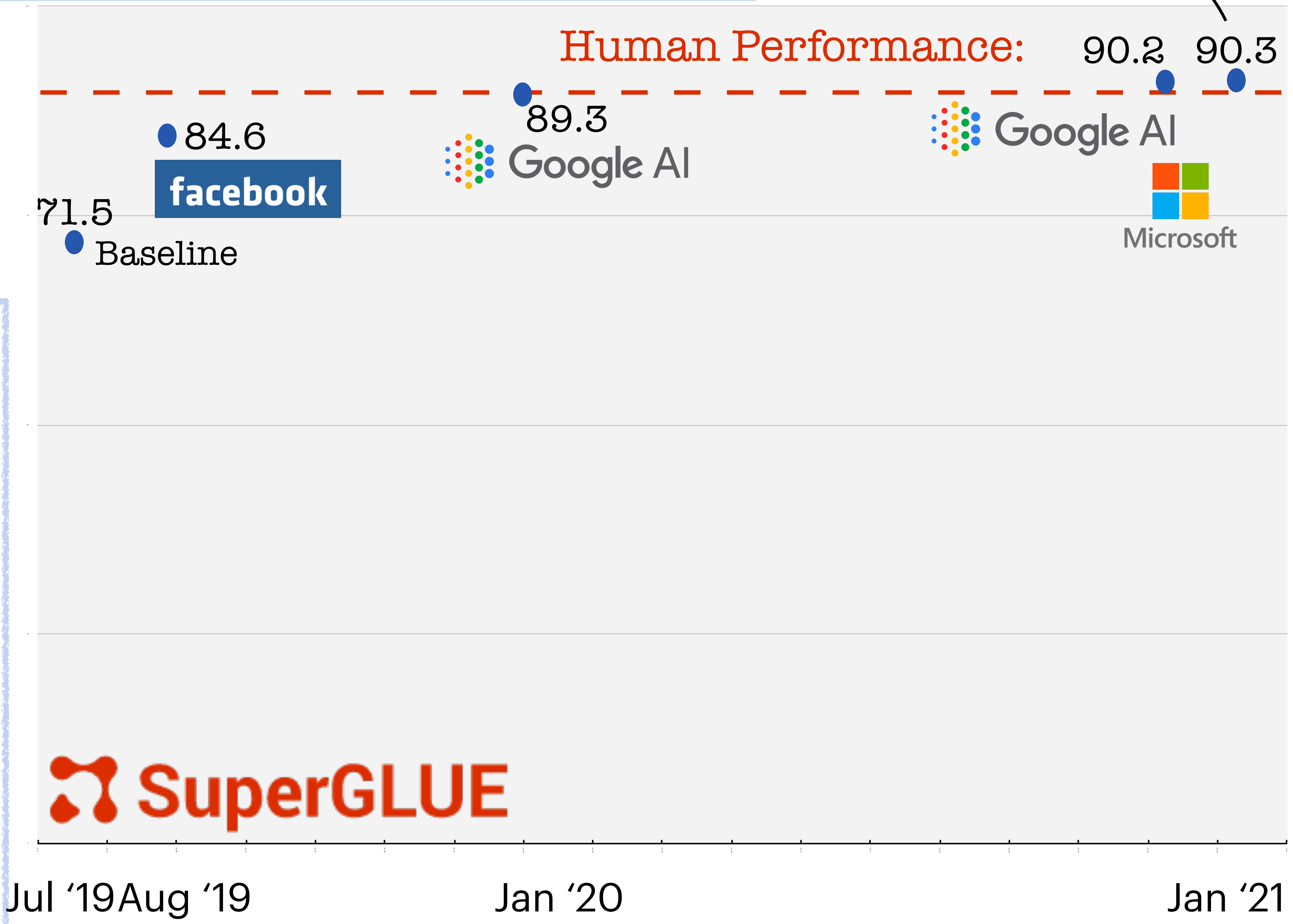
Your next doctor's appointment might be with an AI

A new wave of chatbots are replacing physicians and providing frontline medical advice—but are they as good as the real thing?

by Will Douglas Heaven October 16, 2018

Does this mean language understanding is nearly solved?

What are the remaining challenges?



Is Natural Language Understanding Nearly Solved?

Pre-training

Google AI

Is Natural Language Understanding Nearly Solved?

Pre-training

Google AI

- ✅ Syntax
- ✅ Word meanings
- ✅ Factual Knowledge
- ✅ ...

Is Natural Language Understanding Nearly Solved?

Pre-training

Google AI

Fine-tuning:

Language Model

✅ Syntax
✅ Word meanings
✅ Factual Knowledge
✅ ...

Is Natural Language Understanding Nearly Solved?

Pre-training

Google AI

WIKIPEDIA
The Free Encyclopedia

→ **Fine-tuning:**

Language Model

↑ ↑ ↑ ↑ ↑

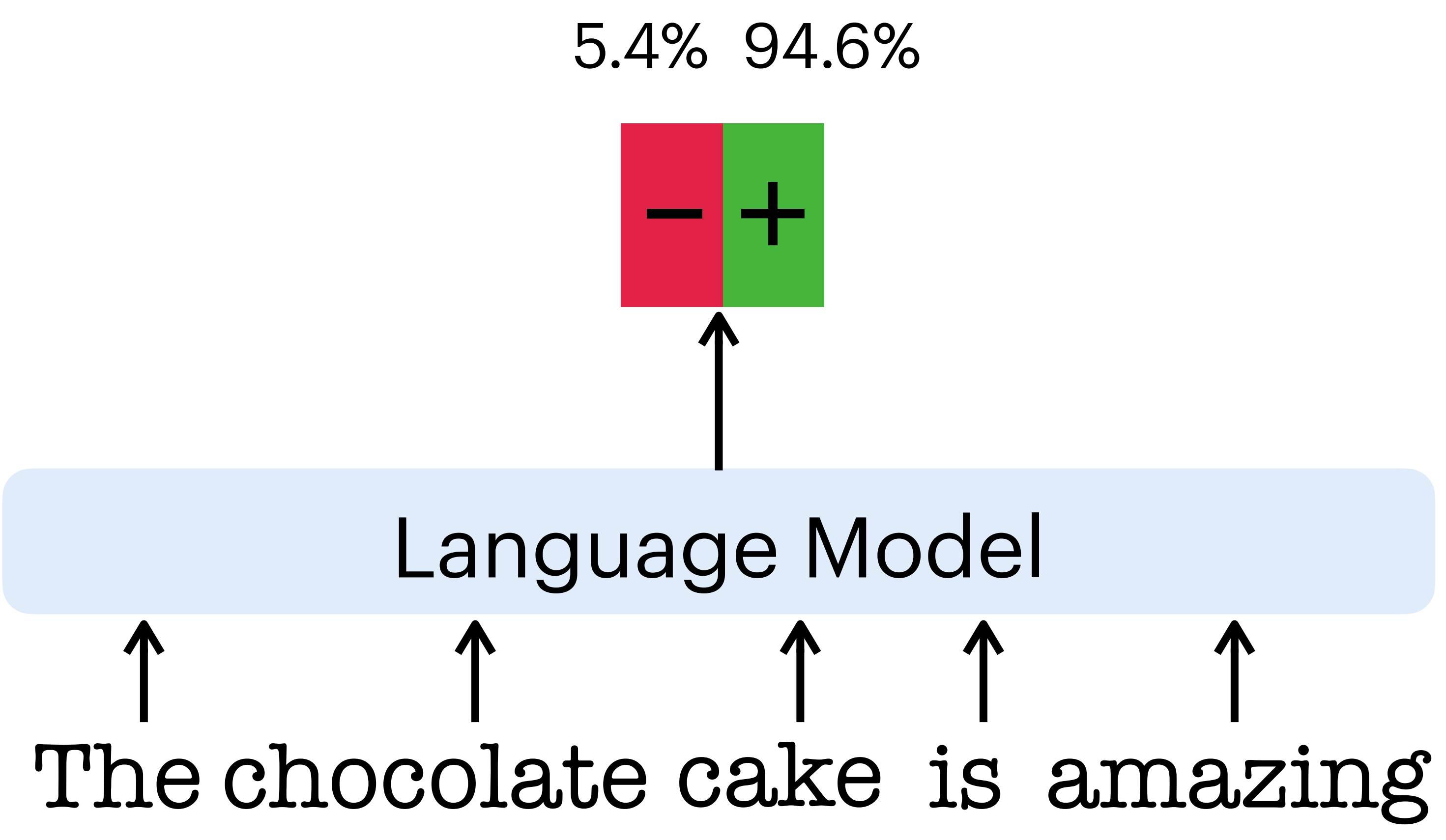
The chocolate cake is amazing

- ✅ Syntax
- ✅ Word meanings
- ✅ Factual Knowledge
- ✅ ...

Is Natural Language Understanding Nearly Solved?

Pre-training

Fine-tuning:



- ✅ Syntax
- ✅ Word meanings
- ✅ Factual Knowledge
- ✅ ...

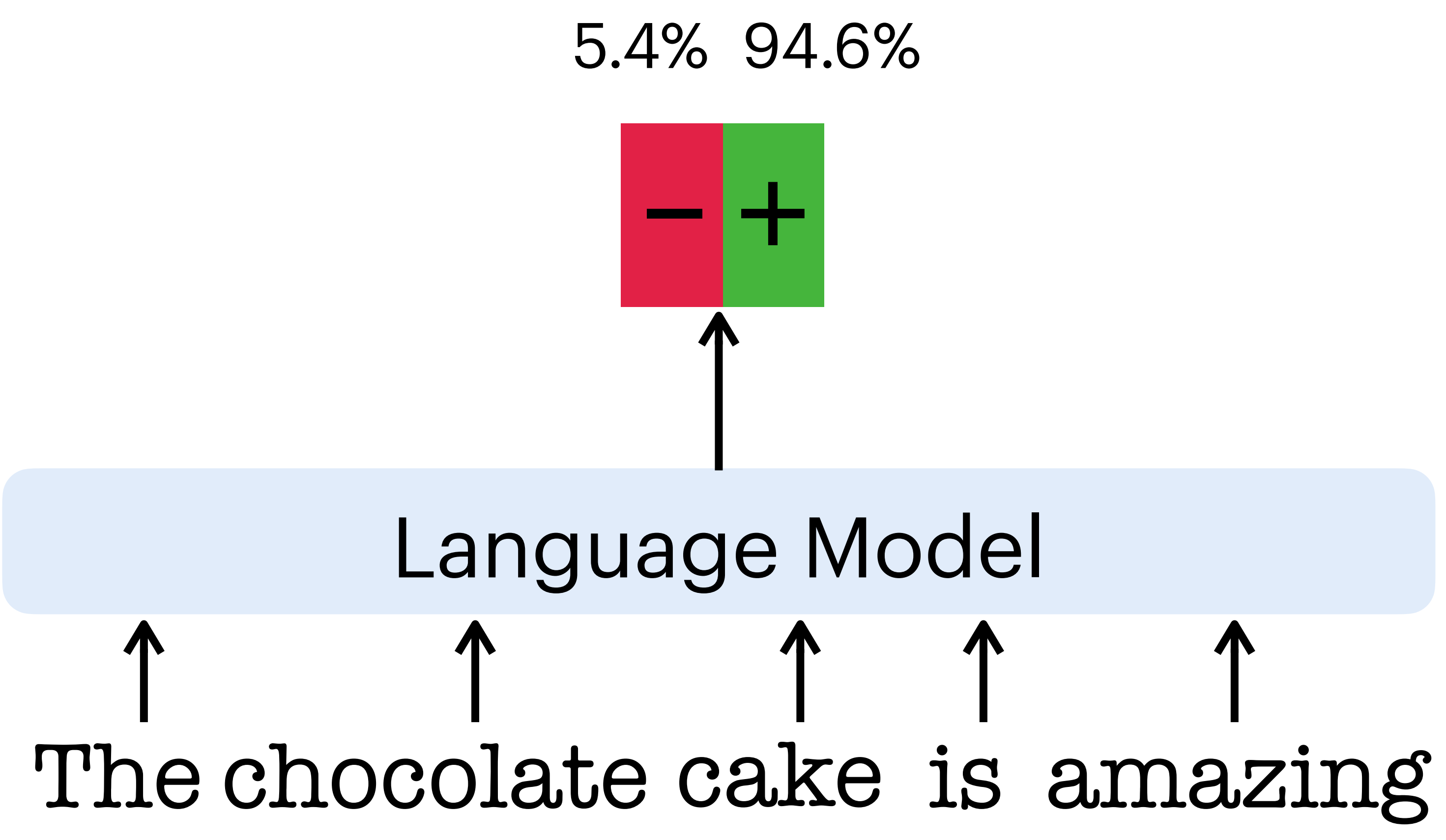
Is Natural Language Understanding Nearly Solved?

Pre-training

WIKIPEDIA
The Free Encyclopedia

Google AI

→ **Fine-tuning:**



✅ Syntax
✅ Word meanings
✅ Factual Knowledge
✅ ...

✅ Understanding the task
✅ Learning to solve the task

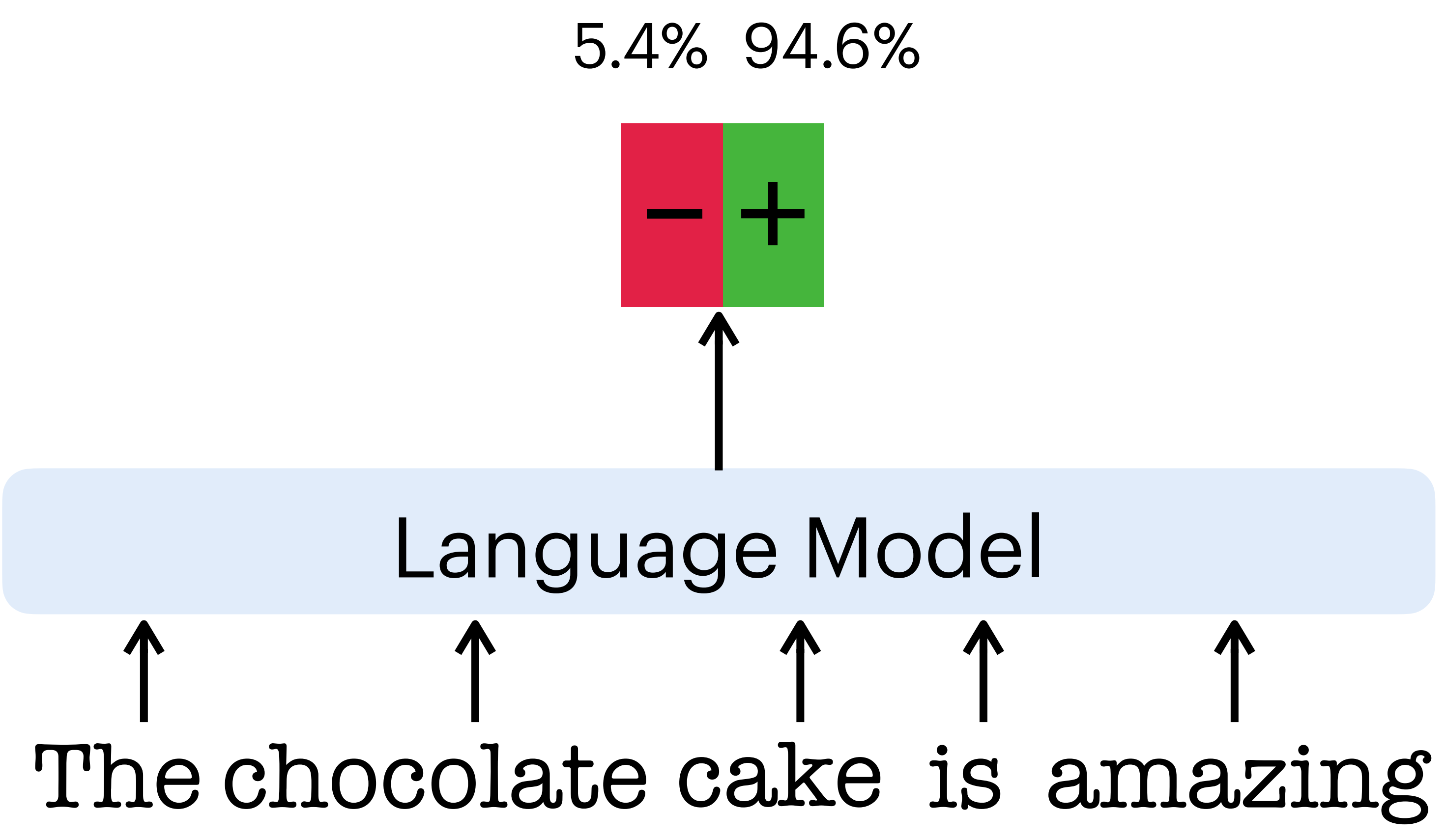
Is Natural Language Understanding Nearly Solved?

Pre-training

WIKIPEDIA
The Free Encyclopedia

Google AI

Fine-tuning:



✅ Syntax
✅ Word meanings
✅ Factual Knowledge
✅ ...

✅ Understanding the task
✅ Learning to solve the task

What are the remaining challenges?

❓ Generalization to unknown situations

Overfitting to Data-specific Spurious Correlations

Analyzing the Behavior of Visual Question Answering Models. Aishwarya Agrawal, Dhruv Batra, and Devi Parikh. EMNLP 2016.

Overfitting to Data-specific Spurious Correlations

How many zebras?

🤖: 2

Analyzing the Behavior of Visual Question Answering Models. Aishwarya Agrawal, Dhruv Batra, and Devi Parikh. EMNLP 2016.

Overfitting to Data-specific Spurious Correlations

How many zebras?

🤖: 2

How many giraffes? 2

How many zebras? 2

How many dogs? 2

Analyzing the Behavior of Visual Question Answering Models. Aishwarya Agrawal, Dhruv Batra, and Devi Parikh. EMNLP 2016.

Overfitting to Data-specific Spurious Correlations

How many zebras?

🤖: 2

How many giraffes? 2

How many zebras? 2

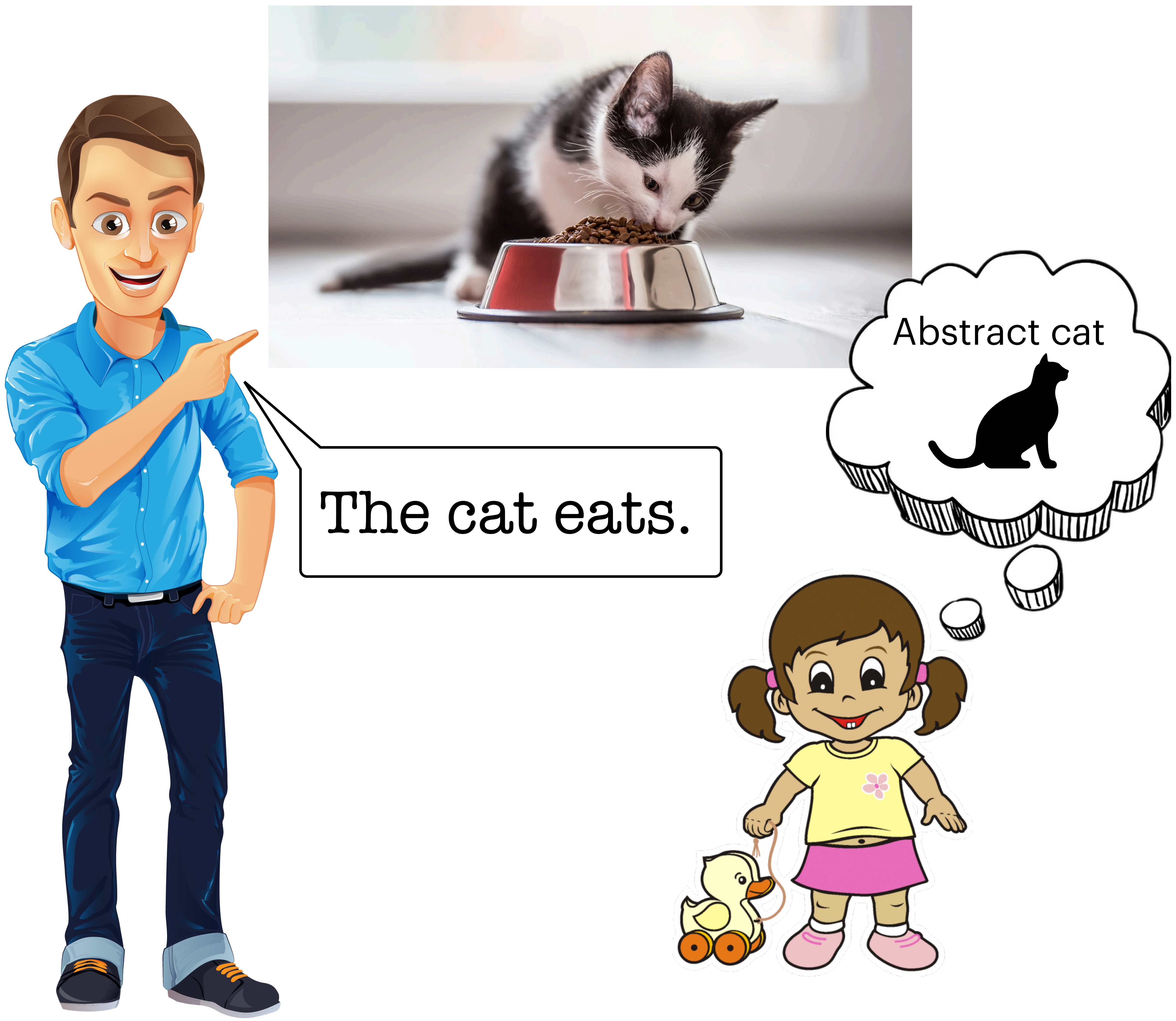
How many dogs? 2

...Solving *datasets* but not underlying *tasks*!

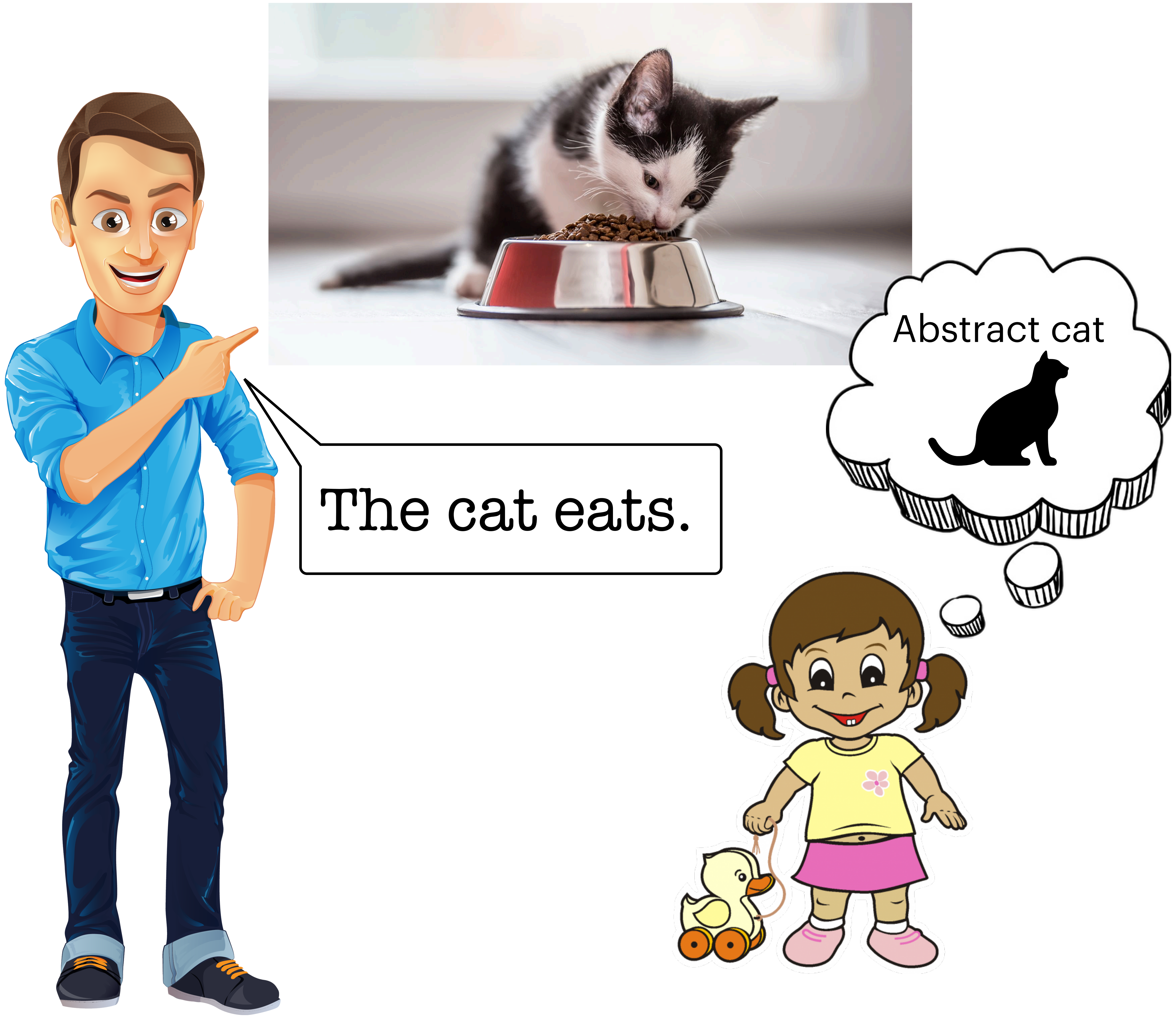
Analyzing the Behavior of Visual Question Answering Models. Aishwarya Agrawal, Dhruv Batra, and Devi Parikh. EMNLP 2016.

Humans generalize from few examples

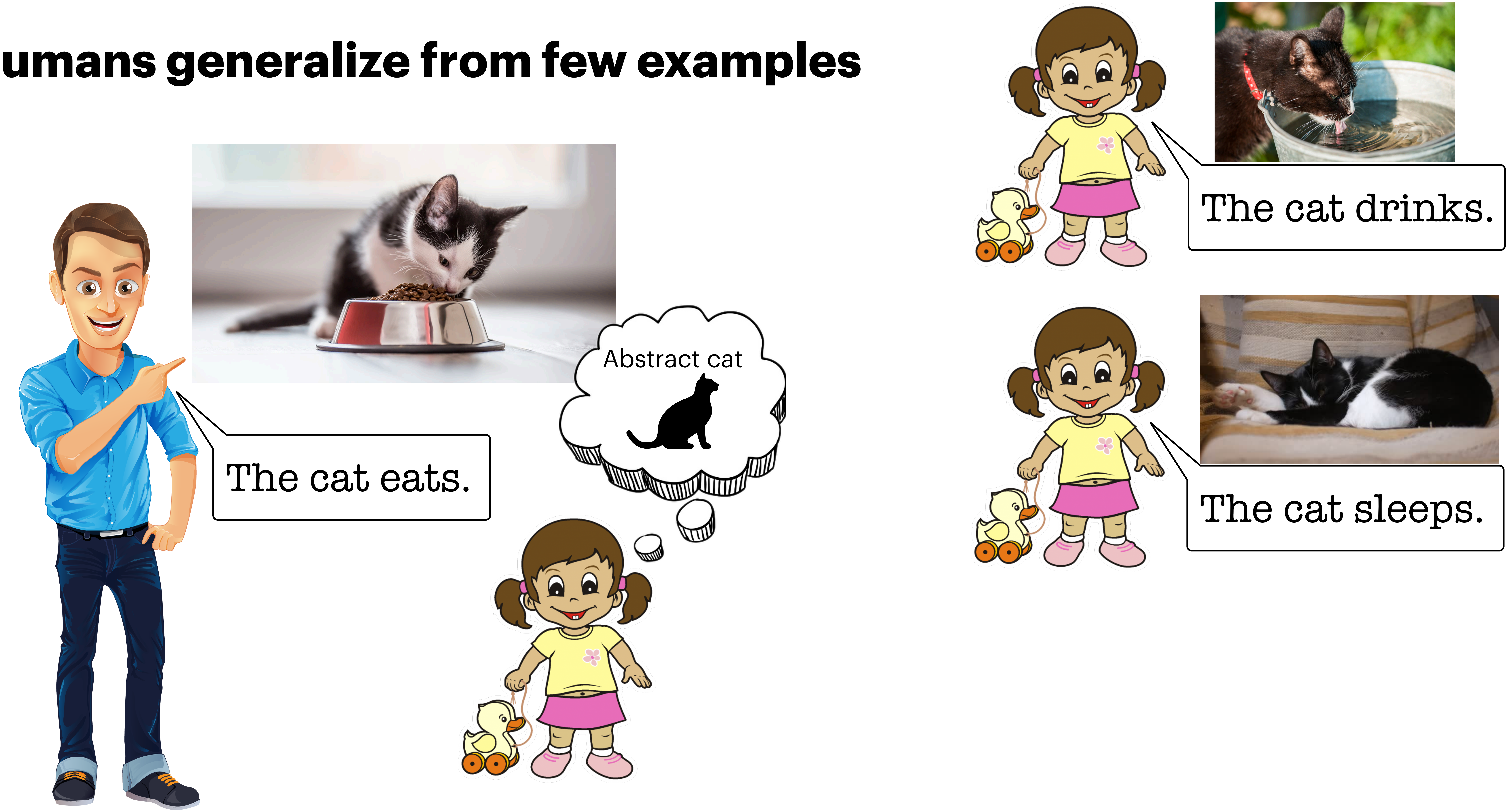
Humans generalize from few examples



Humans generalize from few examples



Humans generalize from few examples



Humans generalize from few examples

Commonsense Reasoning
in Natural Language Processing

Commonsense Reasoning

Commonsense Reasoning

Natural language is...

Commonsense Reasoning

Natural language is...

Ambiguous

Stevie Wonder announces he'll be having kidney surgery during London concert

Commonsense Reasoning

Natural language is...

Ambiguous

Stevie Wonder announces he'll be having kidney surgery during London concert

Q: When is the surgery?
A: During London concert ❌

Commonsense Reasoning

Natural language is...

Ambiguous

Stevie Wonder announces he'll be having kidney surgery during London concert

Q: When is the surgery?
A: During London concert ❌

🤔 Kidney surgery is performed under general anesthesia
🤔 People are unconscious under general anesthesia
🤔 Performing actions requires being conscious

Commonsense Reasoning

Natural language is...

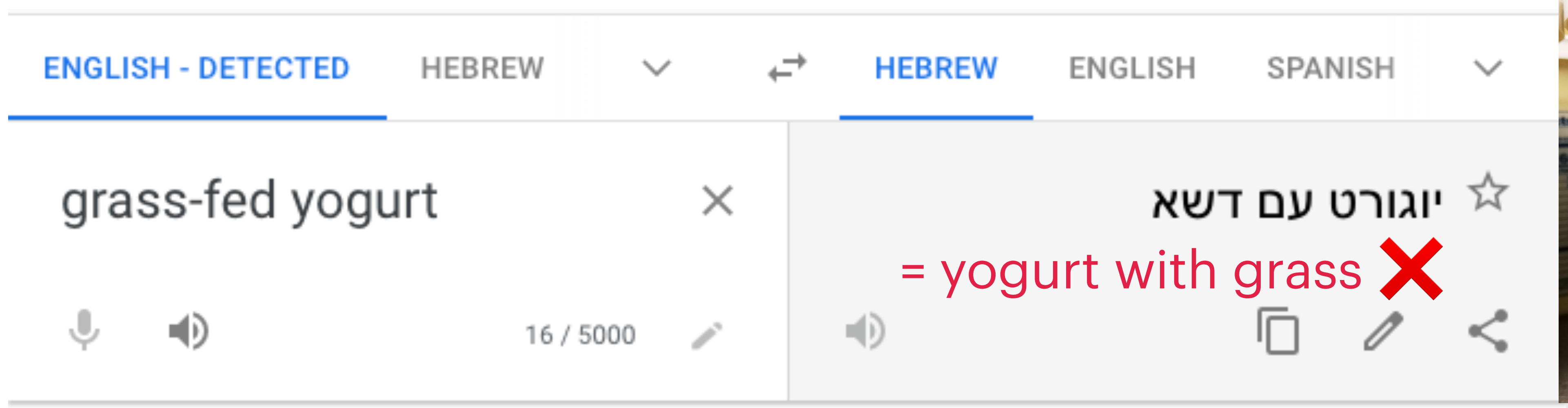
Ambiguous

Stevie Wonder announces he'll be having kidney surgery during London concert

Q: When is the surgery?
A: During London concert ❌

🤔 Kidney surgery is performed under general anesthesia
🤔 People are unconscious under general anesthesia
🤔 Performing actions requires being conscious

Under-Specified



Commonsense Reasoning

Natural language is...

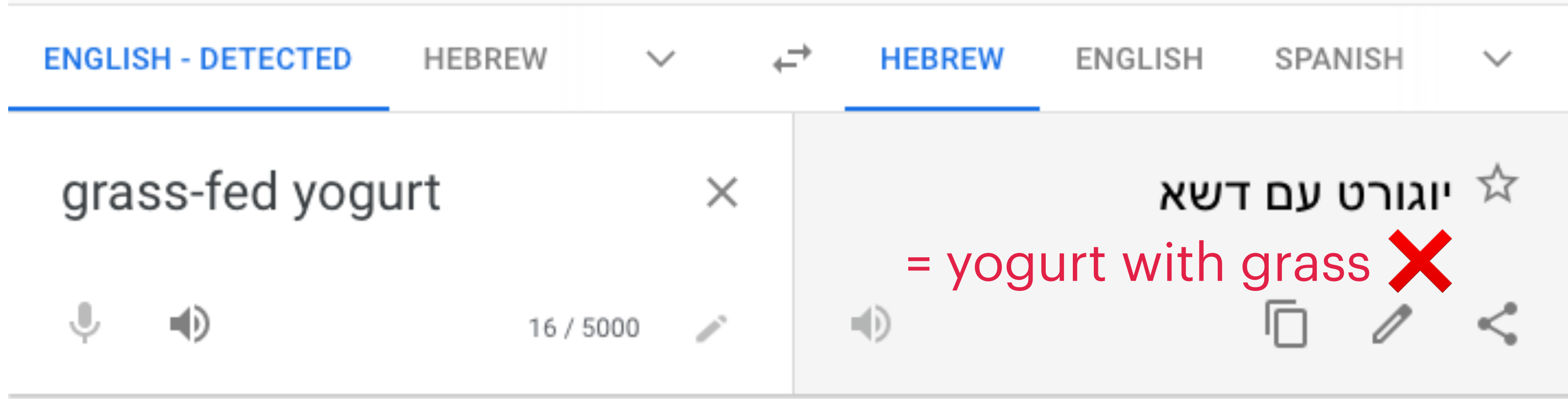
Ambiguous

Stevie Wonder announces he'll be having kidney surgery during London concert

Q: When is the surgery?
A: During London concert ❌

🤔 Kidney surgery is performed under general anesthesia
🤔 People are unconscious under general anesthesia
🤔 Performing actions requires being conscious

Under-Specified



🤔 Yogurt is typically made of cow milk
🤔 Cows eat grass

What is Commonsense?

The basic level of **practical knowledge** and **reasoning** concerning **everyday situations** and **events** that are **commonly** shared among **most** people.

Introductory Tutorial on Commonsense Reasoning. Maarten Sap, **Vered Shwartz**, Antoine Bosselut, Dan Roth, and Yejin Choi. ACL 2020.

https://web.njit.edu/~ronkowit/eliza.html

What is Commonsense?

It's a bad idea to touch a hot stove.

The basic level of **practical knowledge** and **reasoning** concerning **everyday situations** and **events** that are **commonly** shared among **most** people.

What is Commonsense?

The basic level of **practical knowledge** and **reasoning** concerning **everyday situations** and **events** that are **commonly** shared among **most** people.

It's a bad idea to touch a hot stove.

It's impolite to comment on people's weight.

Introductory Tutorial on Commonsense Reasoning. Maarten Sap, **Vered Shwartz**, Antoine Bosselut, Dan Roth, and Yejin Choi. ACL 2020.

https://web.njit.edu/~ronkowit/eliza.html

What is Commonsense?

The basic level of **practical knowledge** and **reasoning** concerning **everyday situations** and **events** that are **commonly** shared among **most** people.

It's a bad idea to touch a hot stove.

It's impolite to comment on people's weight.

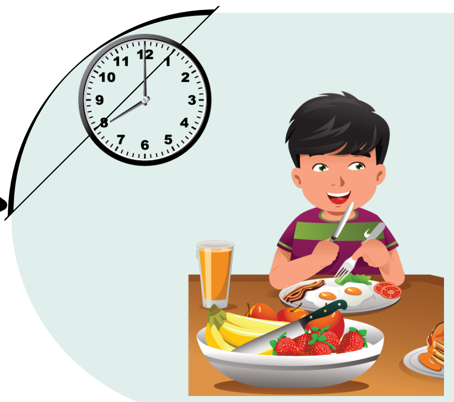

Eating dinner comes before going to bed.

What is Commonsense?

The basic level of **practical knowledge** and **reasoning** concerning **everyday situations** and **events** that are **commonly** shared among **most** people.

It's a bad idea to touch a hot stove.

It's impolite to comment on people's weight.



Eating dinner comes before going to bed.

…

Introductory Tutorial on Commonsense Reasoning. Maarten Sap, **Vered Shwartz**, Antoine Bosselut, Dan Roth, and Yejin Choi. ACL 2020.

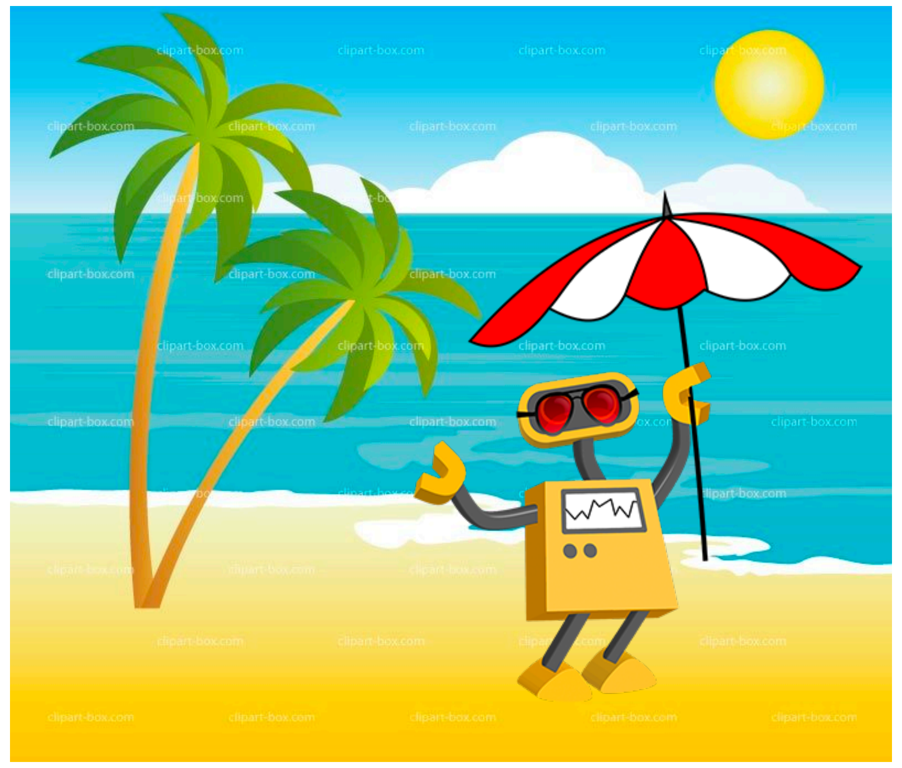
https://web.njit.edu/~ronkowit/eliza.html

Commonsense Timeline

1956

Commonsense Timeline

1956



1974


File F2
Directory F3
Disk F4
View F
TEXT
CALENDAR


File F2
Directory F3
Disk F4
View
TEXT
CALENDAR

Commonsense Timeline

- Reasoning by search → combinatorial explosion
- Lack of commonsense knowledge and reasoning abilities
- Rigidity of symbolic reasoning

1956

1974

1980

Commonsense Timeline

1956

1974

1980

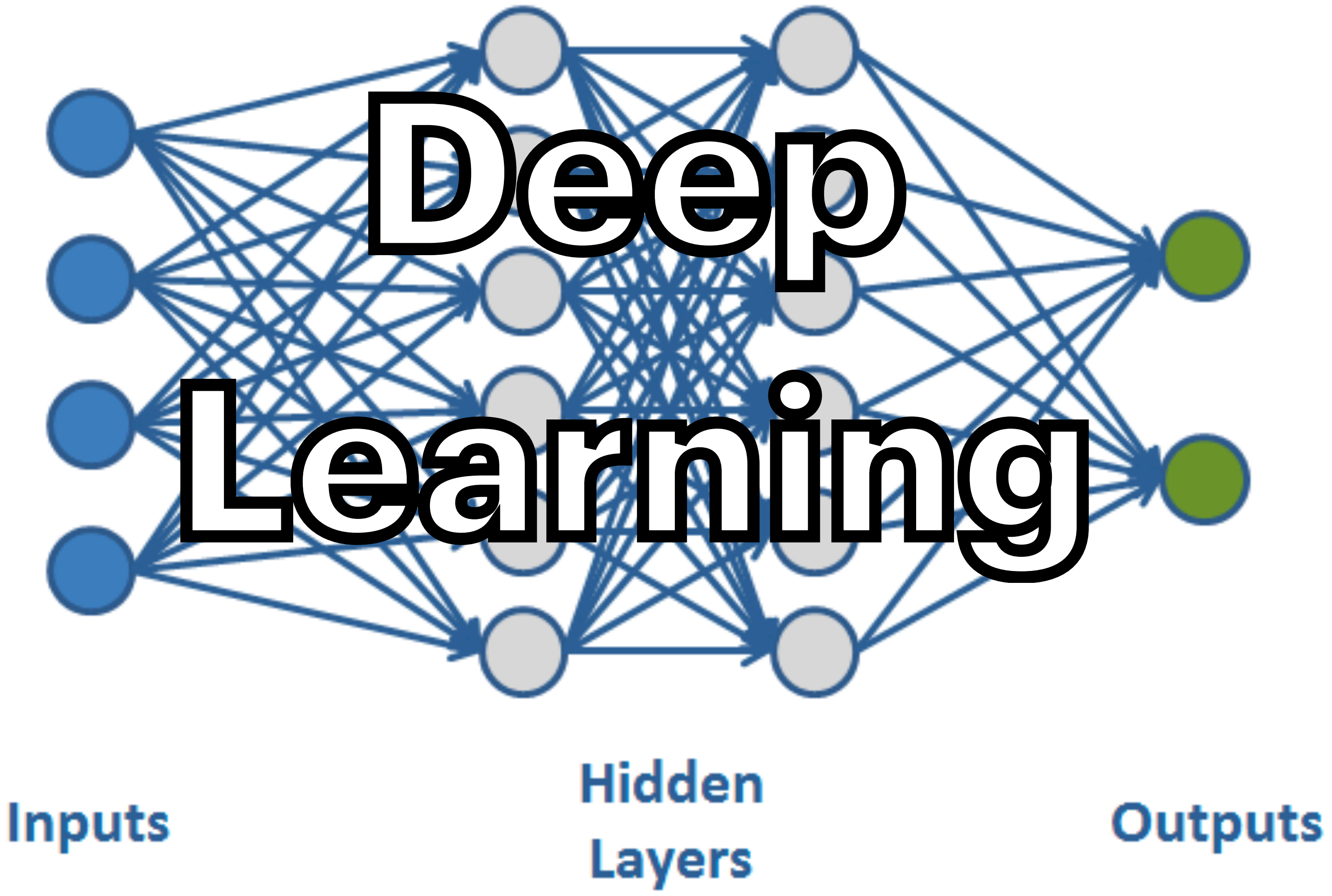
- Reasoning by search → combinatorial explosion
- Lack of commonsense knowledge and reasoning abilities
- Rigidity of symbolic reasoning

- weak computing power
- not enough data (and no crowdsourcing)
- weaker computational models

Commonsense Timeline

- Expert systems
- Slow progress

Commonsense Timeline



1956

1974

1980

2011

Path to commonsense?

Brute force larger networks with deeper layers?

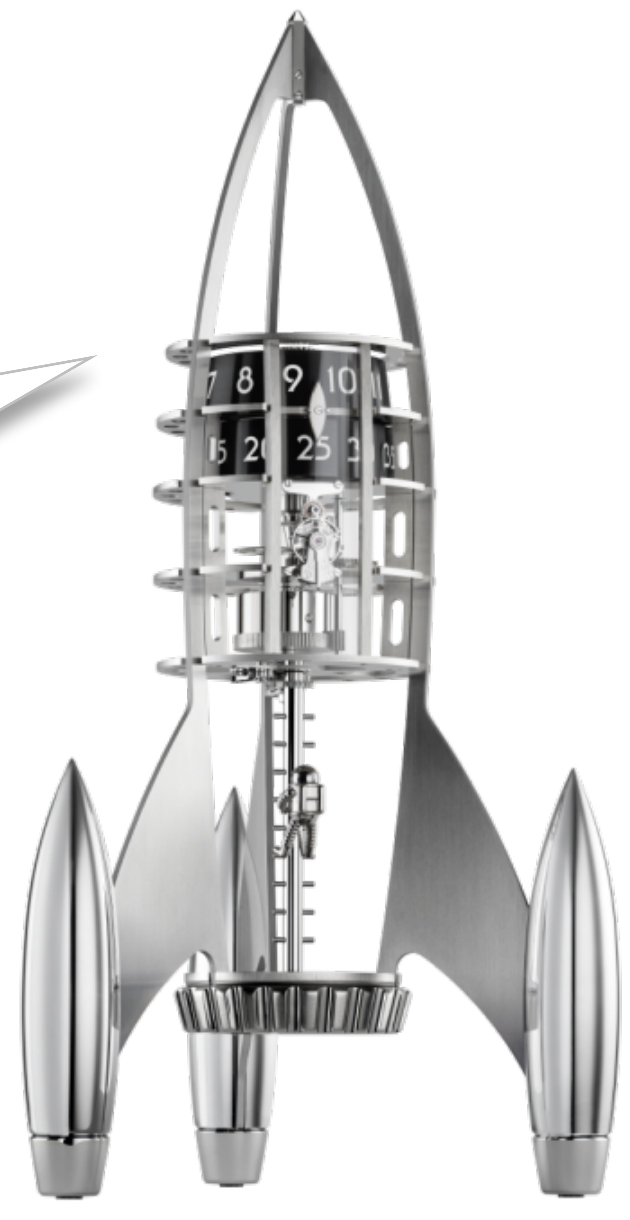
Path to commonsense?

Brute force larger networks with deeper layers?

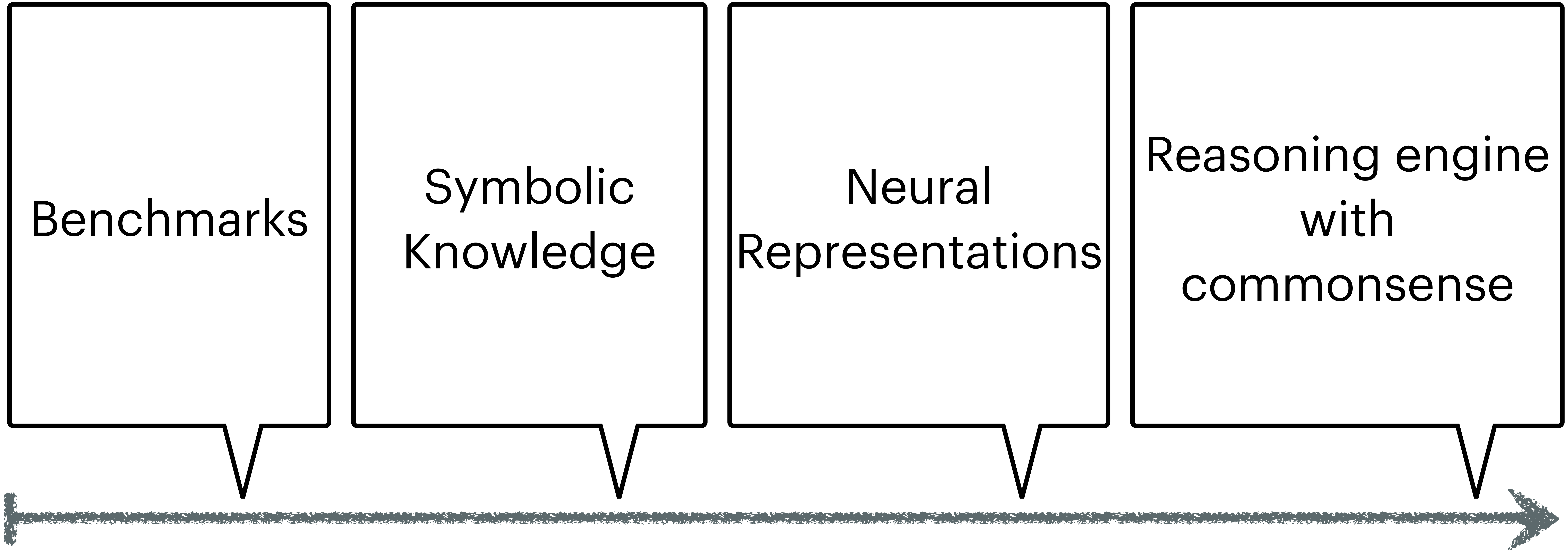
Path to commonsense?

Brute force larger networks with deeper layers?

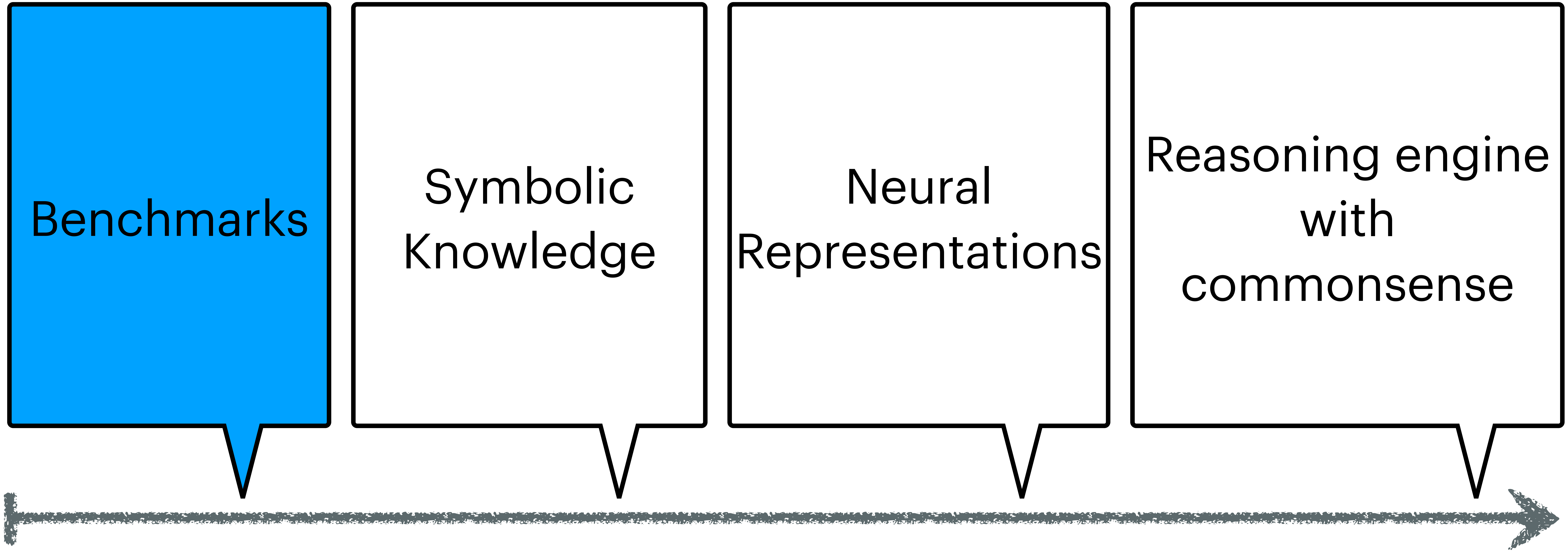
You don’t reach the moon
by making the tallest building in the world taller



Path to commonsense



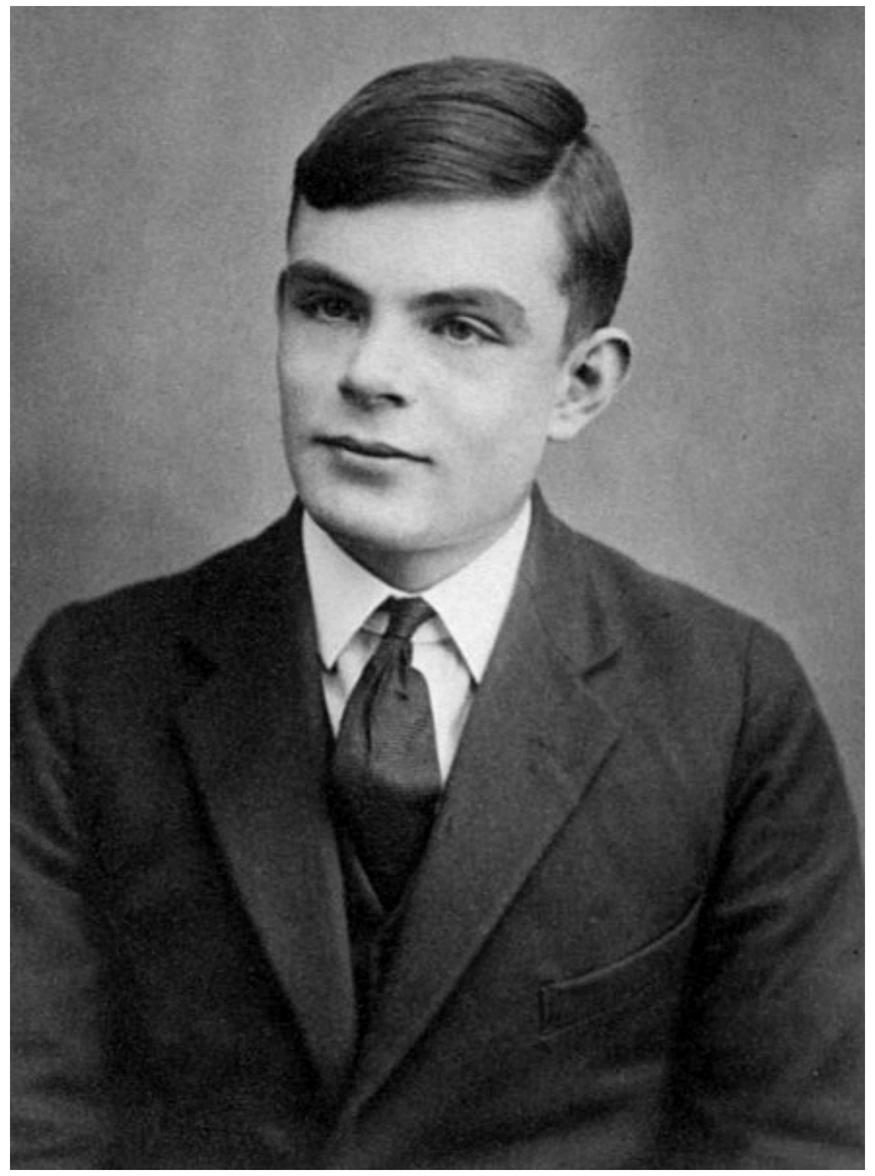
Path to commonsense



1950: Turing Test

~~Can machines think?~~

Can a human judge distinguish between a human and a machine following a short conversation with each?

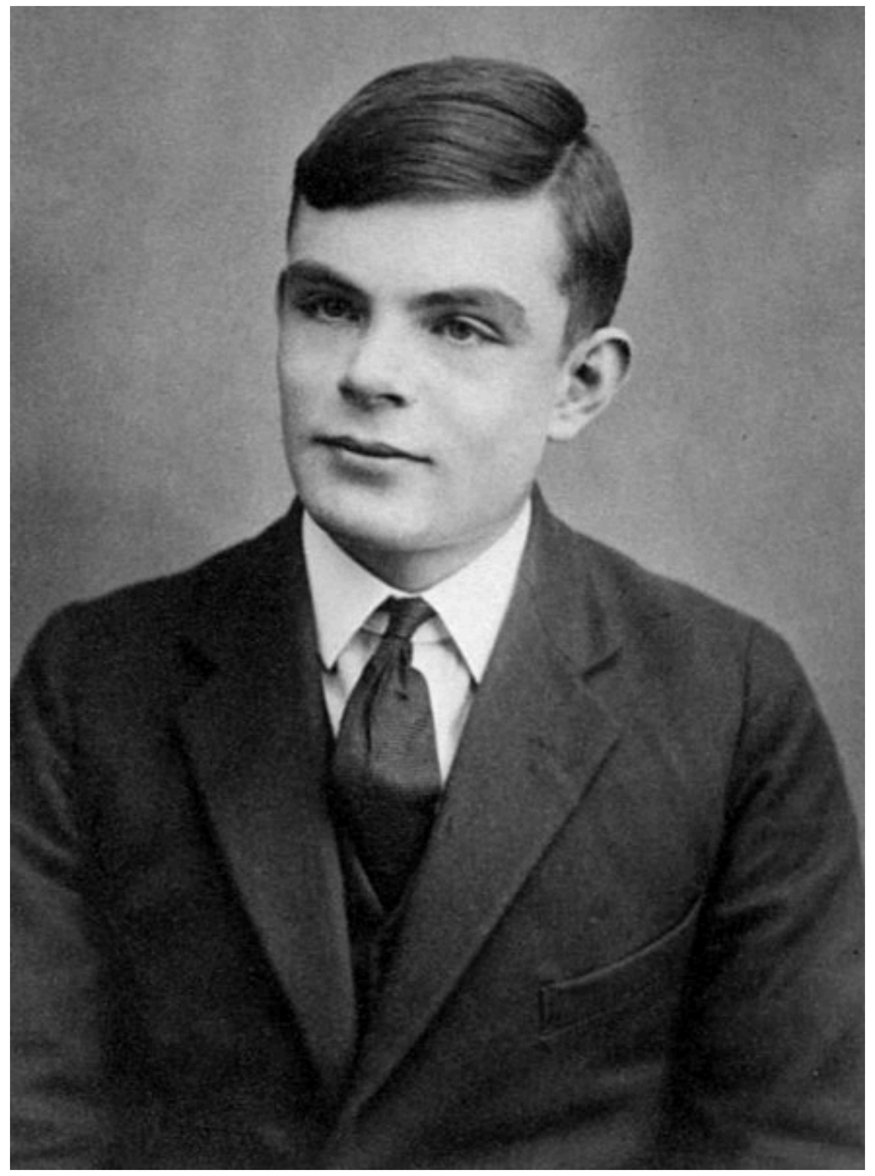


Alan Turing

1950: Turing Test

~~Can machines think?~~

Can a human judge distinguish between a human and a machine following a short conversation with each?



Alan Turing

- Loebner Prize (since 1990s)
- Winner of 2014: a bot named “Eugene Goostman”, simulating a 13-year-old Ukrainian boy, won
- Recommended reading: https://artistdetective.wordpress.com/, “The most human human”

Winograd Schema Challenge (WSC)

The winograd schema challenge. Hector Levesque, Ernest Davis, and Leora Morgenstern. AAAI 2012.

Winograd Schema Challenge (WSC)

The city councilmen refused the demonstrators a permit because *they* **advocated** violence. Who is “*they*”?

(a) The city councilmen
(b) The demonstrators

The winograd schema challenge. Hector Levesque, Ernest Davis, and Leora Morgenstern. AAAI 2012.

Winograd Schema Challenge (WSC)

The city councilmen refused the demonstrators a permit because *they* **advocated** violence. Who is "*they*"?

(a) The city councilmen
(b) The demonstrators

The winograd schema challenge. Hector Levesque, Ernest Davis, and Leora Morgenstern. AAAI 2012.

Winograd Schema Challenge (WSC)

The city councilmen refused the demonstrators a permit because *they* **advocated** violence. Who is "*they*"?

(a) The city councilmen
(b) The demonstrators

The city councilmen refused the demonstrators a permit because *they* **feared** violence. Who is "*they*"?

(a) The city councilmen
(b) The demonstrators

The winograd schema challenge. Hector Levesque, Ernest Davis, and Leora Morgenstern. AAAI 2012.

Winograd Schema Challenge (WSC)

The city councilmen refused the demonstrators a permit because *they* **advocated** violence. Who is "*they*"?

(a) The city councilmen
(b) The demonstrators

The city councilmen refused the demonstrators a permit because *they* **feared** violence. Who is "*they*"?

(a) The city councilmen
(b) The demonstrators

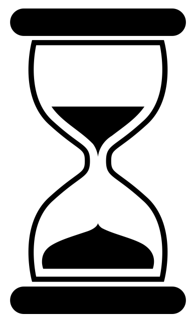
The winograd schema challenge. Hector Levesque, Ernest Davis, and Leora Morgenstern. AAAI 2012.

More benchmarks

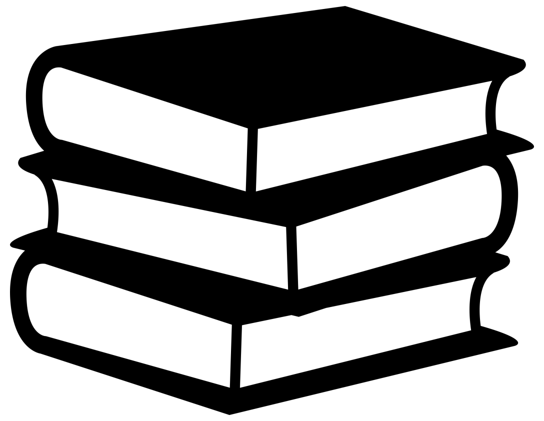
- Naïve Psychology
- ROC story
- Social IQa
- Physical IQa
- HellaSwag
- SWAG
- WSC
- COPA

- Abductive NLI

- VCR
- WinoGrande
- CommonsenseQA
- JHU Ordinal Commonsense



- MCTaco
- ReCORD
- CosmosQA



- MultiRC

Even more benchmarks: https://commonsense.run/

More benchmarks

Social commonsense

Naïve Psychology

ROC story

Social IQa

WSC

COPA

VCR

WinoGrande

Physical IQa

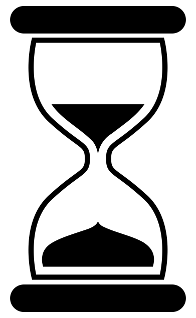
HellaSwag

Abductive NLI

SWAG

CommonsenseQA

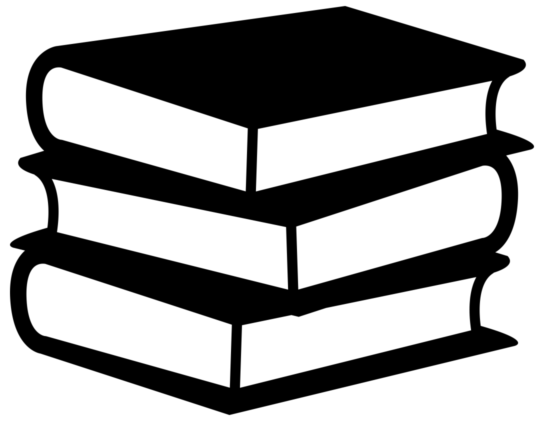
JHU Ordinal Commonsense



MCTaco

ReCORD

CosmosQA



MultiRC

Even more benchmarks: https://commonsense.run/

More benchmarks

Social commonsense

- Naïve Psychology
- ROC story
- Social IQa
- WSC
- COPA
- VCR
- WinoGrande

Abductive NLI

CommonsenseQA

Physical commonsense

- Physical IQa
- HellaSwag
- SWAG
- JHU Ordinal Commonsense

MCTaco

ReCORD

CosmosQA

MultiRC

Even more benchmarks: https://commonsense.run/

More benchmarks

Social commonsense

- Naïve Psychology
- ROC story
- Social IQa
- WSC
- COPA
- VCR
- WinoGrande

Physical commonsense

- Physical IQa
- HellaSwag
- SWAG
- JHU Ordinal Commonsense

Abductive NLI

CommonsenseQA

Temporal commonsense

- MCTaco

ReCORD

CosmosQA

MultiRC

Even more benchmarks: https://commonsense.run/

More benchmarks

Social commonsense

- Naïve Psychology
- ROC story
- Social IQa
- WSC
- COPA
- VCR
- WinoGrande

Physical commonsense

- Physical IQa
- HellaSwag
- SWAG
- JHU Ordinal Commonsense

Abductive NLI

CommonsenseQA

Temporal commonsense

- MCTaco

Commonsense reading comprehension

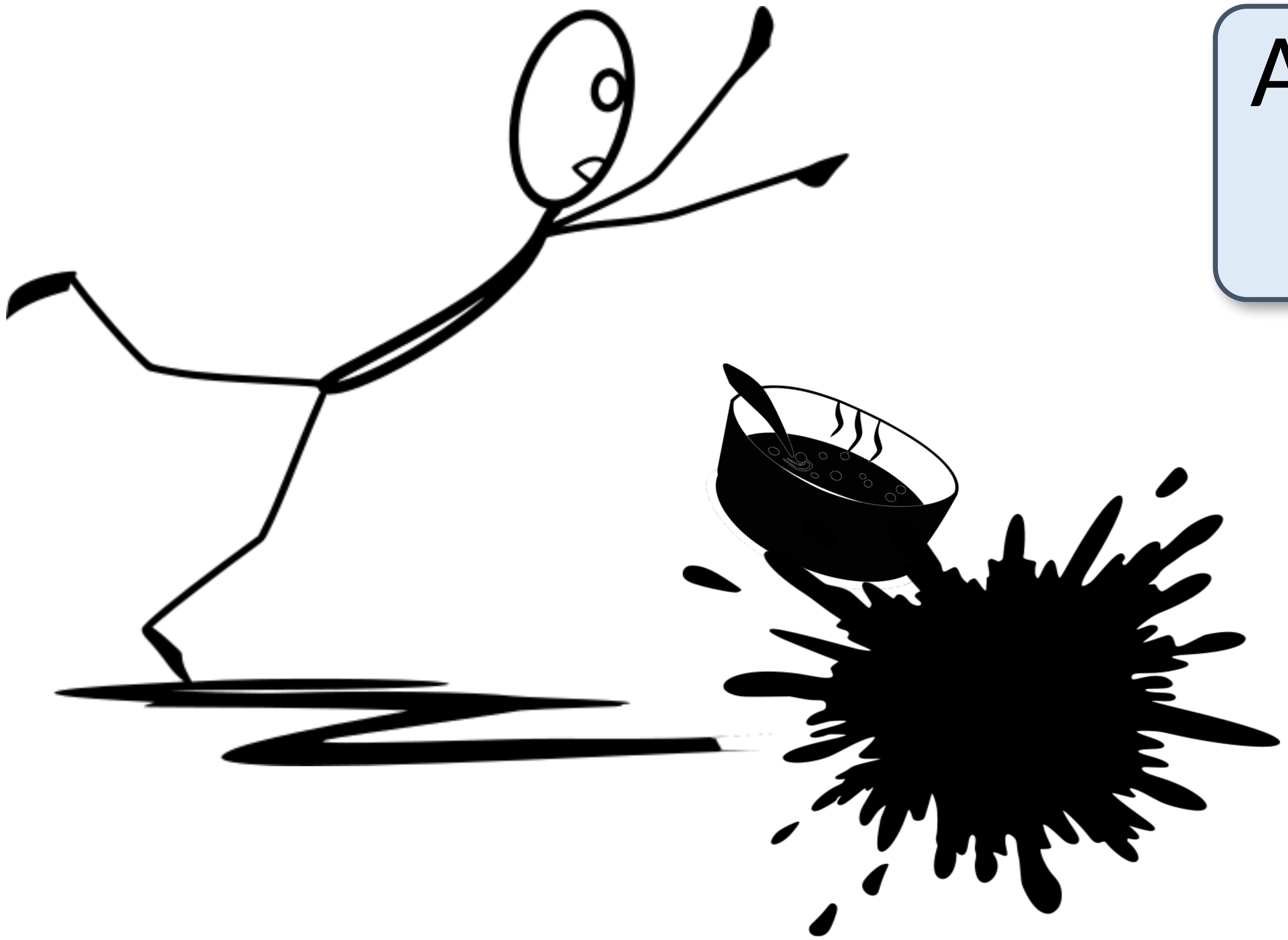
- ReCORD
- CosmosQA
- MultiRC

Even more benchmarks: https://commonsense.run/

Reasoning about Social Situations

https://leaderboard.allenai.org/socialiqa

Reasoning about Social Situations



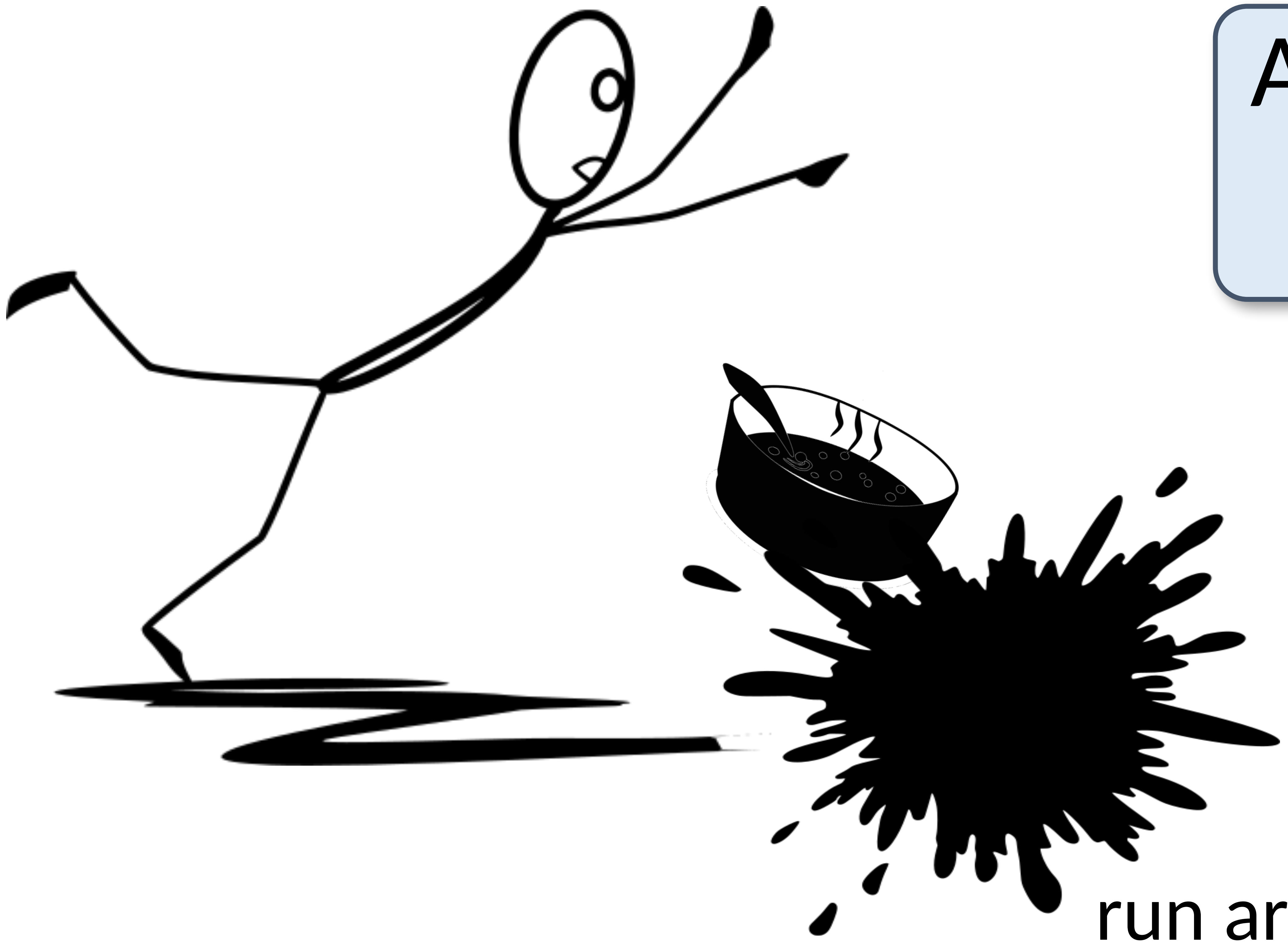
Alex spilt food all over the floor and it made a huge mess.

What will Alex want to do next?

https://leaderboard.allenai.org/socialiqa

Social IQa

Reasoning about Social Situations



Alex spilt food all over the floor and it made a huge mess.

What will Alex want to do next?

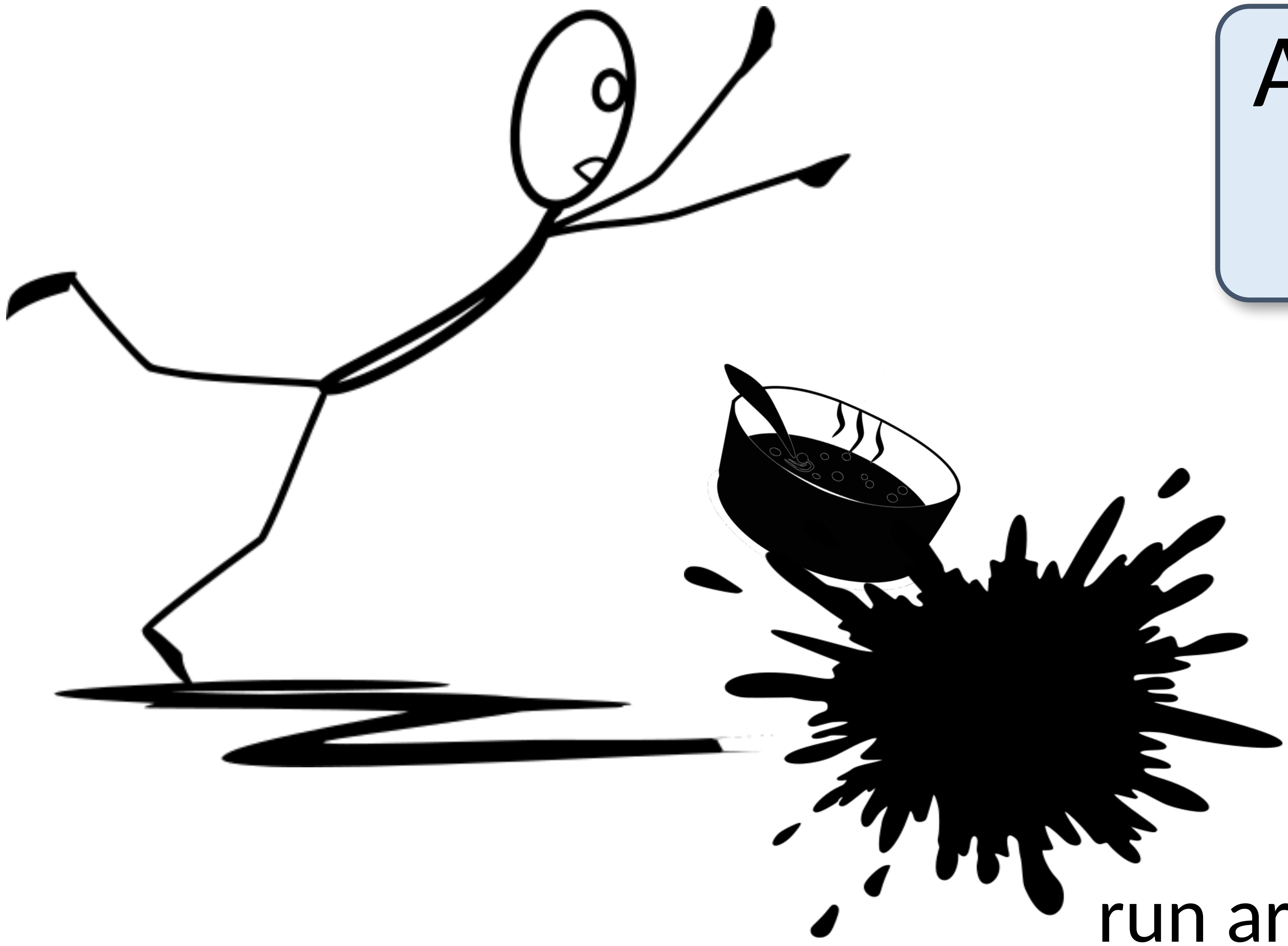
run around in the mess

mop up the mess

https://leaderboard.allenai.org/socialiqa

Reasoning about Social Situations

Social IQa



Alex spilt food all over the floor and it made a huge mess.

What will Alex want to do next?

run around in the mess

less likely

mop up the mess

more likely

https://leaderboard.allenai.org/socialiqa

Physical IQa

Reasoning about Physical Properties of the World

To separate egg whites from the yolk using a water bottle, you should

www.youtube.com › watch

Separating Egg Yolks With A Water Bottle - YouTube

EZTV ONLINE is the "How To" channel that combines entertainment with information. **We'll** show **you** the ...

Oct. 19, 2015 · Uploaded by eztv online

Squeeze the water bottle and press it against the yolk. **Release**, which creates suction and lifts the yolk.

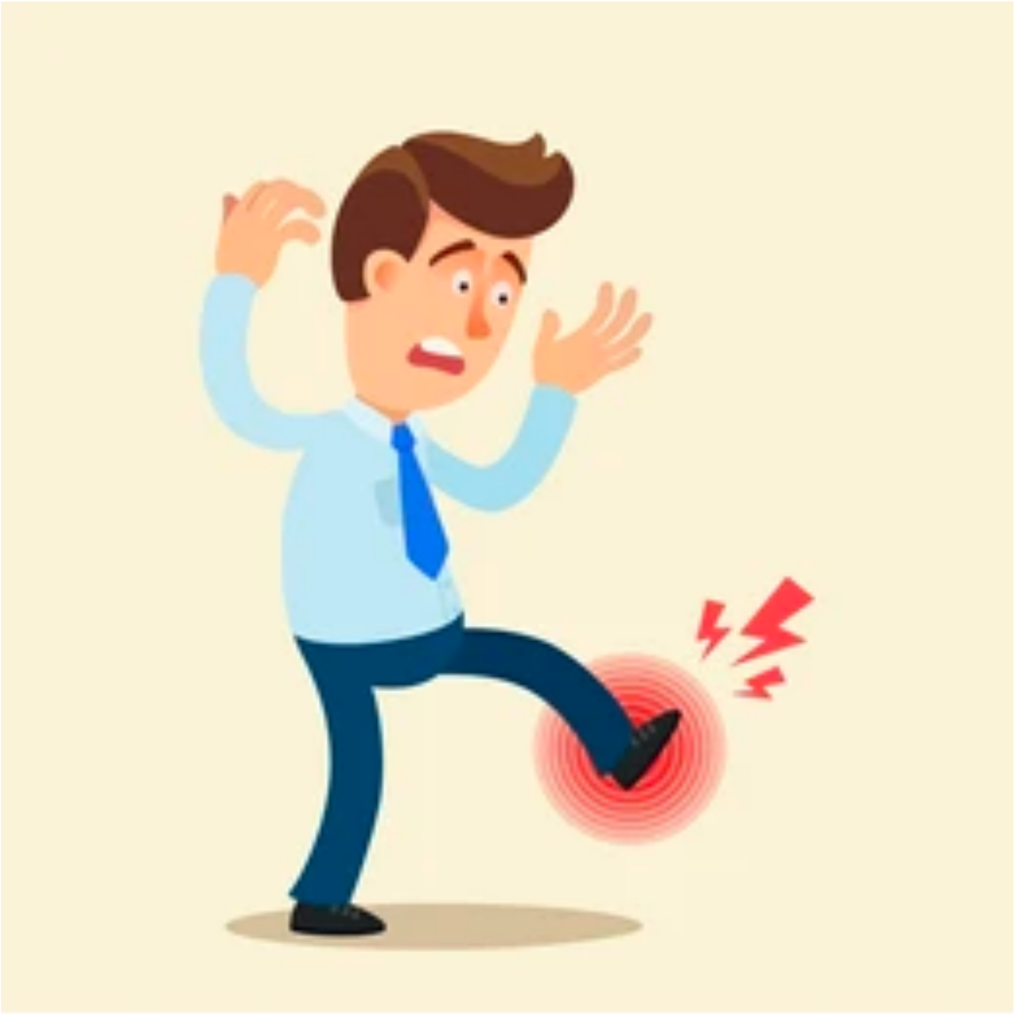
less likely

Place the water bottle and press it against the yolk. **Keep pushing**, which creates suction and lifts the yolk.

more likely

https://leaderboard.allenai.org/physicaliqa/

COPA: Choice of Plausible Alternatives



The man broke his toe.

What was the cause?

He got a hole in his sock.

less likely

He dropped a hammer on his foot.

more likely

RocStories

Karen was assigned a roommate her first year of college. Her roommate asked her to go to a nearby city for a concert. Karen agreed happily. The show was absolutely exhilarating.

Karen hated her roommate.

less likely

Karen became good friends with her roommate.

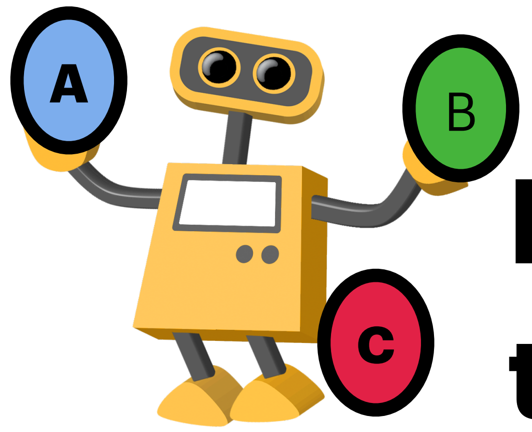
more likely

https://www.cs.rochester.edu/nlp/rocstories/

Discussion: Advantages and Disadvantages of Multiple-Choice Benchmarks

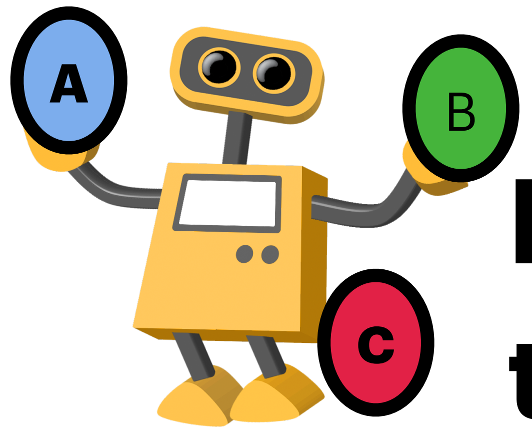
Reliable Evaluation

Reliable Evaluation



Discriminative tasks:

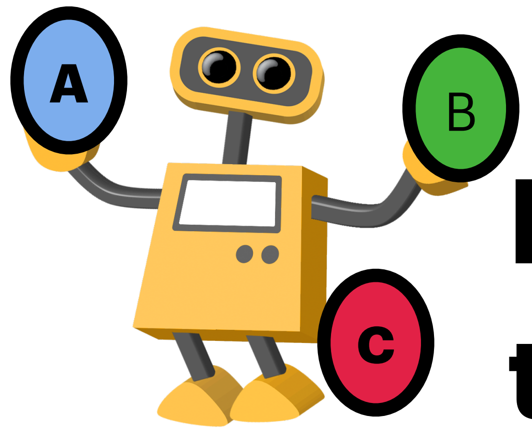
Reliable Evaluation



Discriminative tasks:

✓ Easy to evaluate

Reliable Evaluation

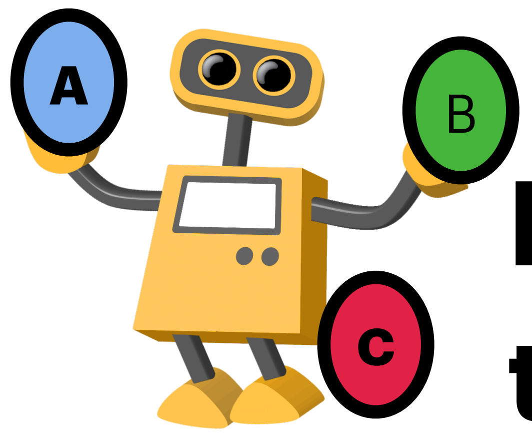


Discriminative tasks:

✓ Easy to evaluate

✗ Models are right for the wrong reasons

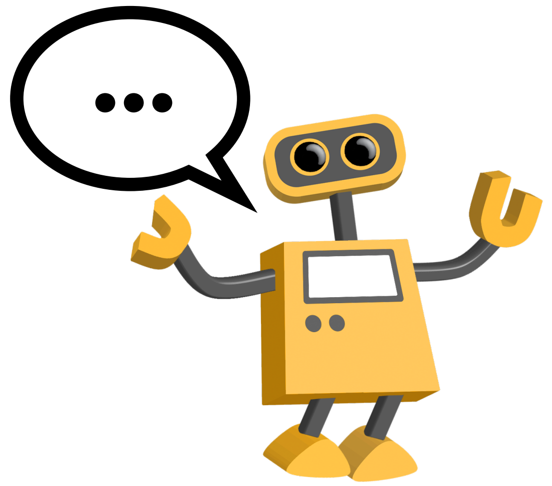
Reliable Evaluation



Discriminative tasks:

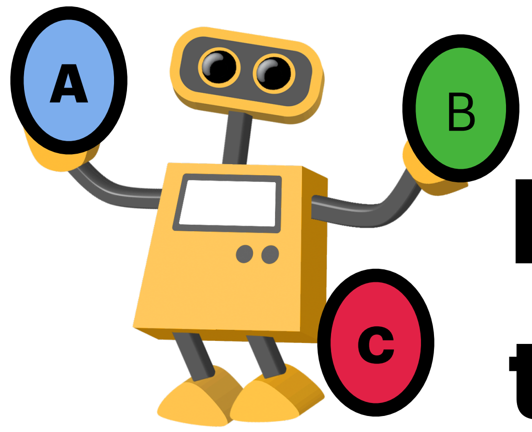
✓ Easy to evaluate

✗ Models are right for the wrong reasons



Generative tasks:

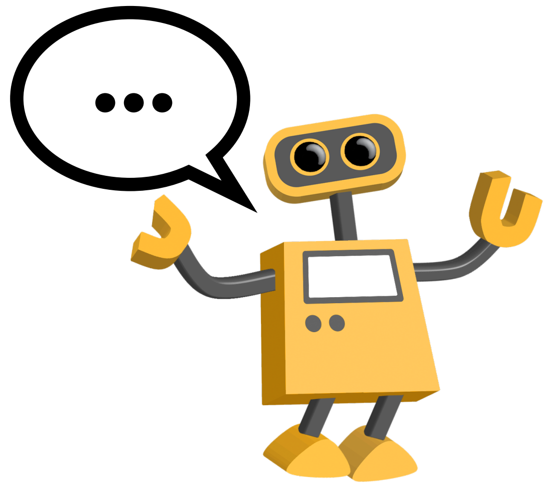
Reliable Evaluation



Discriminative tasks:

✓ Easy to evaluate

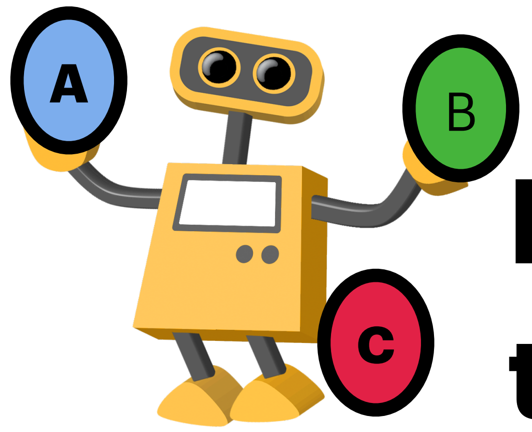
✗ Models are right for the wrong reasons



Generative tasks:

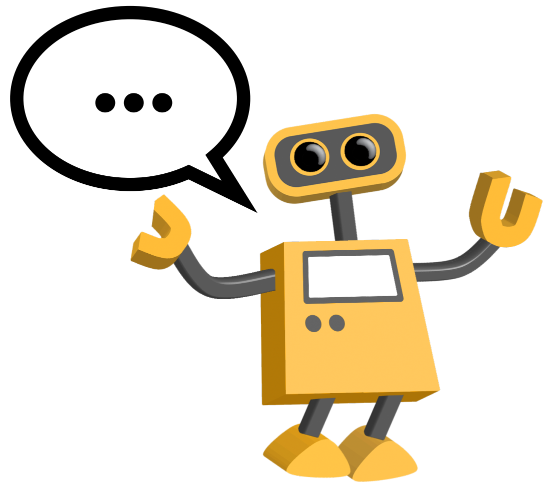
✓ More nuanced & flexible than pre-defined labels

Reliable Evaluation



Discriminative tasks:

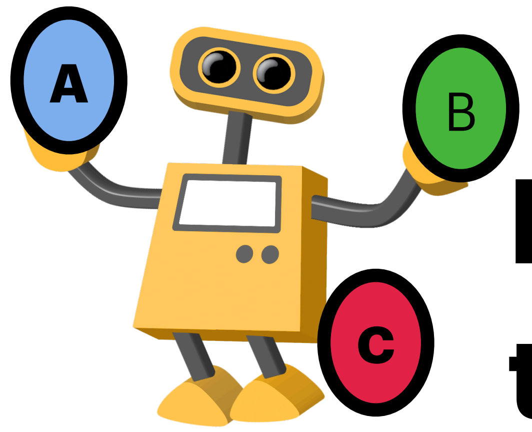
- ✓ Easy to evaluate
- ✗ Models are right for the wrong reasons



Generative tasks:

- ✓ More nuanced & flexible than pre-defined labels
- ✓ More similar to human reasoning process (no “answer choices”)

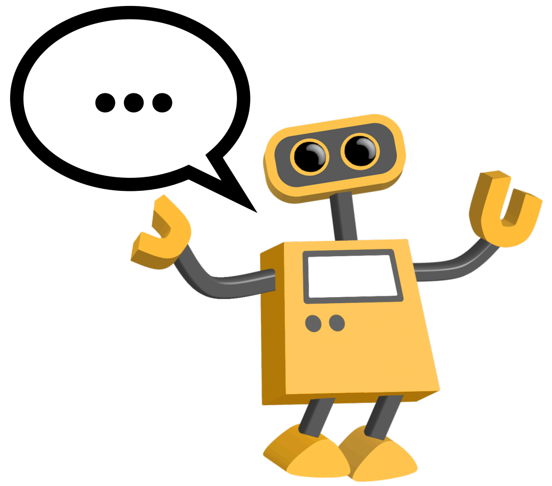
Reliable Evaluation



Discriminative tasks:

✓ Easy to evaluate

✗ Models are right for the wrong reasons



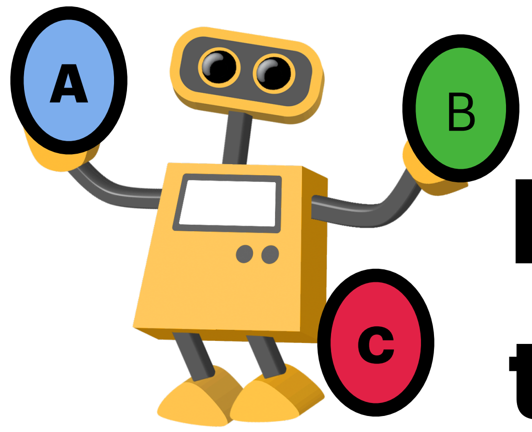
Generative tasks:

✓ More nuanced & flexible than pre-defined labels

✓ More similar to human reasoning process (no “answer choices”)

✓ Infinite answer space (no “guessing” of correct answer)

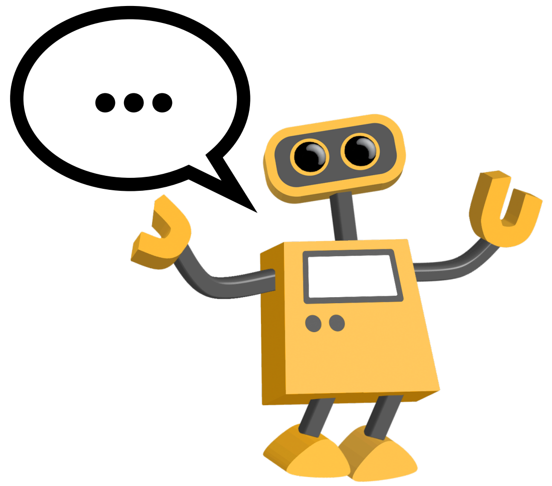
Reliable Evaluation



Discriminative tasks:

✓ Easy to evaluate

✗ Models are right for the wrong reasons



Generative tasks:

✓ More nuanced & flexible than pre-defined labels

✓ More similar to human reasoning process (no "answer choices")

✓ Infinite answer space (no "guessing" of correct answer)

✗ No reliable automatic evaluation metric

CommonGen

Concept-Set: a collection of objects/actions.

dog | frisbee | catch | throw

Generative Commonsense Reasoning

Expected Output: everyday scenarios covering all given concepts.

[Humans]

- A dog leaps to catch a thrown frisbee.
- The dog catches the frisbee when the boy throws it.
- A man throws away his dog 's favorite frisbee expecting him to catch it in the air.

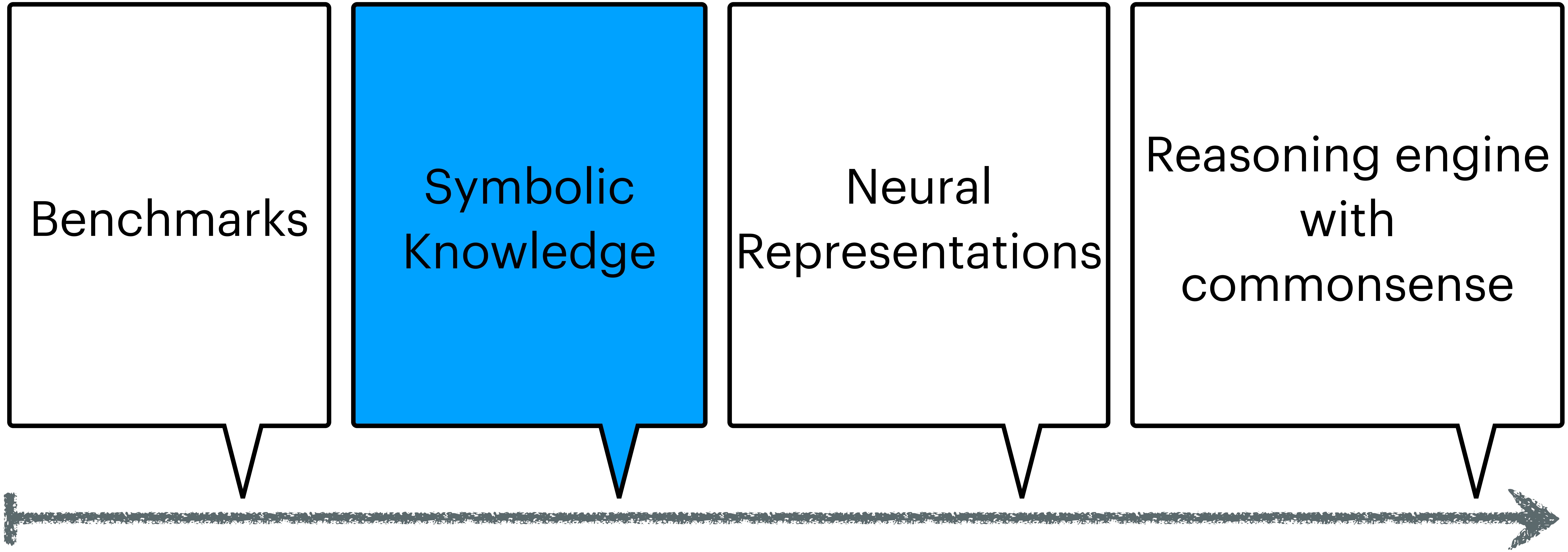
[Machines]

GPT2: A dog throws a frisbee at a football player.
UniLM: Two dogs are throwing frisbees at each other .
BART: A dog throws a frisbee and a dog catches it.
T5: dog catches a frisbee and throws it to a dog

https://inklab.usc.edu/CommonGen/

CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning.
Bill Yuchen Lin, Wangchunshu Zhou, Ming Shen, Pei Zhou, Chandra Bhagavatula, Yejin Choi, and Xiang Ren. Findings of EMNLP 2020.

Path to commonsense



Grandma's glasses

Tom's grandma was reading a new book, when she dropped her glasses.

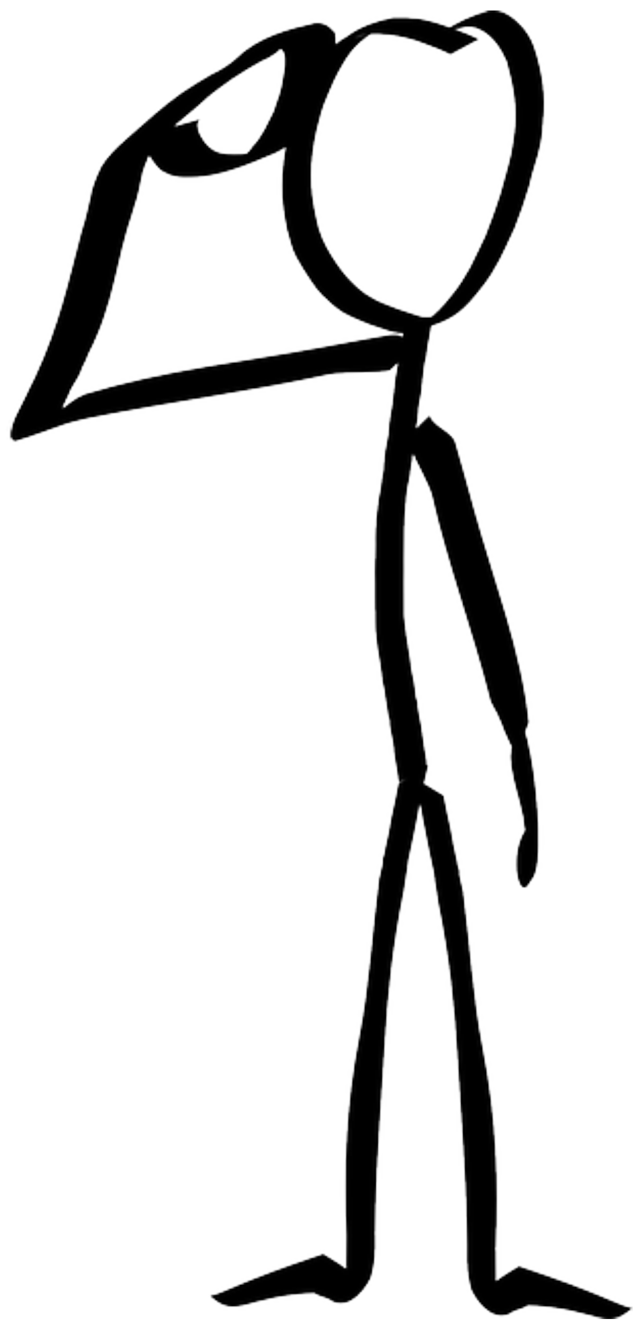
She couldn't pick them up, so she called Tom for help.

Tom rushed to help her look for them, they heard a loud crack.

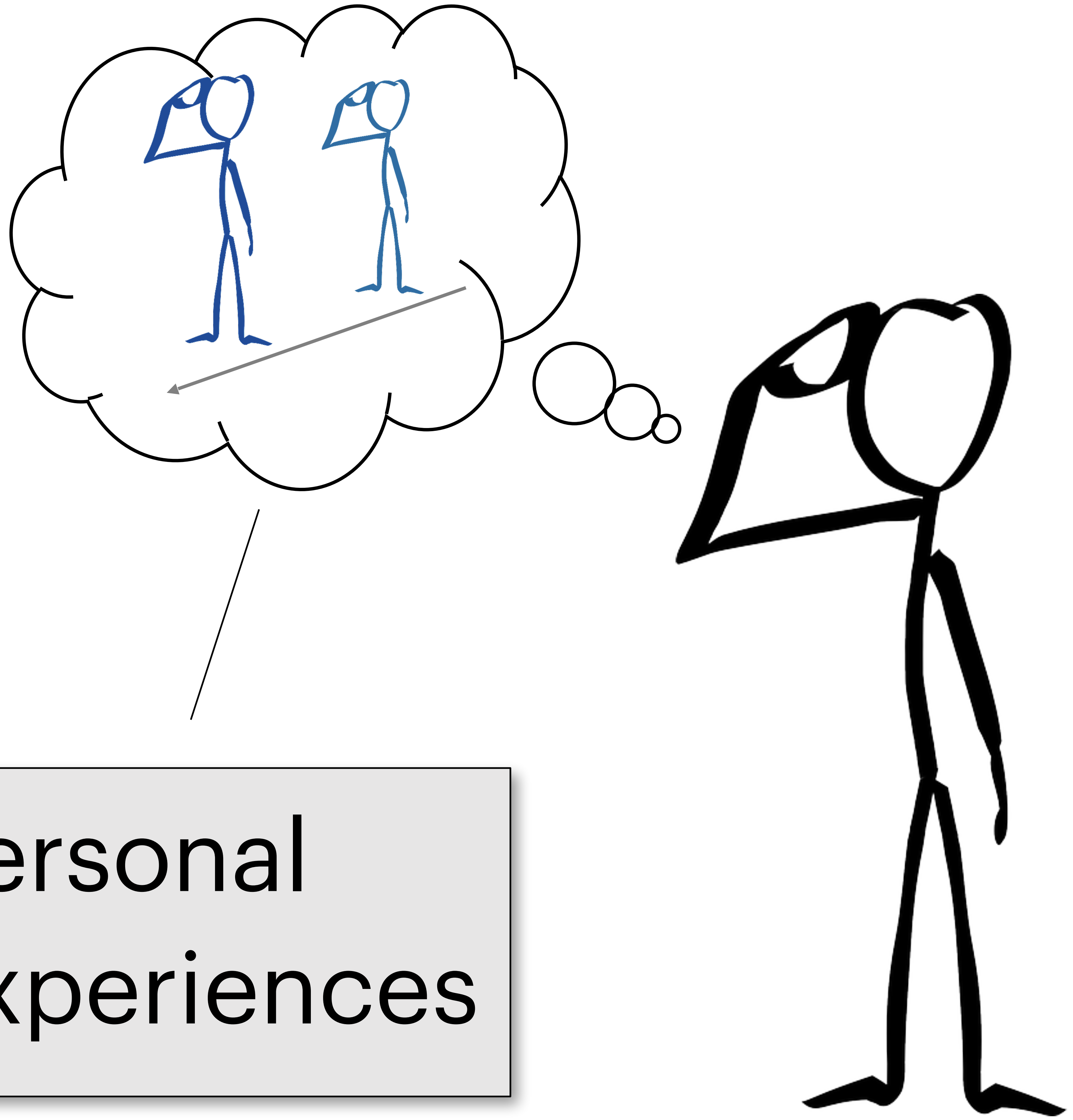
They realized that Tom broke her glasses by stepping on them.

Promptly, his grandma yelled at Tom to go get her a new pair.

Humans reason about the world with mental models [Graesser, 1994]

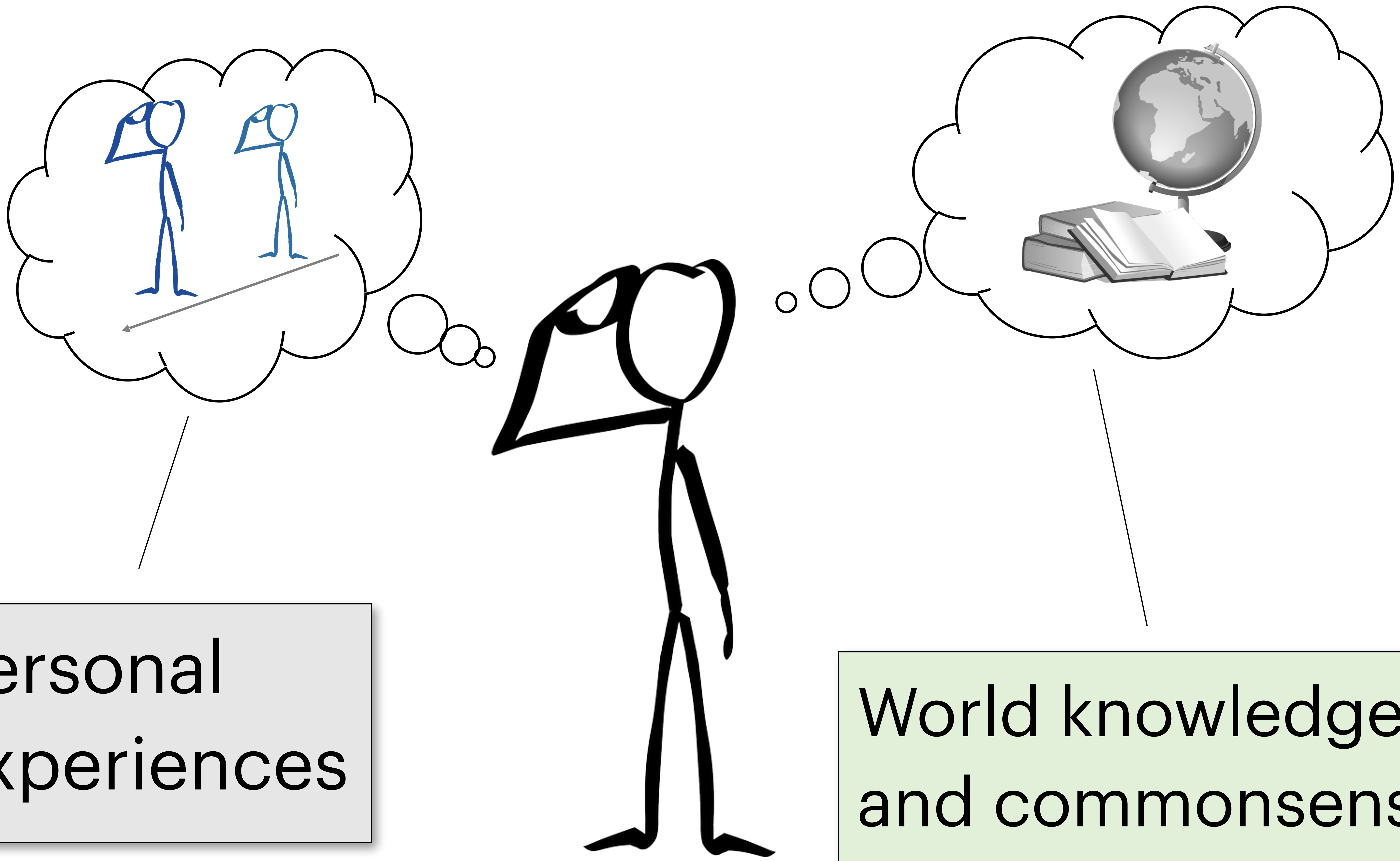


Humans reason about the world with mental models [Graesser, 1994]



Personal experiences

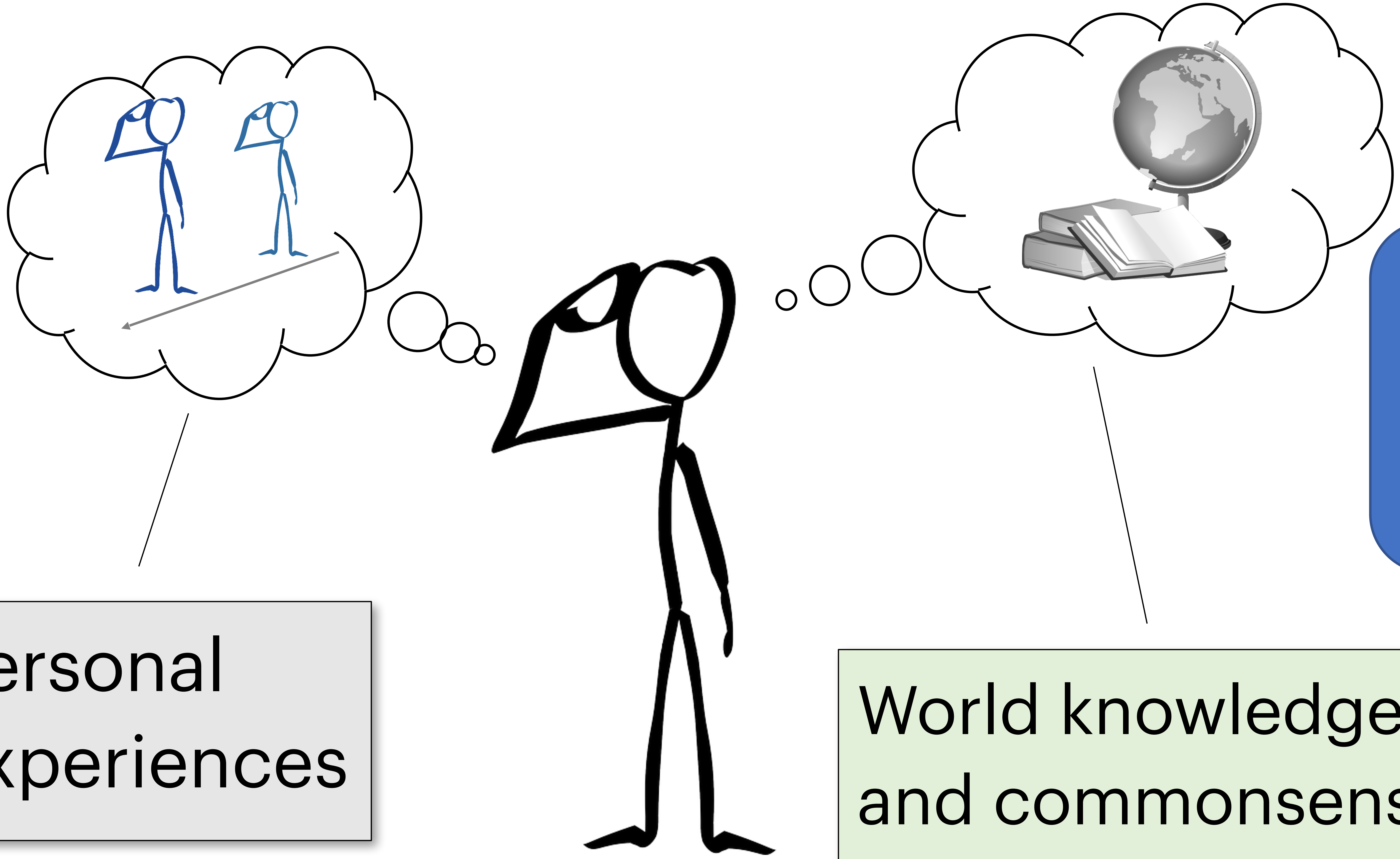
Humans reason about the world with mental models [Graesser, 1994]



Personal experiences

World knowledge and commonsense

Humans reason about the world with mental models [Graesser, 1994]



Personal experiences

World knowledge and commonsense

Commonsense resources aim to be a bank of knowledge for machines to be able to reason about the world in tasks

Tom’s grandma was reading a new book, when she dropped her glasses.

She couldn’t pick them up, so she called Tom for help.

Tom rushed to help her look for them, they heard a loud crack.

They realized that Tom broke her glasses by stepping on them.

Promptly, his grandma yelled at Tom to go get her a new pair.

usedFor

Tom's grandma was reading a new book, when she dropped her glasses.

She couldn't pick them up, so she called Tom for help.

Y will

Tom rushed to help her look for them, they heard a loud crack.

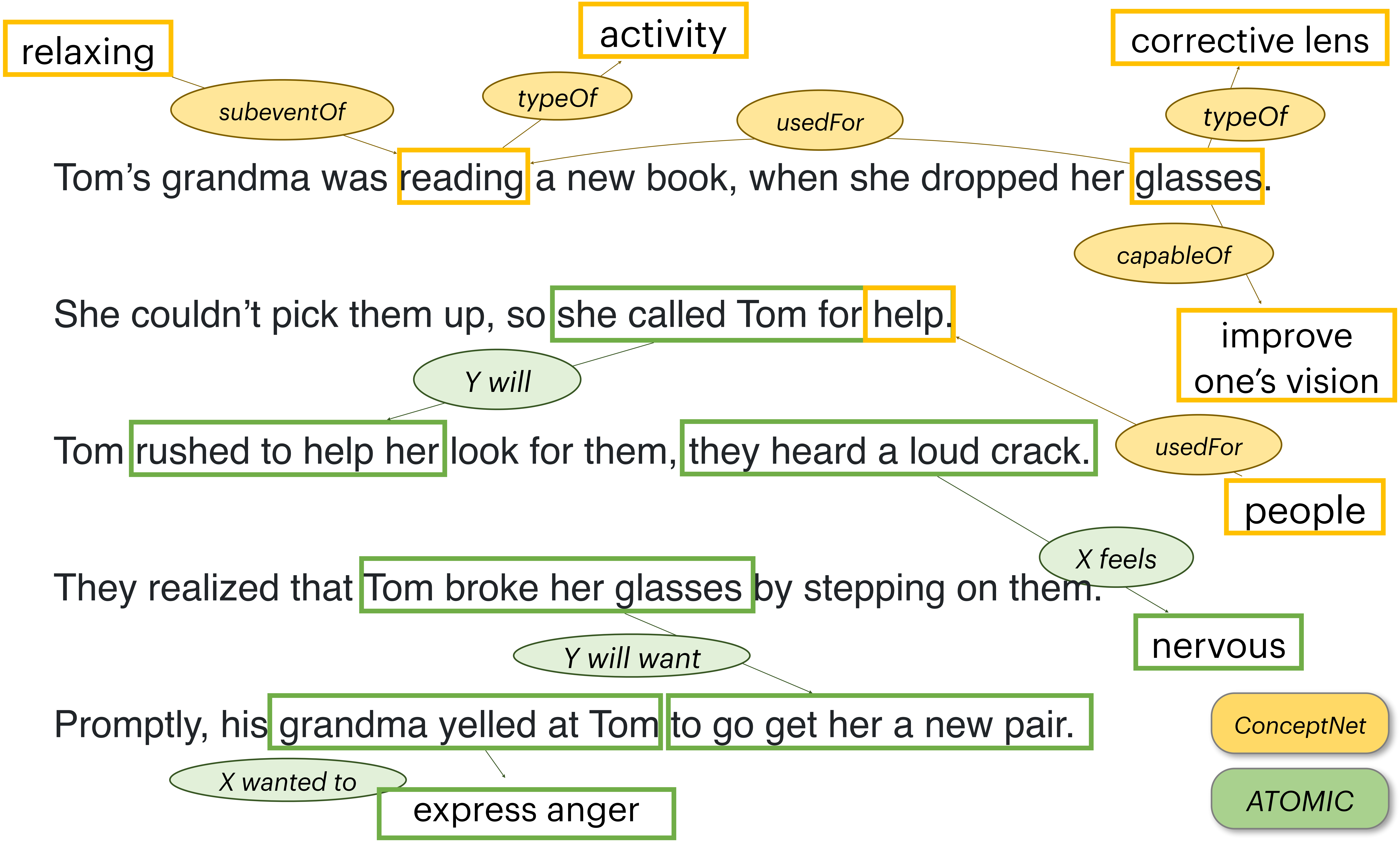
They realized that Tom broke her glasses by stepping on them.

Y will want

Promptly, his grandma yelled at Tom to go get her a new pair.

ConceptNet

ATOMIC


relaxing
activity
corrective lens
subeventOf
typeOf
usedFor
typeOf
Tom's grandma was reading a new book, when she dropped her glasses.
capableOf
She couldn't pick them up, so she called Tom for help.
improve one's vision
Y will
Tom rushed to help her look for them, they heard a loud crack.
usedFor
people
X feels
They realized that Tom broke her glasses by stepping on them.
nervous
Y will want
Promptly, his grandma yelled at Tom to go get her a new pair.
X wanted to
express anger
ConceptNet
ATOMIC

Overview of existing resources

Represented in **symbolic logic**
(e.g., LISP-style logic)

```
(#$implies
  (#$and
    (#$isa ?OBJ ?SUBSET)
    (#$genls ?SUBSET ?SUPERSET))
  (#$isa ?OBJ ?SUPERSET))
```

Cyc
(Lenat et al., 1984)

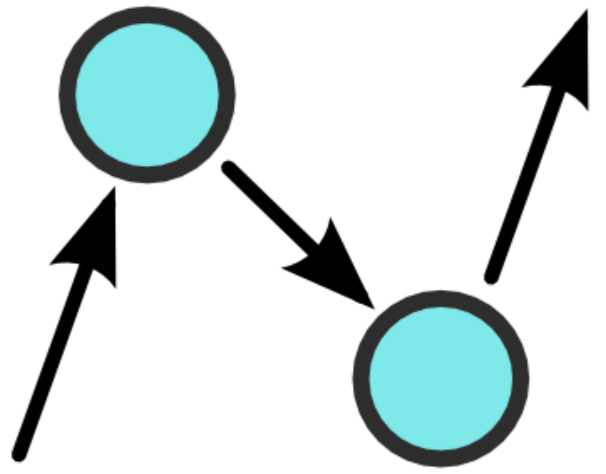
OpenCyc
(Lenat, 2004)

ResearchCyc
(Lenat, 2006)

OpenCyc 4.0
(Lenat, 2012)

today

Overview of existing resources



Open Mind Common Sense
(Minsky, Singh & Havasi, 1999)

ConceptNet
(Liu & Singh, 2004)

ConceptNet 5.5
(Speer et al., 2017)

Cyc
(Lenat et al., 1984)

OpenCyc
(Lenat, 2004)

ResearchCyc
(Lenat, 2006)

OpenCyc 4.0
(Lenat, 2012)

today

Represented in **natural language**
(how humans *talk* and *think*)

en reading
An English term in ConceptNet 5.8

Represented in **natural language**
(how humans *talk* and *think*)

reading is a subevent of...

- en you learn →
- en turning a page →
- en learning →

en **reading**

An English term in ConceptNet 5.8

Related terms

- en book →
- en books →
- en book →

Represented in **natural language**
(how humans *talk* and *think*)

reading is a subevent of...

- en you learn →
- en turning a page →
- en learning →

en **reading**

An English term in ConceptNet 5.8

Represented in **natural language**
(how humans *talk* and *think*)

Related terms

- en book →
- en books →
- en book →

Effects of reading

- en learning →
- en ideas →
- en a headache →

reading is a subevent of...

- en you learn →
- en turning a page →
- en learning →

en reading

An English term in ConceptNet 5.8

Represented in **natural language**
(how humans *talk* and *think*)

en reading

An English term in ConceptNet 5.8

Related terms

- en book →
- en books →
- en book →

Effects of reading

- en learning →
- en ideas →
- en a headache →

reading is a subevent of...

- en you learn →
- en turning a page →
- en learning →

reading is a type of...

- en an activity →
- en a good way to learn →
- en one way of learning →
- en one way to learn →

Represented in **natural language**
(how humans *talk* and *think*)

Related terms

- en book →
- en books →
- en book →

Effects of reading

- en learning →
- en ideas →
- en a headache →

reading is a type of...

- en an activity →
- en a good way to learn →
- en one way of learning →
- en one way to learn →

reading is a subevent of...

- en you learn →
- en turning a page →
- en learning →

en reading

An English term in ConceptNet 5.8

Types of reading

- en browse (n, communication) →
- en bumf (n, communication) →
- en clock time (n, time) →
- en miles per hour (n, time) →

Represented in **natural language**
(how humans *talk* and *think*)

Related terms

- en book →
- en books →
- en book →

Effects of reading

- en learning →
- en ideas →
- en a headache →

reading is a type of...

- en an activity →
- en a good way to learn →
- en one way of learning →
- en one way to learn →

reading is a subevent of...

- en you learn →
- en turning a page →
- en learning →

en reading

An English term in ConceptNet 5.8

Things used for reading

- en article →
- en a library →
- en literature →
- en a paper page →

Types of reading

- en browse (n, communication) →
- en bumf (n, communication) →
- en clock time (n, time) →
- en miles per hour (n, time) →

Represented in **natural language**
(how humans *talk* and *think*)

en reading

An English term in ConceptNet 5.8

Related terms

- en book →
- en books →
- en book →

Effects of reading

- en learning →
- en ideas →
- en a headache →

reading is a type of...

- en an activity →
- en a good way to learn →
- en one way of learning →
- en one way to learn →

reading is a subevent of...

- en you learn →
- en turning a page →
- en learning →

Subevents of reading

- en relaxing →
- en study →
- en studying for a subject →

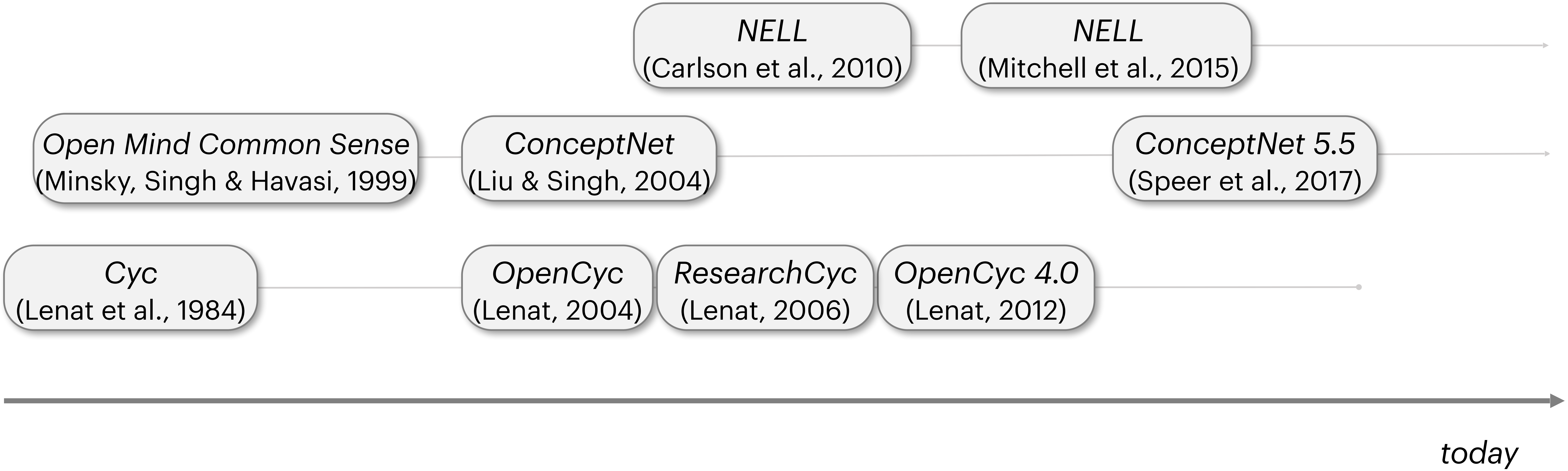
Things used for reading

- en article →
- en a library →
- en literature →
- en a paper page →

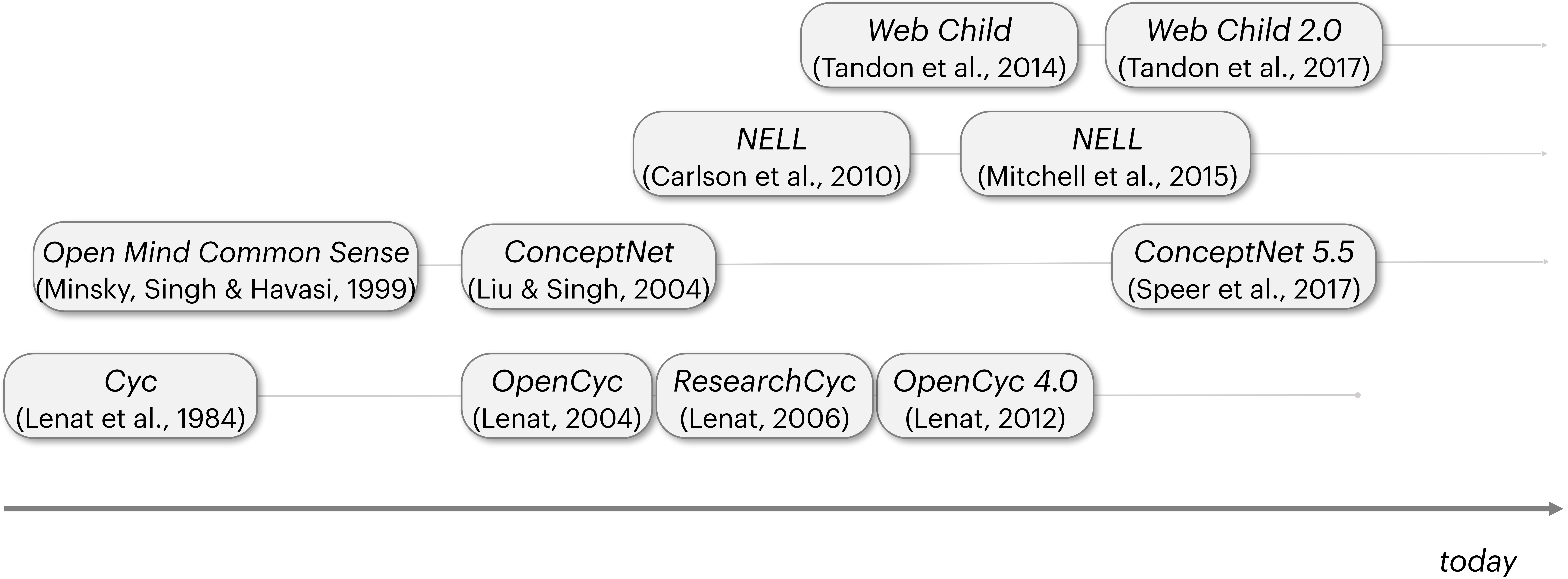
Types of reading

- en browse (n, communication) →
- en bumf (n, communication) →
- en clock time (n, time) →
- en miles per hour (n, time) →

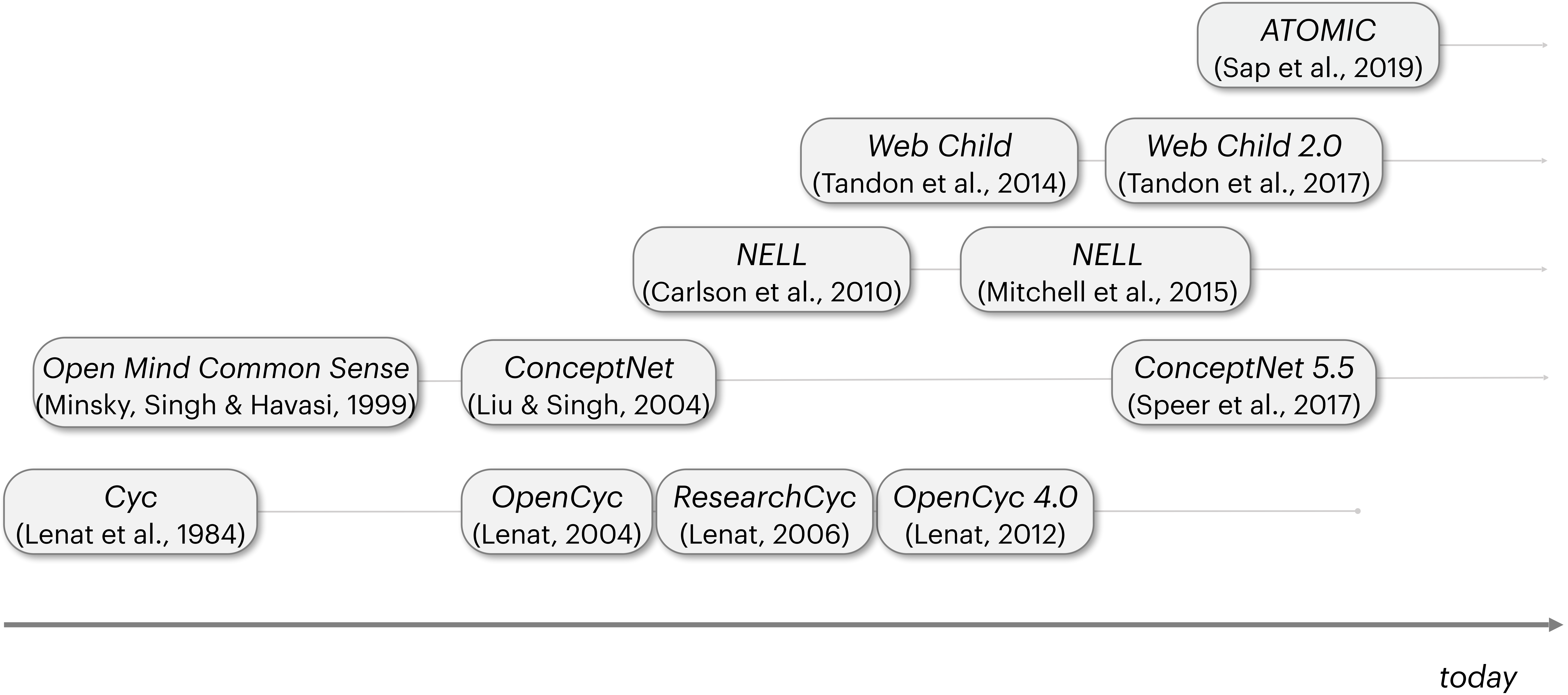
Overview of existing resources



Overview of existing resources

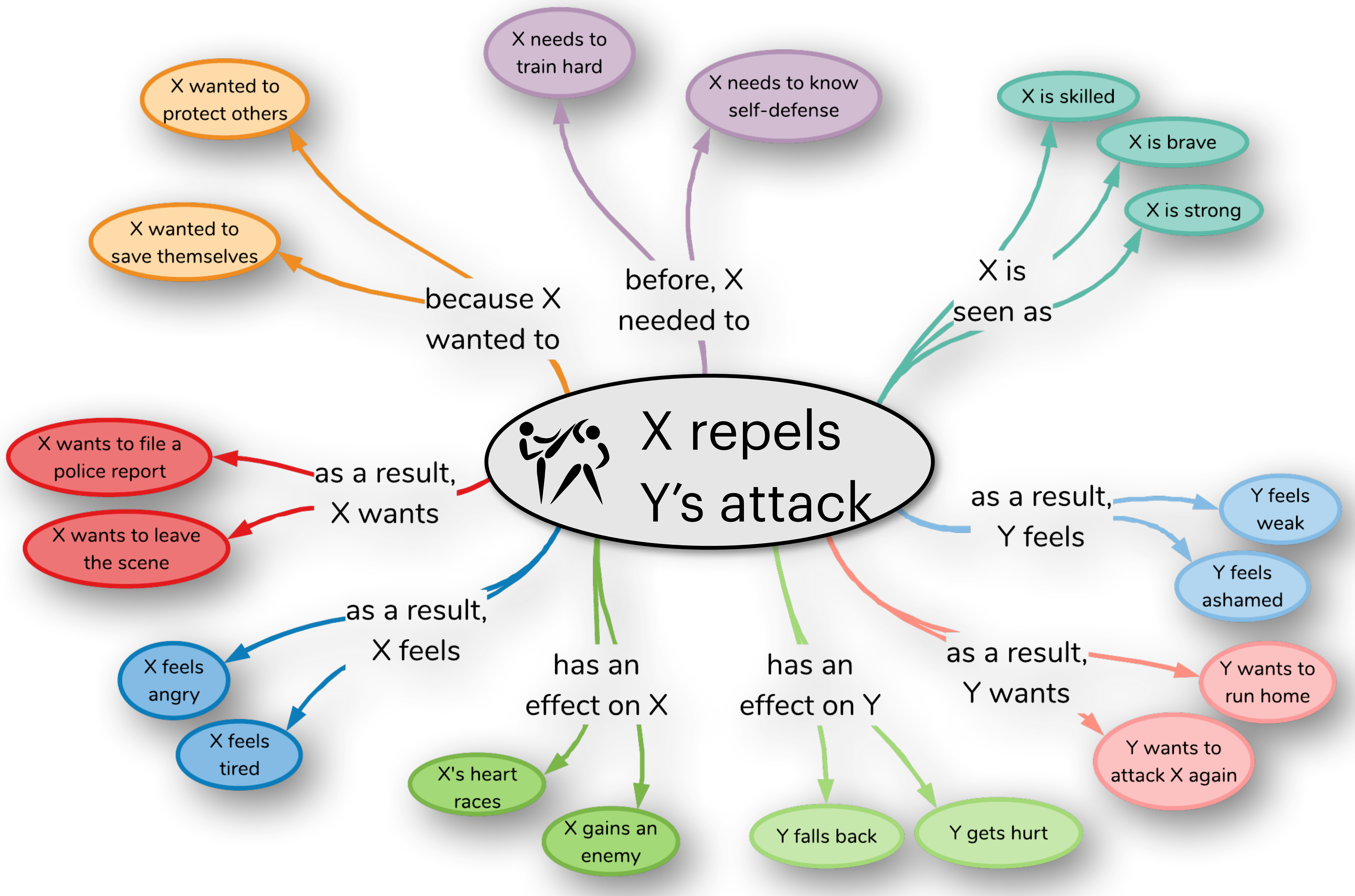


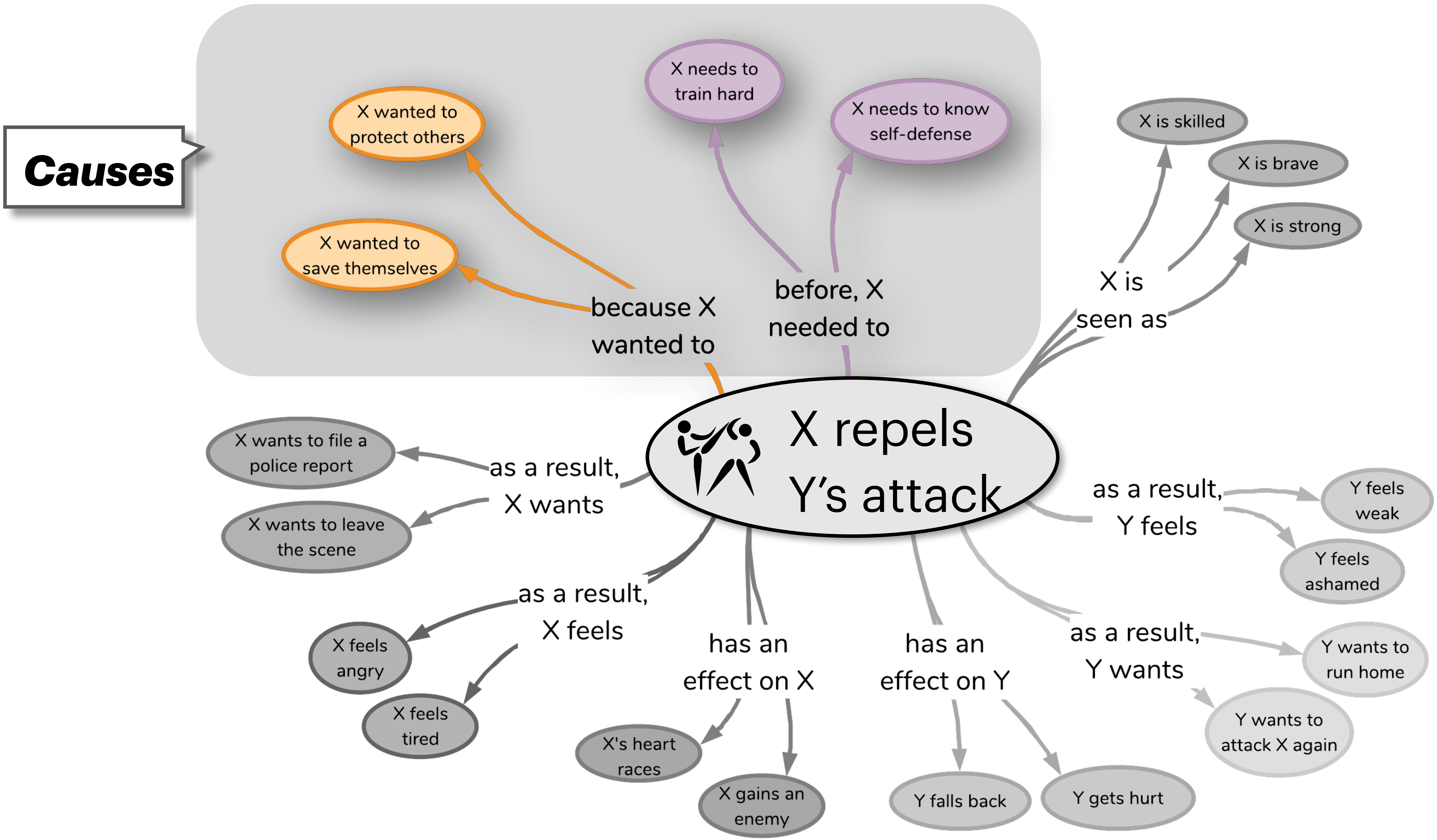
Overview of existing resources

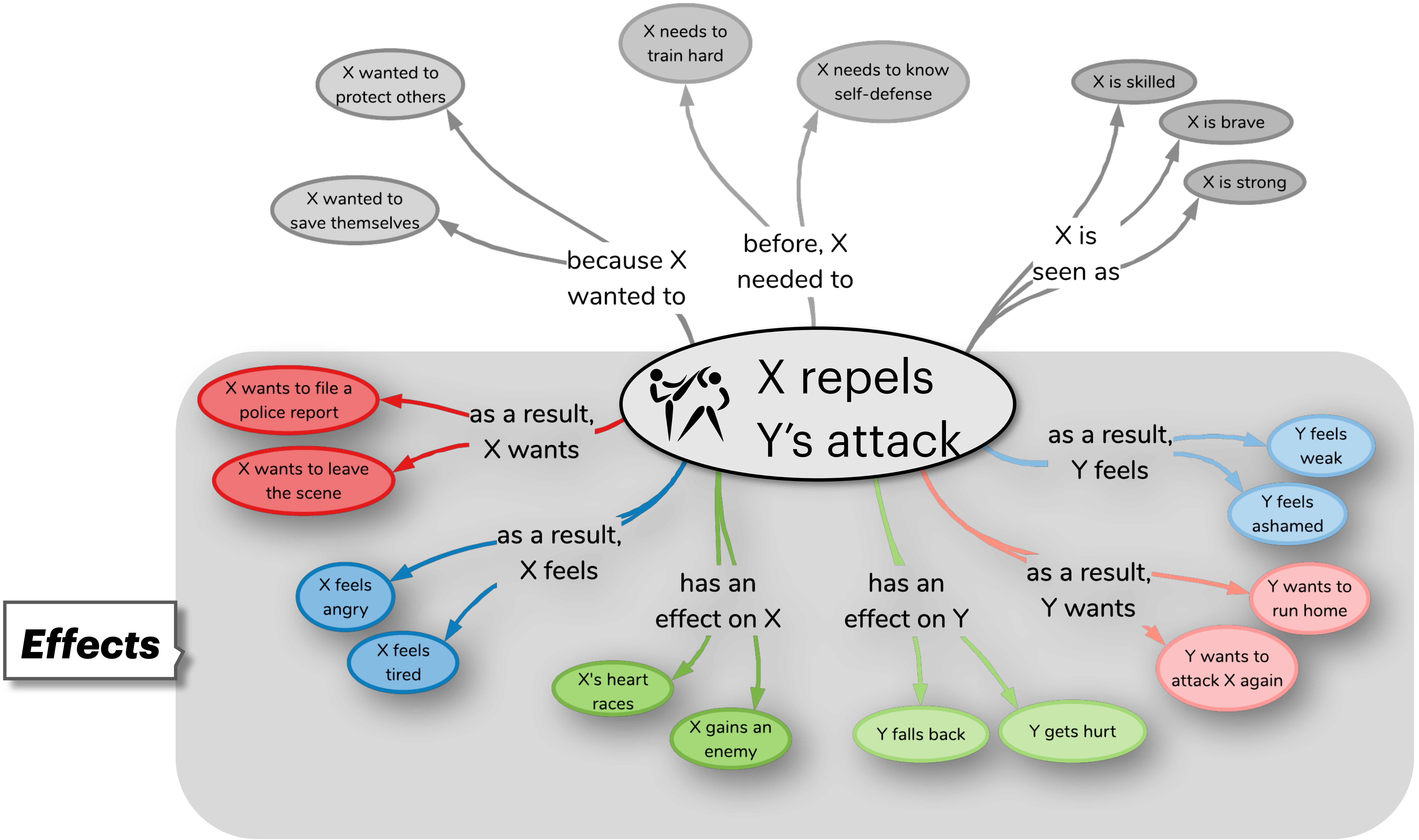


Represented in **natural language**
(how humans *talk* and *think*)

ATOMIC: 880,000 triples for AI systems to reason about *causes* and *effects* of everyday situations


X repels Y's attack
because X wanted to
X wanted to protect others
X wanted to save themselves
before, X needed to
X needs to train hard
X needs to know self-defense
X is seen as
X is skilled
X is brave
X is strong
as a result, Y feels
Y feels weak
Y feels ashamed
as a result, Y wants
Y wants to run home
Y wants to attack X again
has an effect on Y
Y falls back
Y gets hurt
has an effect on X
X's heart races
X gains an enemy
as a result, X feels
X feels angry
X feels tired
as a result, X wants
X wants to file a police report
X wants to leave the scene


Causes
X wanted to protect others
X wanted to save themselves
because X wanted to
X needs to train hard
X needs to know self-defense
before, X needed to
X is skilled
X is brave
X is strong
X is seen as
X repels Y's attack
X wants to file a police report
X wants to leave the scene
as a result, X wants
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
has an effect on Y
Y falls back
Y gets hurt
as a result, Y wants
Y wants to run home
Y wants to attack X again
as a result, Y feels
Y feels weak
Y feels ashamed


X wanted to protect others
X needs to train hard
X needs to know self-defense
X is skilled
X is brave
X is strong
X wanted to save themselves
because X wanted to
before, X needed to
X is seen as
X repels Y's attack
X wants to file a police report
X wants to leave the scene
as a result, X wants
as a result, Y feels
Y feels weak
Y feels ashamed
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
has an effect on Y
Y falls back
Y gets hurt
as a result, Y wants
Y wants to run home
Y wants to attack X again
Effects

Decisions when building a new resource

Decisions when building a new resource

1. Representation Tradeoff between **expressivity** and **ease of collection**

Decisions when building a new resource

1. **Representation** Tradeoff between **expressivity** and **ease of collection**

2. **Knowledge Type**

Decisions when building a new resource

1. Representation Tradeoff between **expressivity** and **ease of collection**

2. Knowledge Type

3. Acquisition Method

Discussion:
Tradeoffs between collecting knowledge from people and extracting from text

3. Acquisition Method

1️⃣ from people

2️⃣ from text

✘ Expensive, takes a long time 💲

3. Acquisition Method

1️⃣ **from people**

2️⃣ **from text**

✘ Expensive, takes a long time

Reporting bias and knowledge acquisition. Jonathan Gordon and Benjamin Van Durme. AKBC 2013.

3. Acquisition Method

1️⃣ **from people**

✘ Expensive, takes a long time

2️⃣ **from text**

✘ Reporting bias

Reporting bias and knowledge acquisition. Jonathan Gordon and Benjamin Van Durme. AKBC 2013.

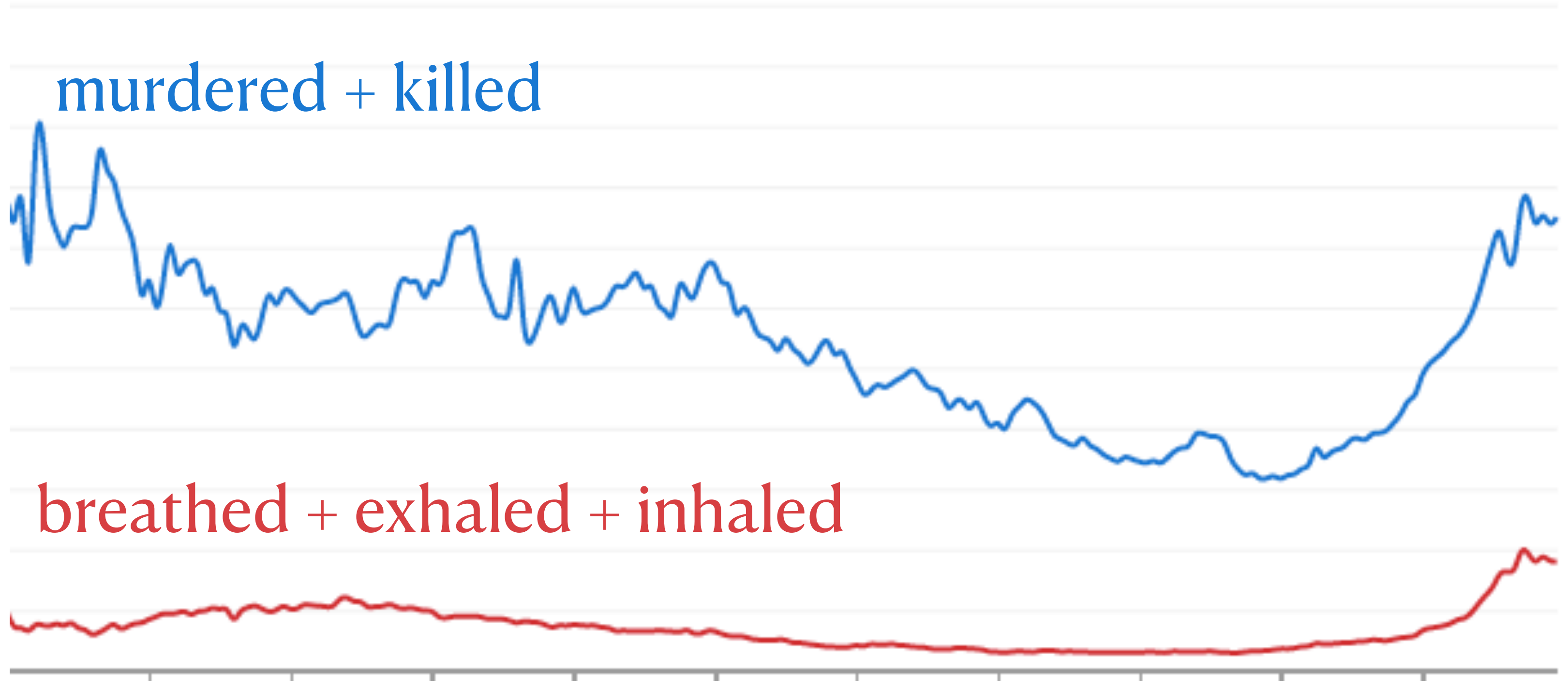
3. Acquisition Method

1️⃣ from people

✘ Expensive, takes a long time

2️⃣ from text

✘ Reporting bias



Reporting bias and knowledge acquisition. Jonathan Gordon and Benjamin Van Durme. AKBC 2013.

3. Acquisition Method

1️⃣ from people

✘ Expensive, takes a long time 💲

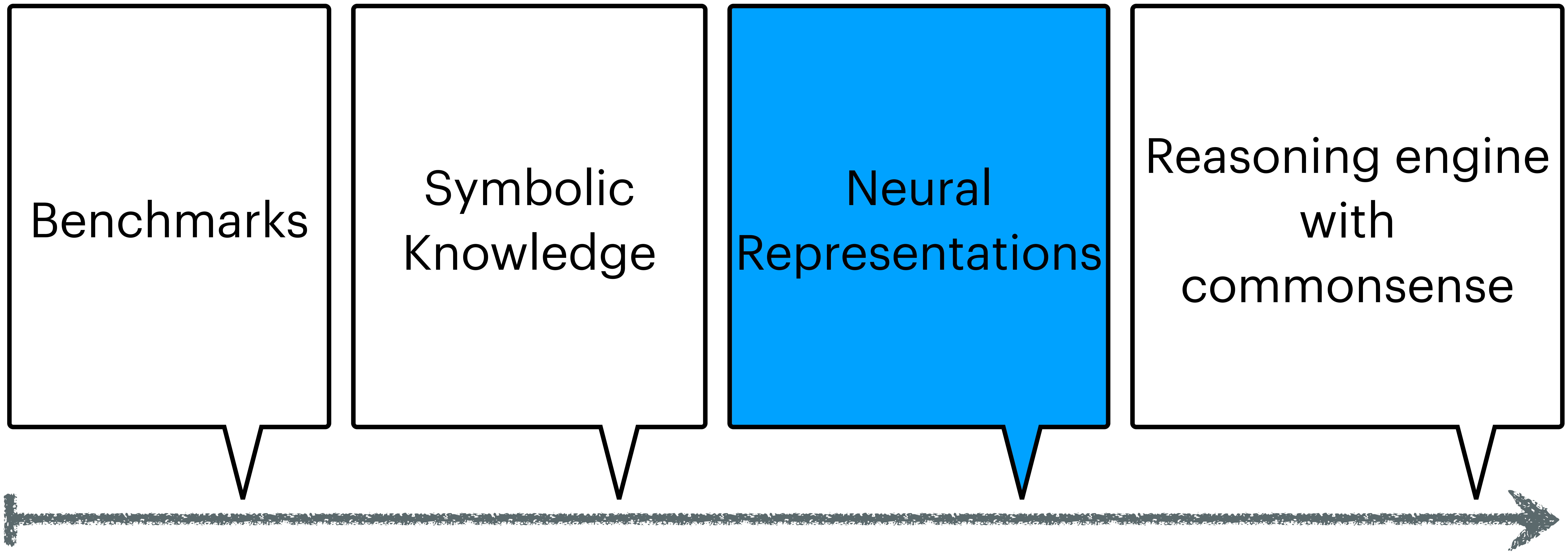
2️⃣ from text

✘ Reporting bias

✘ What is NOT true

Reporting bias and knowledge acquisition. Jonathan Gordon and Benjamin Van Durme. AKBC 2013.

Path to commonsense



✓ Knowledge in Pre-trained LMs

WIKIPEDIA
The Free Encyclopedia

A Primer in BERTology: What we know about how BERT works. Anna Rogers, Olga Kovaleva, and Anna Rumshisky. TACL 2020.

✔ Knowledge in Pre-trained LMs

✅ Syntax:

- Encode information about parts of speech, syntactic chunks and roles
- Syntax trees can be recovered from the representation
- Subject-verb agreement (e.g. tense, plurality)

A Primer in BERTology: What we know about how BERT works. Anna Rogers, Olga Kovaleva, and Anna Rumshisky. TACL 2020.

✔ Knowledge in Pre-trained LMs

✅ Syntax:

- Encode information about parts of speech, syntactic chunks and roles
- Syntax trees can be recovered from the representation
- Subject-verb agreement (e.g. tense, plurality)

✅ Semantics:

- Semantic roles
- Entity types

A Primer in BERTology: What we know about how BERT works. Anna Rogers, Olga Kovaleva, and Anna Rumshisky. TACL 2020.

✔ Knowledge in Pre-trained LMs

✅ Syntax:

- Encode information about parts of speech, syntactic chunks and roles
- Syntax trees can be recovered from the representation
- Subject-verb agreement (e.g. tense, plurality)

✅ Semantics:

- Semantic roles
- Entity types

✅ Factual knowledge

Domain-specific facts Most people don't know

The native language of Mammootty is [MASK].

Malayalam ✅

A Primer in BERTology: What we know about how BERT works. Anna Rogers, Olga Kovaleva, and Anna Rumshisky. TACL 2020.

❌ Knowledge in Pre-trained LMs

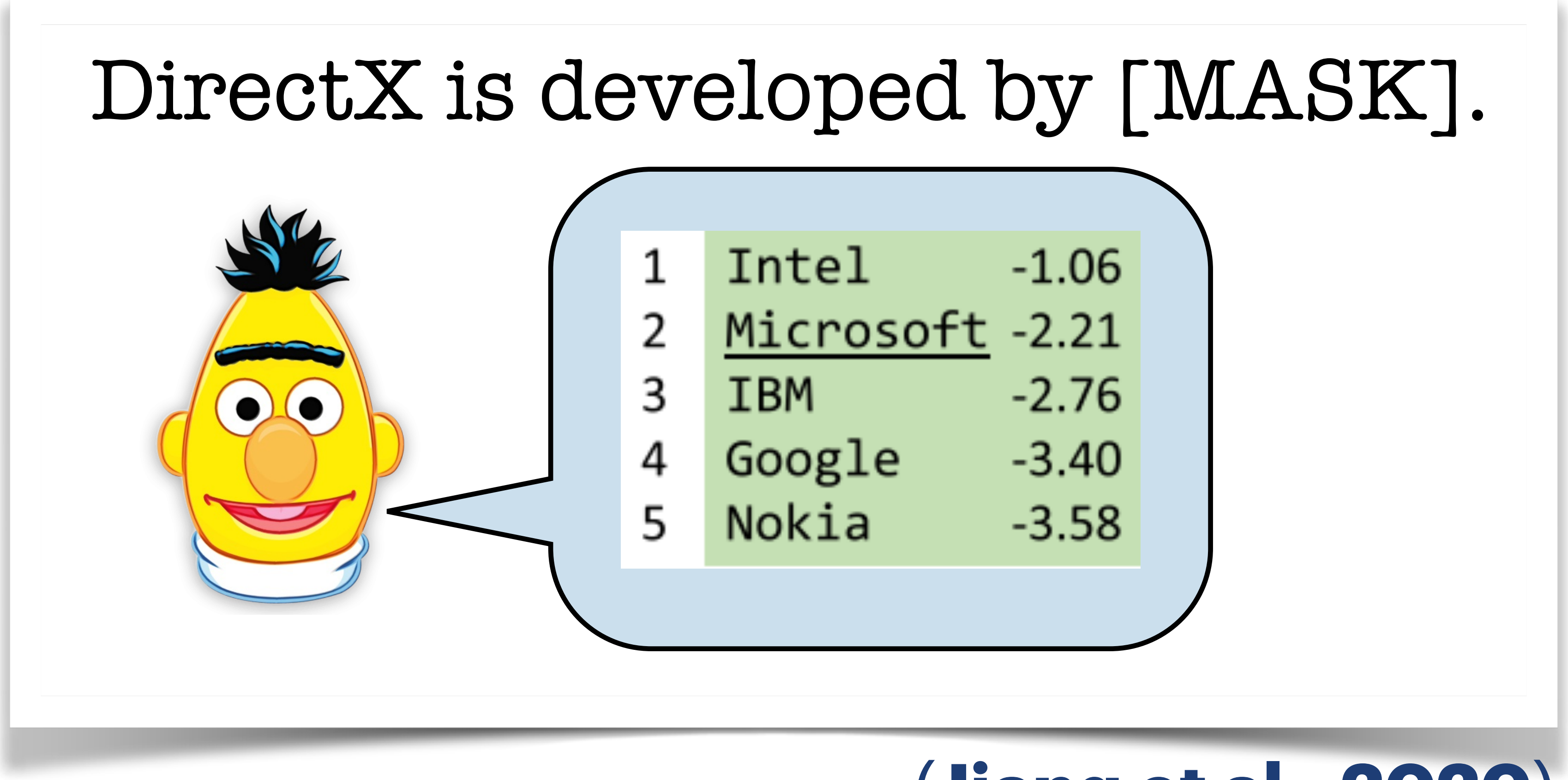
How can we know what language models know? Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. TACL 2020

Negated and Misprimed Probes for Pretrained Language Models: Birds Can Talk, But Cannot Fly. Nora Kassner and Hinrich Schütze. ACL 2020

What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. Allyson Ettinger. TACL 2020

❌ Knowledge in Pre-trained LMs

❌ Confuse semantically-similar mutually-exclusive terms



(Jiang et al., 2020)

How can we know what language models know? Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. TACL 2020

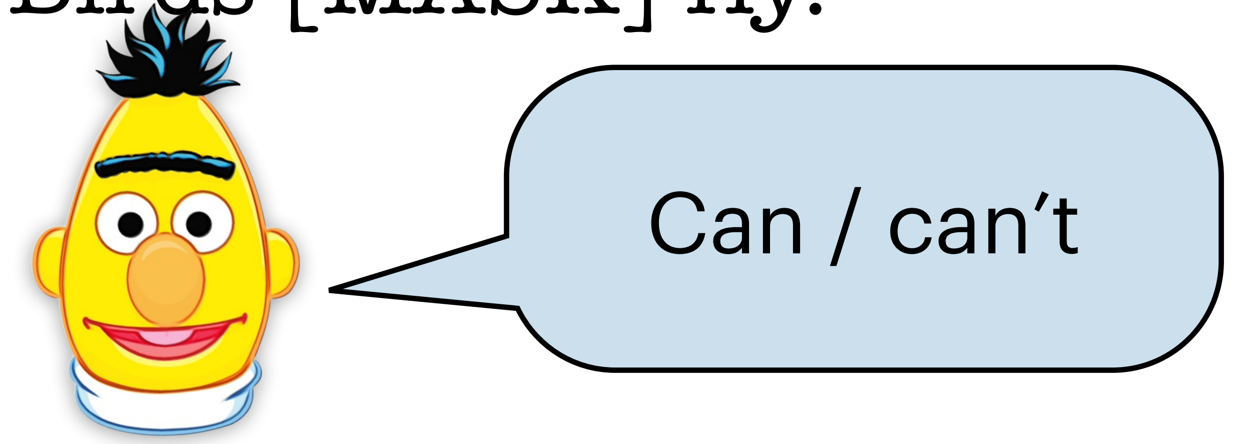
Negated and Misprimed Probes for Pretrained Language Models: Birds Can Talk, But Cannot Fly. Nora Kassner and Hinrich Schütze. ACL 2020

What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. Allyson Ettinger. TACL 2020

❌ Knowledge in Pre-trained LMs

- ❌ Confuse semantically-similar mutually-exclusive terms
- ❌ Are really bad with negation

Birds [MASK] fly.



(**Kassner et al. 2020; Ettinger, 2020**)

How can we know what language models know? Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. TACL 2020

Negated and Misprimed Probes for Pretrained Language Models: Birds Can Talk, But Cannot Fly. Nora Kassner and Hinrich Schütze. ACL 2020

What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. Allyson Ettinger. TACL 2020

❌ Knowledge in Pre-trained LMs

❌ Confuse semantically-similar mutually-exclusive terms

❌ Are really bad with negation

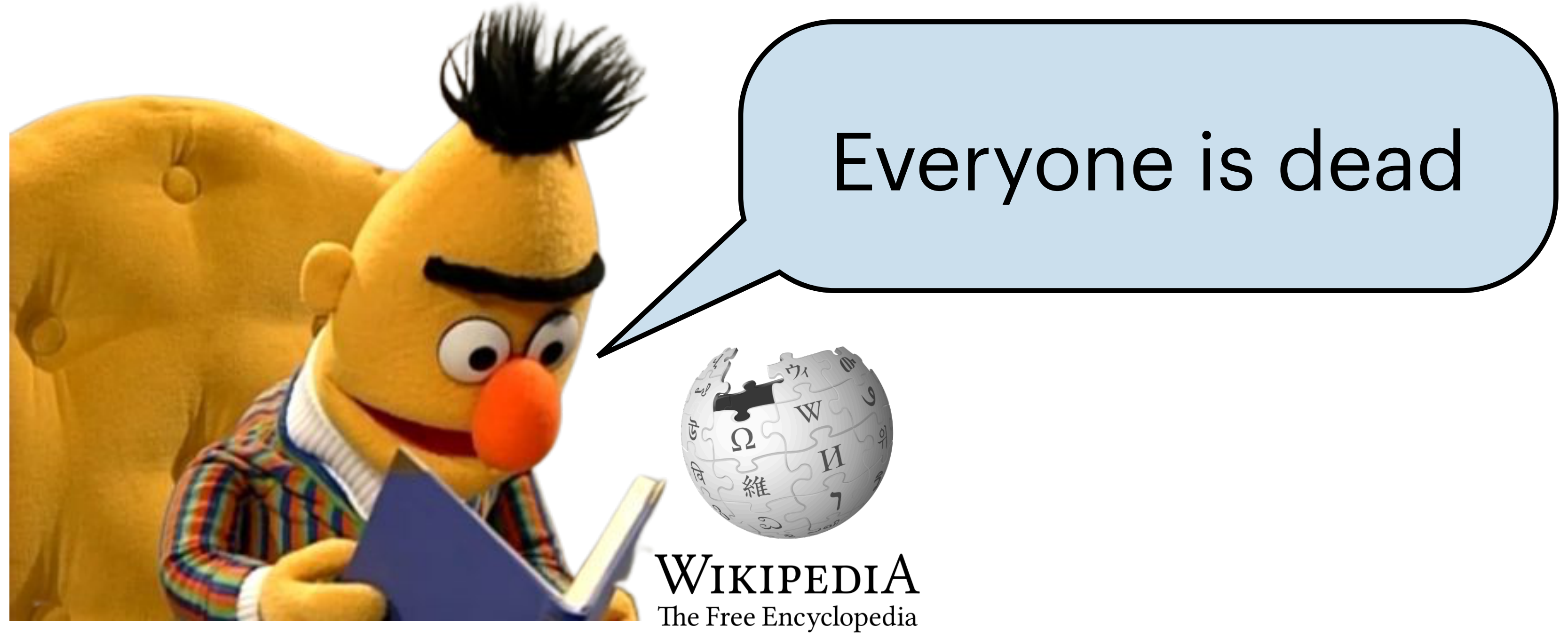
❌ Lack perceptual knowledge (people don’t talk about it)

How can we know what language models know? Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. TACL 2020

Negated and Misprimed Probes for Pretrained Language Models: Birds Can Talk, But Cannot Fly. Nora Kassner and Hinrich Schütze. ACL 2020

What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. Allyson Ettinger. TACL 2020

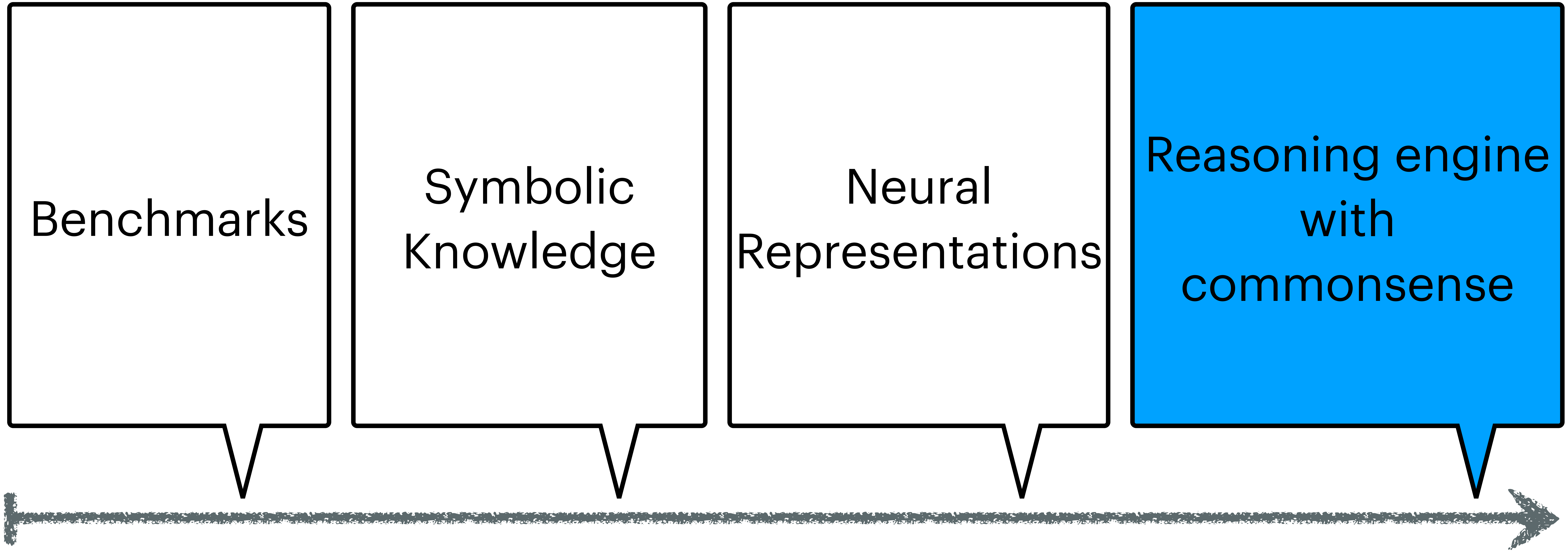
❌ Knowledge in Pre-trained LMs



- ❌ Confuse semantically-similar mutually-exclusive terms
- ❌ Are really bad with negation
- ❌ Lack perceptual knowledge (people don’t talk about it)
- ❌ Also suffer from reporting bias!

How can we know what language models know? Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. TACL 2020
Negated and Misprimed Probes for Pretrained Language Models: Birds Can Talk, But Cannot Fly. Nora Kassner and Hinrich Schütze. ACL 2020
What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. Allyson Ettinger. TACL 2020

Path to commonsense



Winograd Schema Challenge (WSC)

The city councilmen refused the demonstrators a permit because *they* **advocated** violence. Who is "*they*"?

(a) The city councilmen
(b) The demonstrators

The city councilmen refused the demonstrators a permit because *they* **feared** violence. Who is "*they*"?

(a) The city councilmen
(b) The demonstrators

The winograd schema challenge. Hector Levesque, Ernest Davis, and Leora Morgenstern. AAAI 2012.

Supervised Approach

[CLS] The city councilmen refused the demonstrators a permit because [SEP] the city councilmen advocated violence.

[CLS] The city councilmen refused the demonstrators a permit because [SEP] the demonstrators advocated violence.

0.67
0.33

A Surprisingly Robust Trick for the Winograd Schema Challenge. *Vid Kocijan, Ana-Maria Cretu, Oana-Maria Camburu, Yordan Yordanov, and Thomas Lukasiewic.* ACL 2019.

Unsupervised Approach

$$\text{argmax}_i \, P_{LM}(s_1, s_2)$$

s_1: **The city councilmen refused the demonstrators a permit because the city councilmen advocated violence.**

s_2: **The city councilmen refused the demonstrators a permit because the demonstrators advocated violence.**

A Simple Method for Commonsense Reasoning. *Trieu H. Trinh and Quoc V. Le*. arXiv 2019.

Unsupervised Approach

$$\text{argmax}_i \, P_{LM}(s_1, s_2)$$

s_1: **The city councilmen refused the demonstrators a permit because the city councilmen advocated violence.**

s_2: **The city councilmen refused the demonstrators a permit because the demonstrators advocated violence.**

$$\text{argmax}_i \sum_j P_{LM_j}(s_1, s_2)$$

A Simple Method for Commonsense Reasoning. *Trieu H. Trinh and Quoc V. Le*. arXiv 2019.

Katrina had the financial means to afford a new car while Monica did not, since ____ had a high paying job.

WINOGRANDE: An Adversarial Winograd Schema Challenge at Scale. *Keisuke Sakaguchi, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi.* AAAI 2020.

Sentence:

Katrina had the financial means to afford a new car while Monica did not, since [MASK] had a high paying job.

Predictions:

11.8% ↵

8.8% **She**

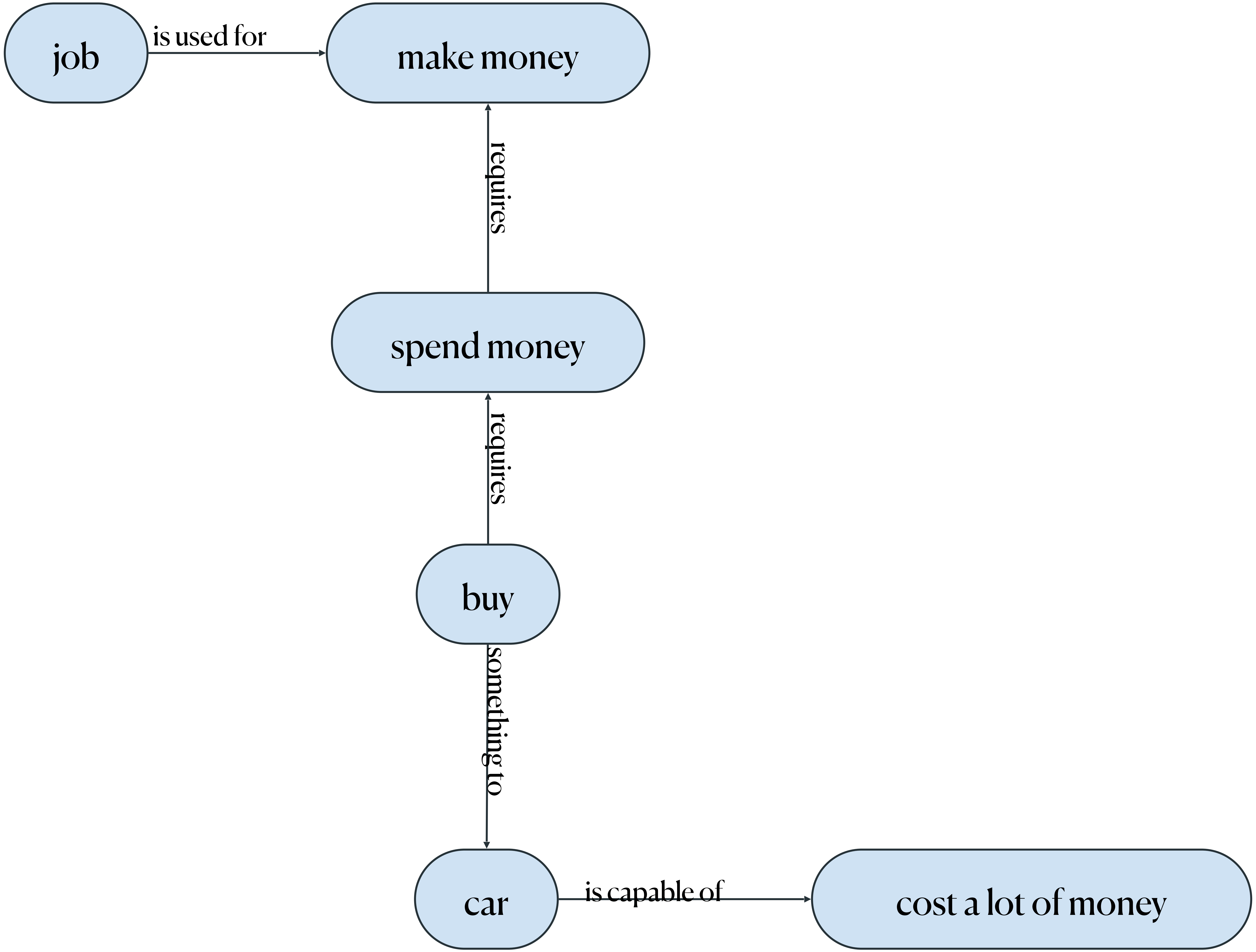
6.3% **I**

6.2% **So**

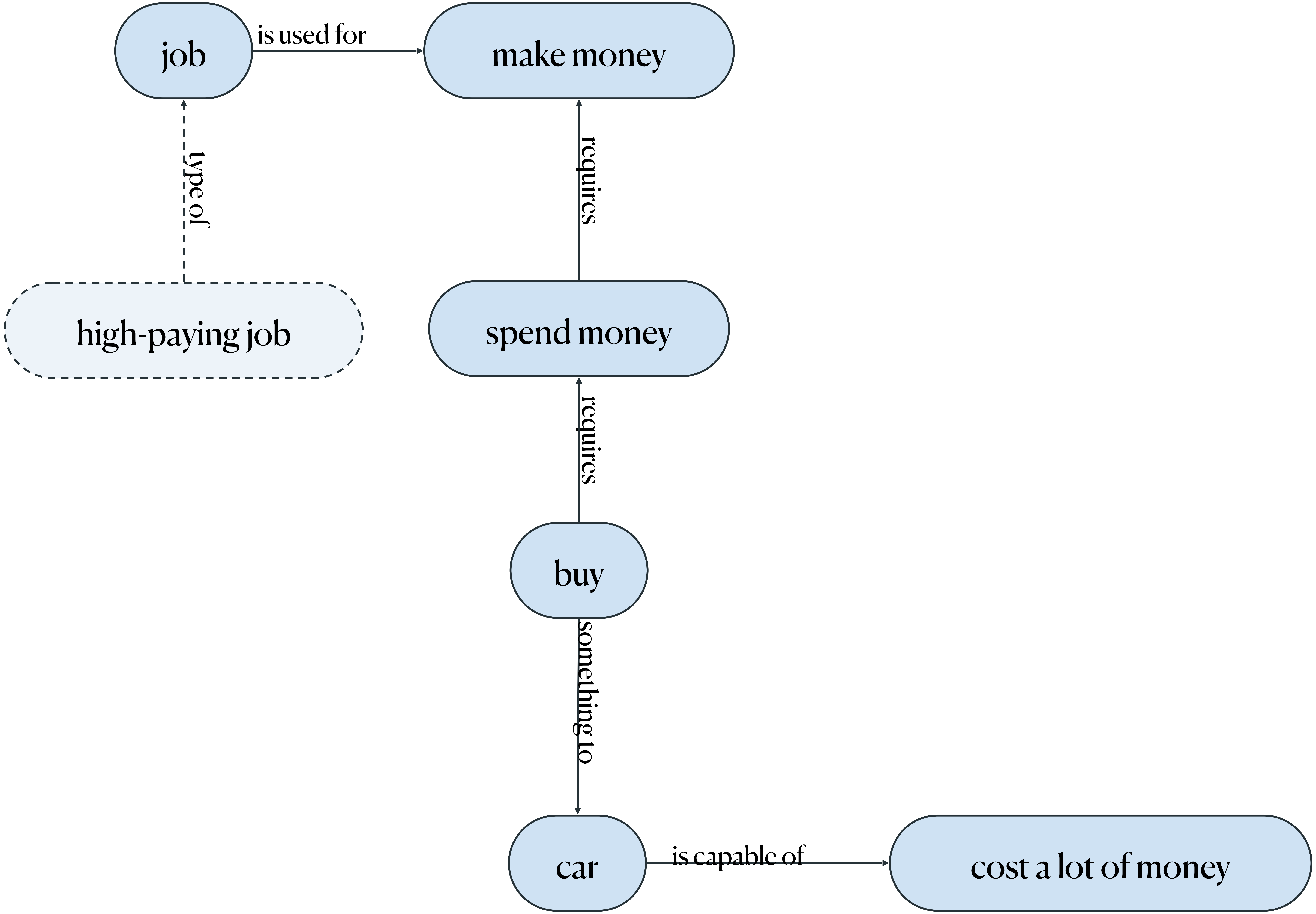
5.2% **Monica**

← **Undo**

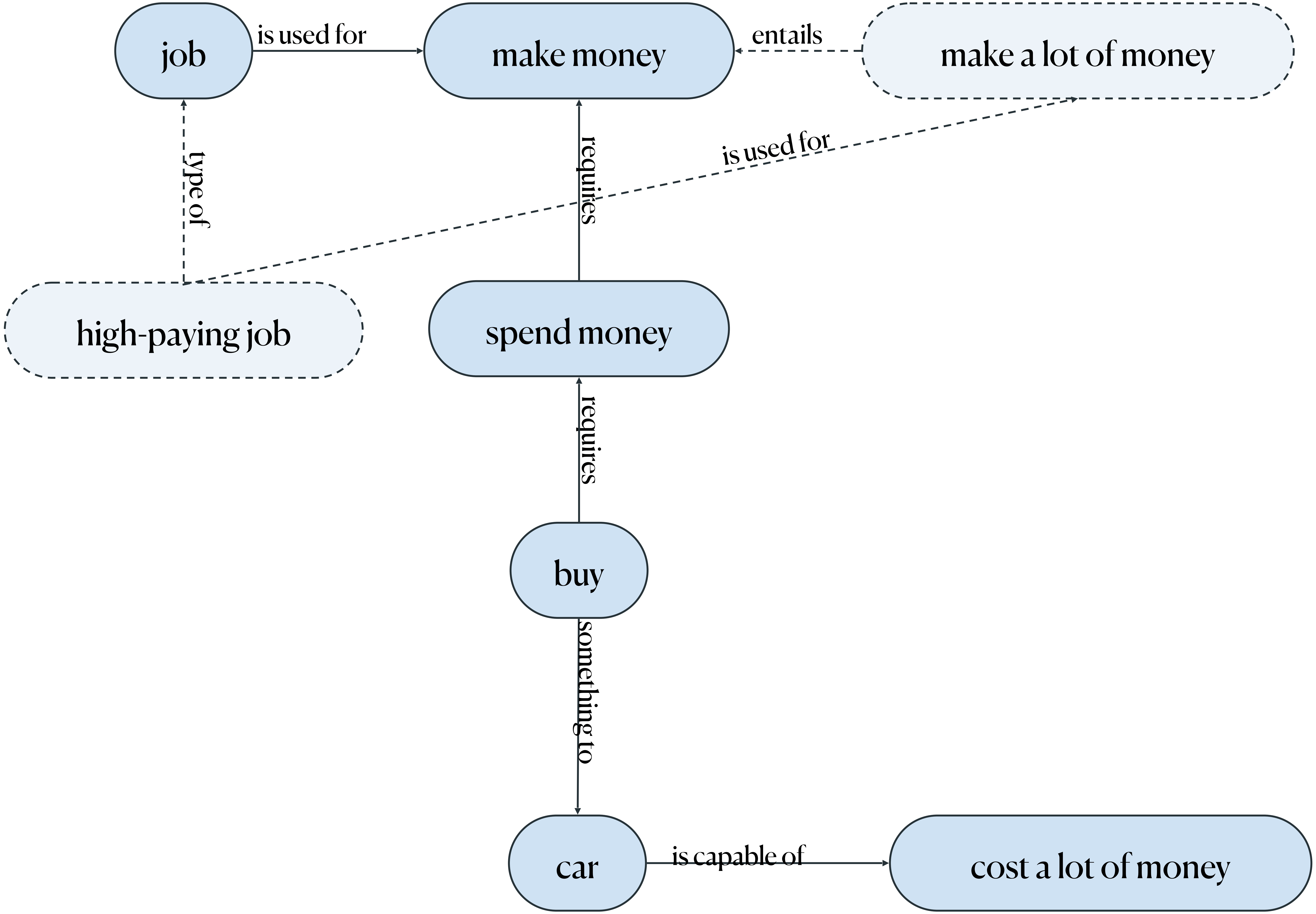
https://demo.allennlp.org/masked-lm



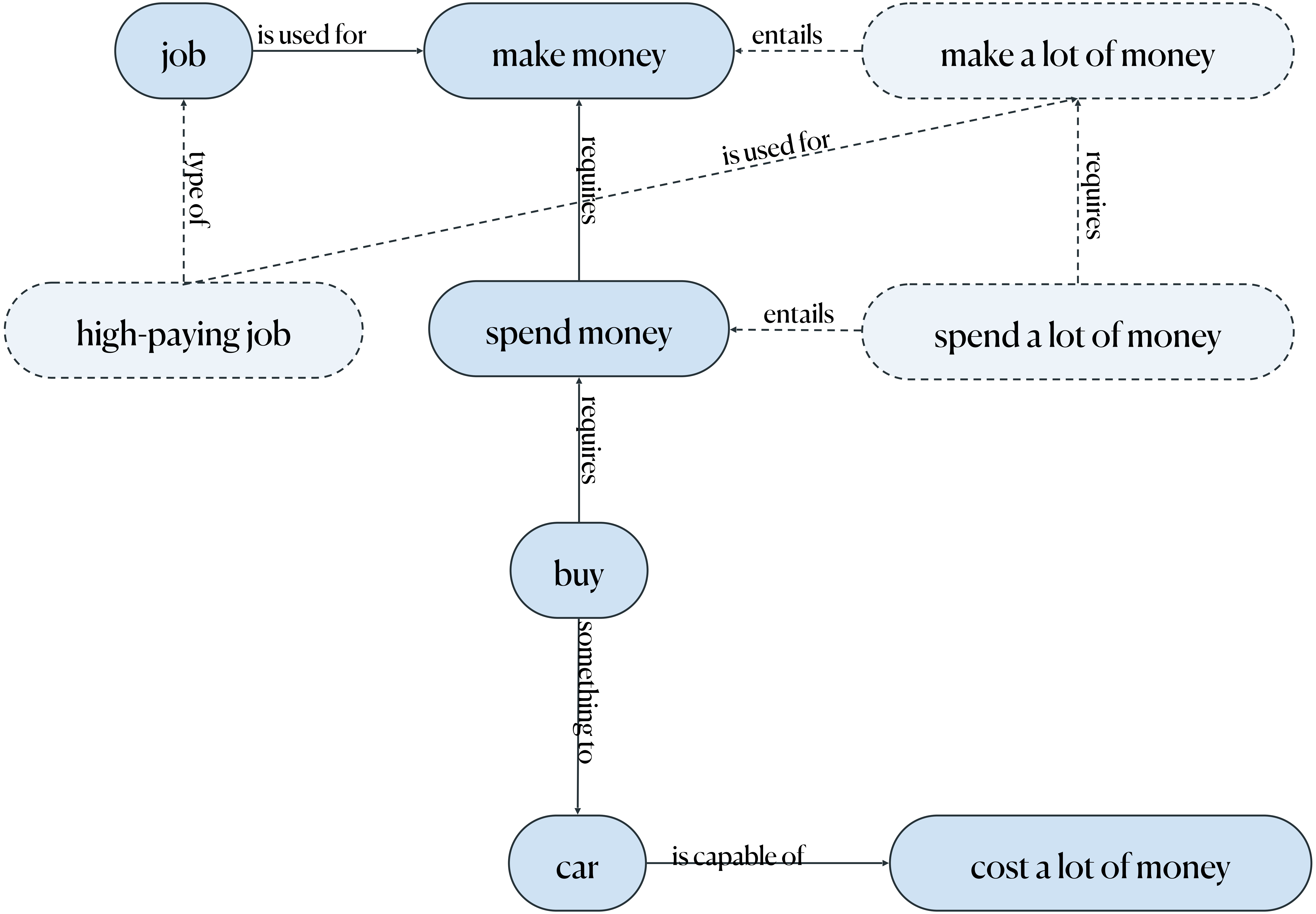
http://conceptnet5.media.mit.edu/



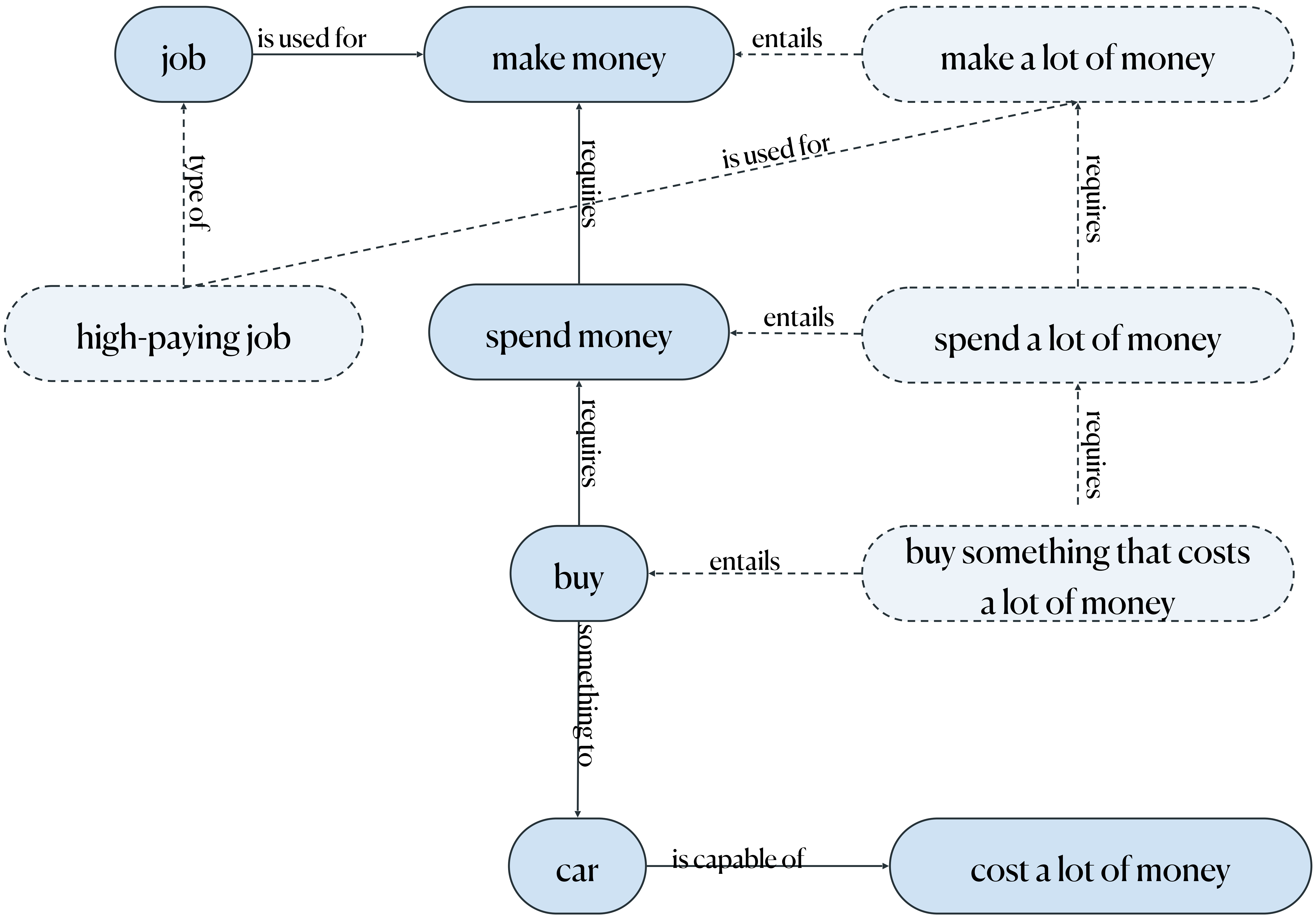
http://conceptnet5.media.mit.edu/



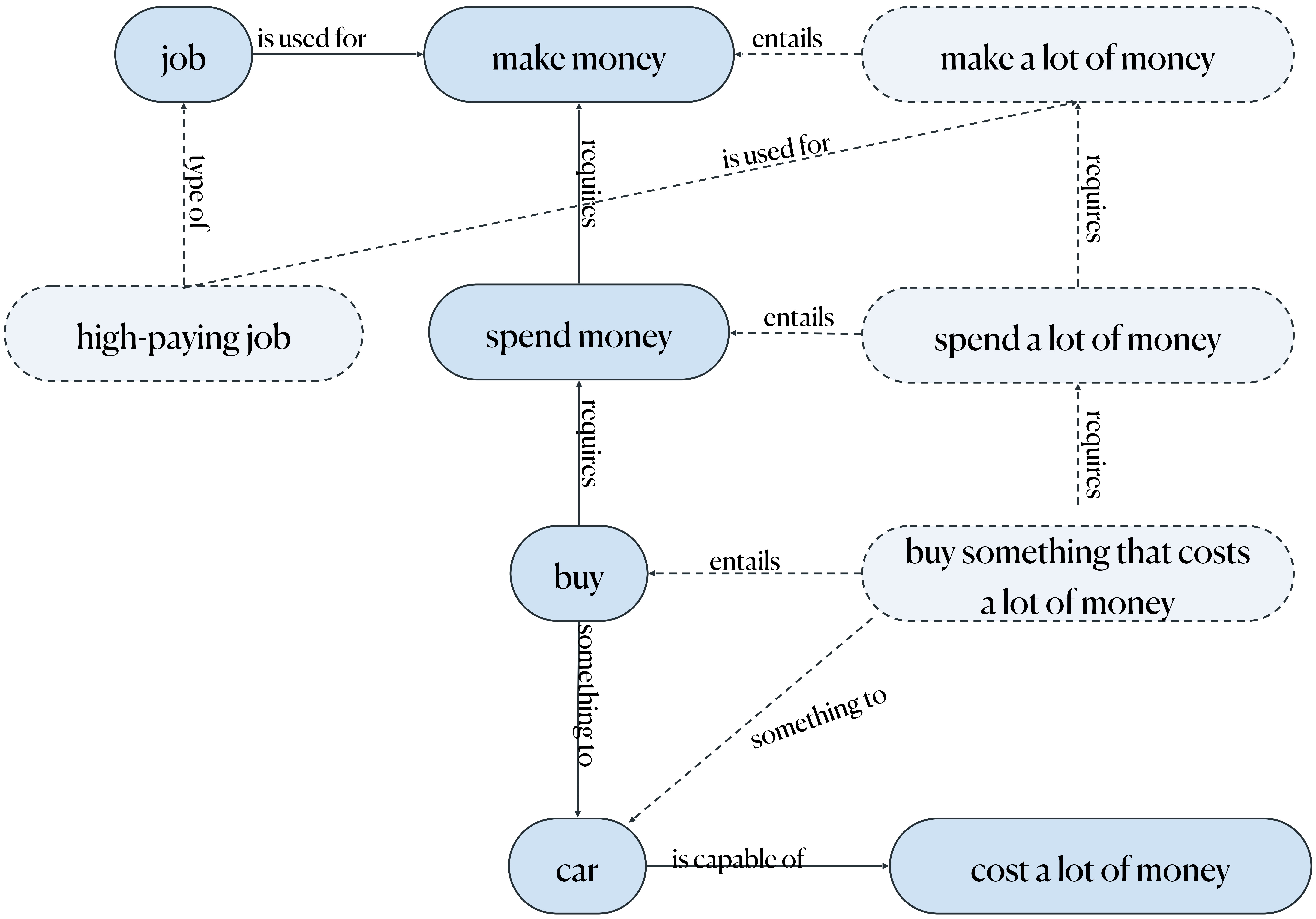
http://conceptnet5.media.mit.edu/



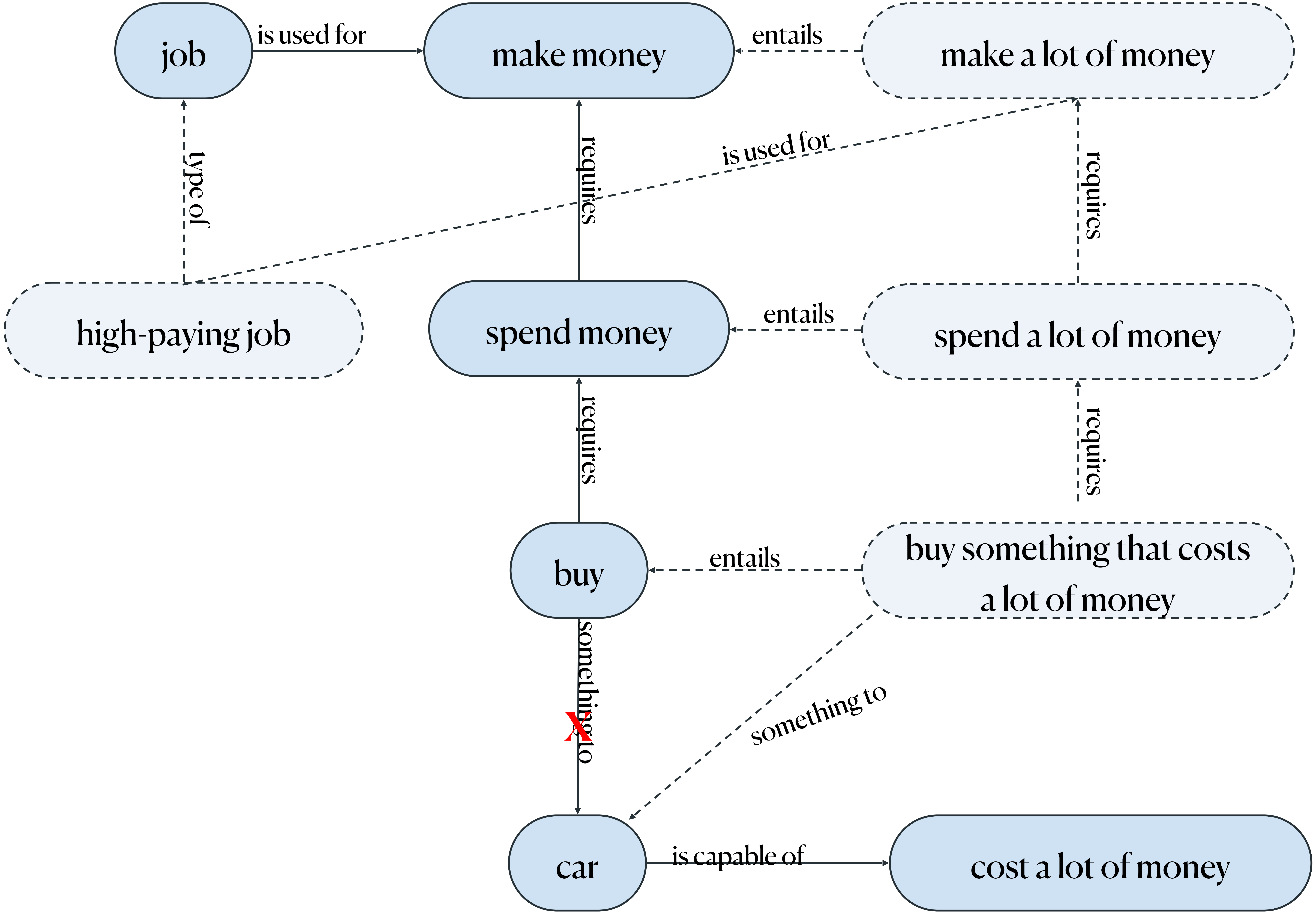
http://conceptnet5.media.mit.edu/



http://conceptnet5.media.mit.edu/

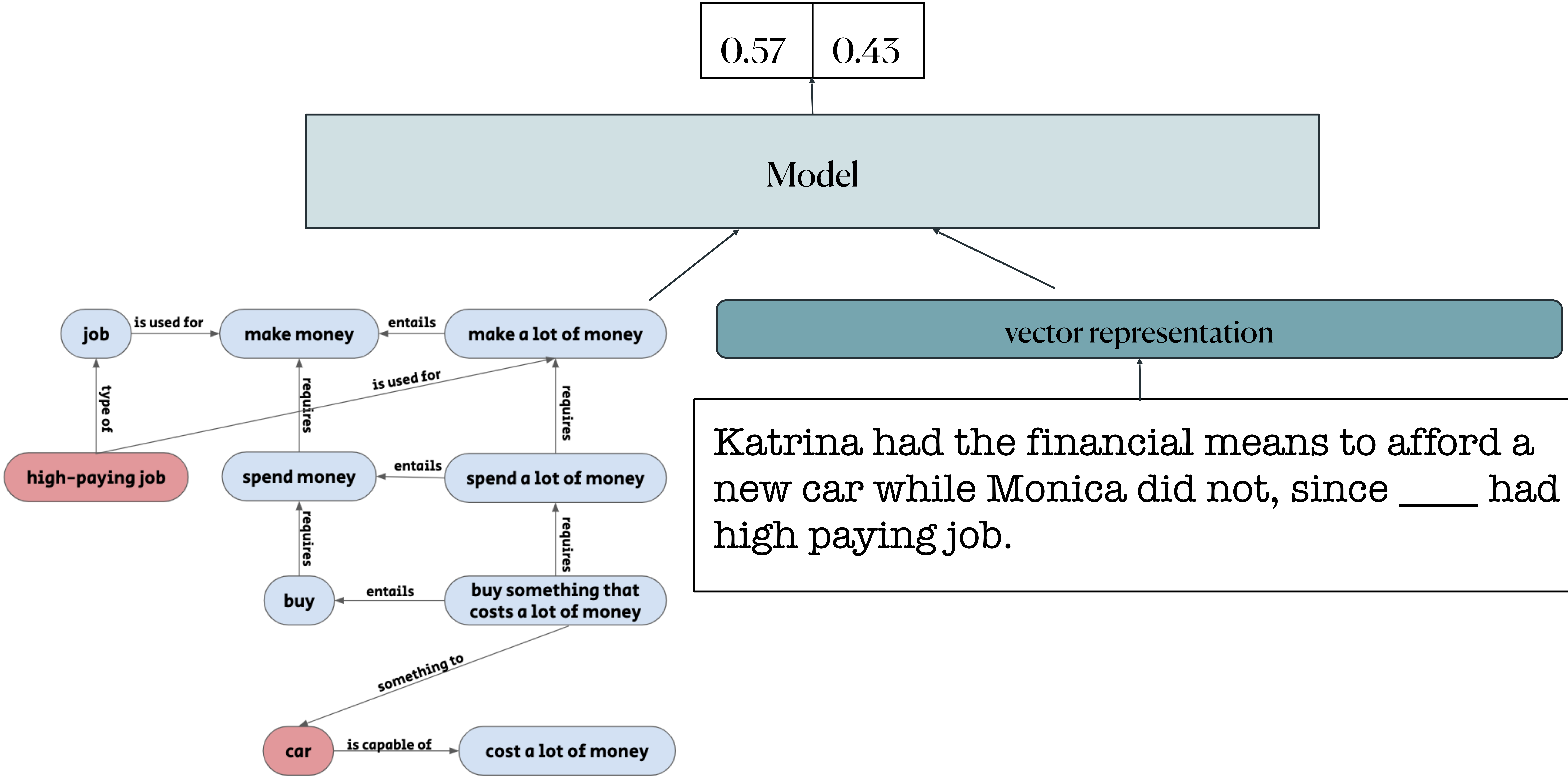


http://conceptnet5.media.mit.edu/



http://conceptnet5.media.mit.edu/

Neurosymbolic Approach



Katrina had the financial means to afford a new car while Monica did not, since ____ had a high paying job.

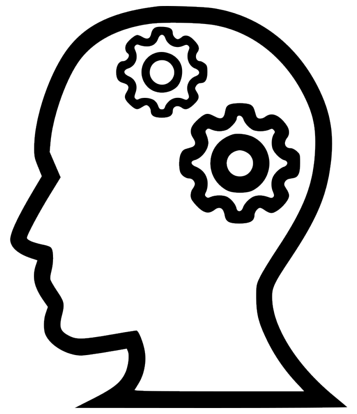
Recipe

Incorporating External Knowledge into Neural Models

Recipe

Knowledge Source

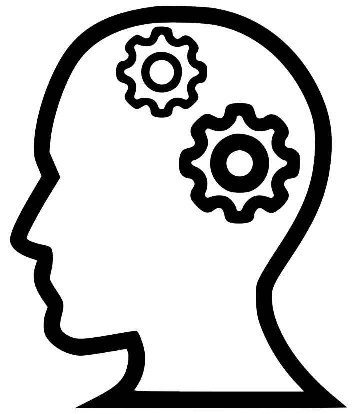
Knowledge bases, extracted from text, hand-crafted rules



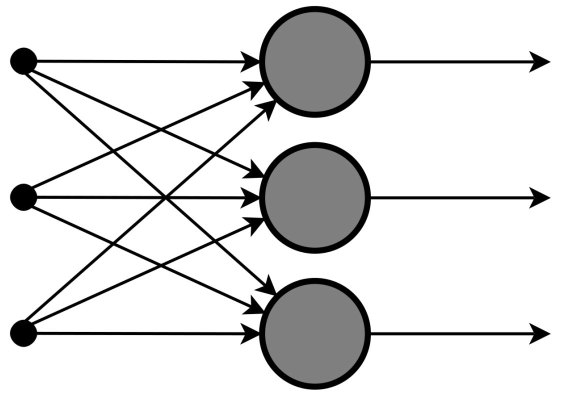
Incorporating External Knowledge into Neural Models

Recipe

Knowledge Source



Knowledge bases, extracted from text, hand-crafted rules



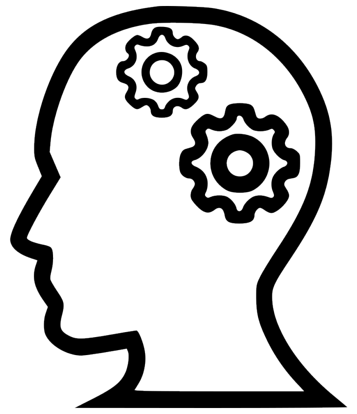
Neural Component

Pre/post pre-trained language models

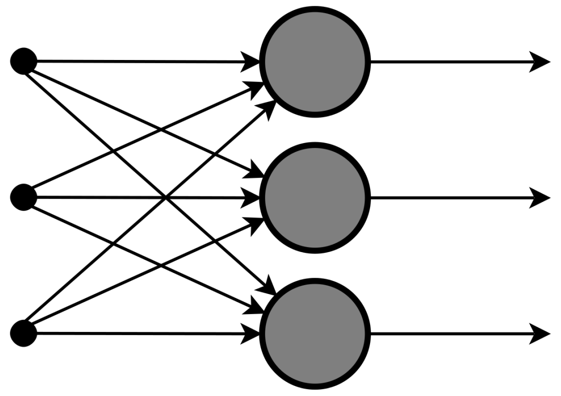
Incorporating External Knowledge into Neural Models

Recipe

Knowledge Source

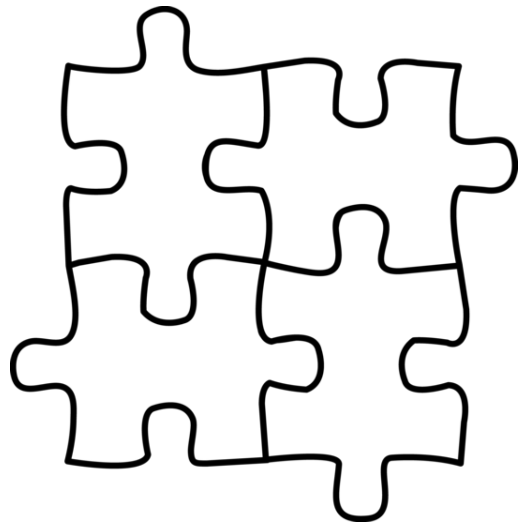


Knowledge bases, extracted from text, hand-crafted rules



Neural Component

Pre/post pre-trained language models



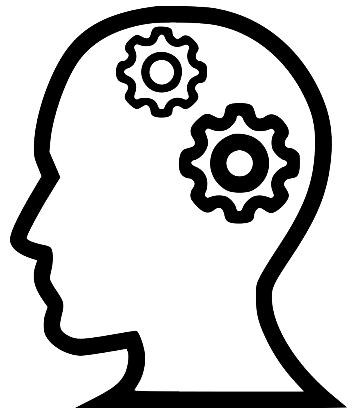
Combination Method

Attention, pruning, word embeddings, multi-task learning

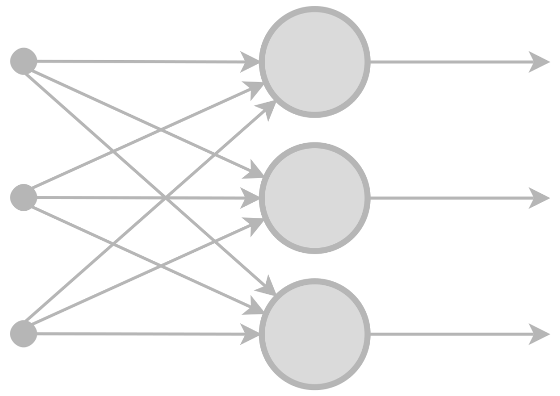
Incorporating External Knowledge into Neural Models

Recipe

Knowledge Source

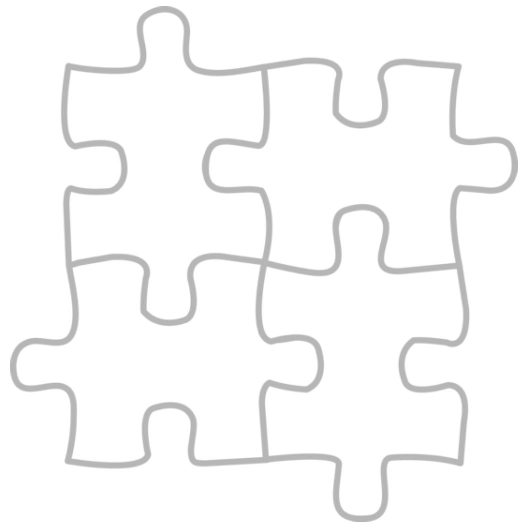


Knowledge bases, extracted from text, hand-crafted rules



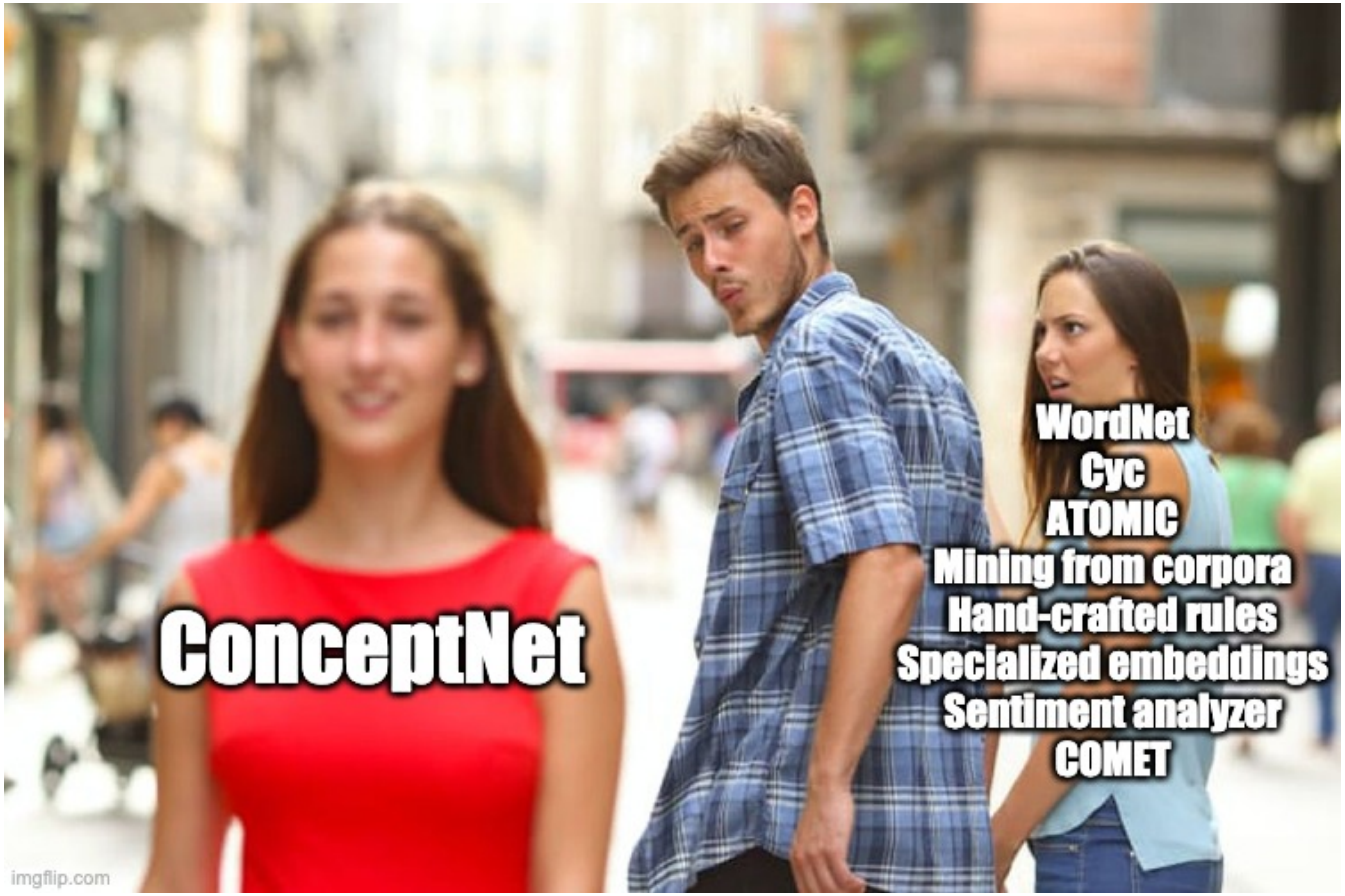
Neural Component

Pre/post pre-trained language models



Combination Method

Attention, pruning, word embeddings, multi-task learning

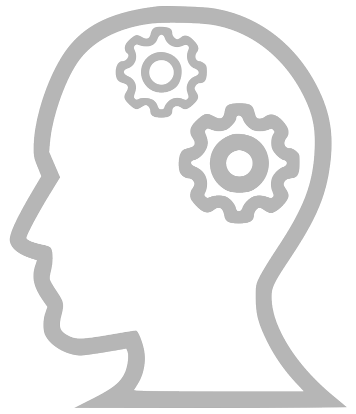

ConceptNet
WordNet
Cyc
ATOMIC
Mining from corpora
Hand-crafted rules
Specialized embeddings
Sentiment analyzer
COMET
imgflip.com

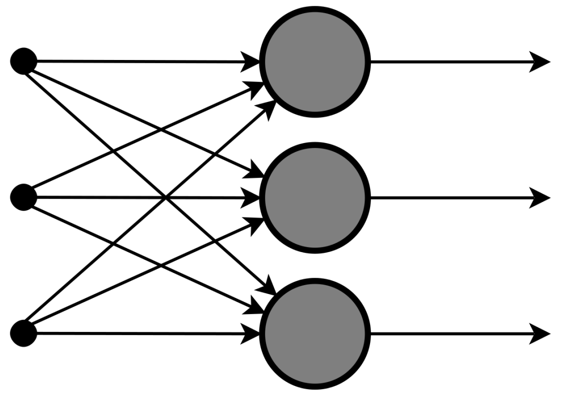
Incorporating External Knowledge into Neural Models

Recipe

Knowledge Source

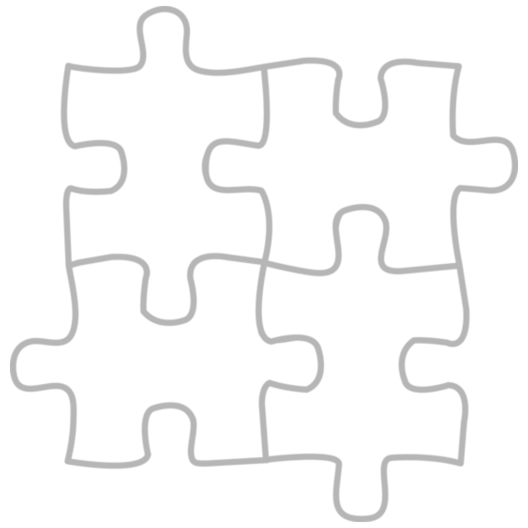
Knowledge bases, extracted from text, hand-crafted rules





Neural Component

Pre/post pre-trained language models



Combination Method

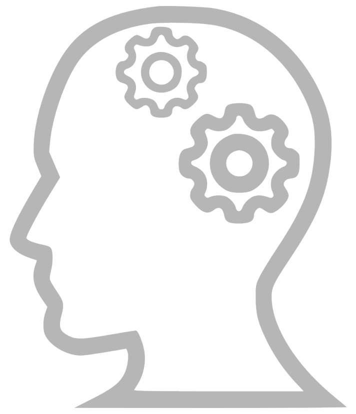
Attention, pruning, word embeddings, multi-task learning

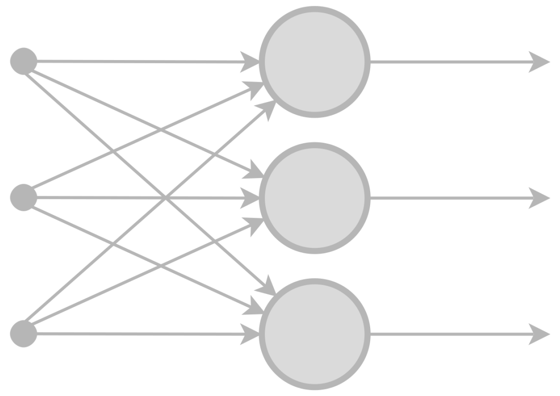
Incorporating External Knowledge into Neural Models

Recipe

Knowledge Source

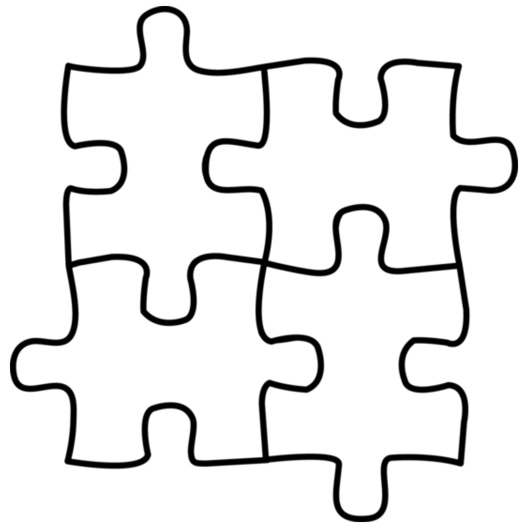
Knowledge bases, extracted from text, hand-crafted rules





Neural Component

Pre/post pre-trained language models

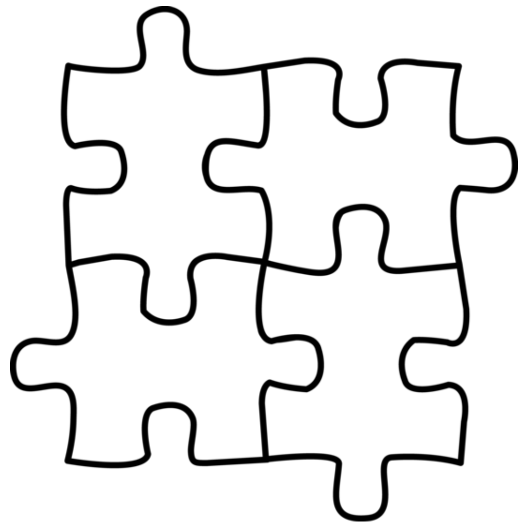


Combination Method

Attention, pruning, word embeddings, multi-task learning

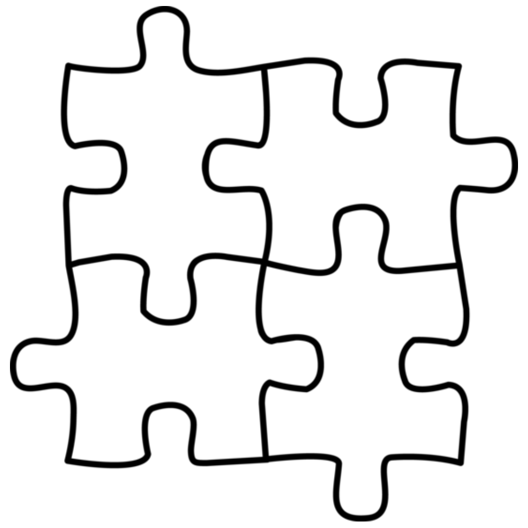
Combination Method

- Incorporate into scoring function
- Multi-task learning
- Symbolic → vector representation (+attention)



Combination Method

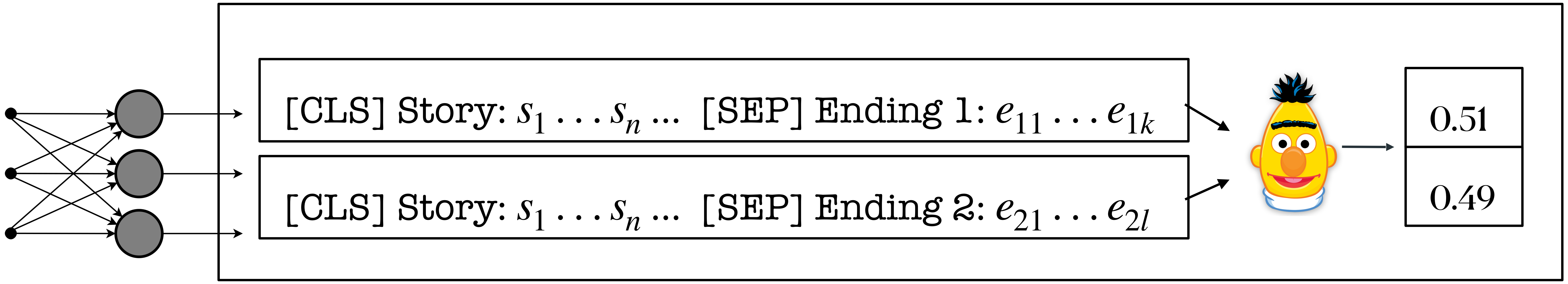
- Incorporate into scoring function
- Multi-task learning
- Symbolic → vector representation (+attention)



Multitask Learning

Incorporating Relation Knowledge into Commonsense Reading Comprehension with Multi-task Learning. *Jiangnan Xia, Chen Wu, and Ming Yan*. CIKM 2019.

Multitask Learning

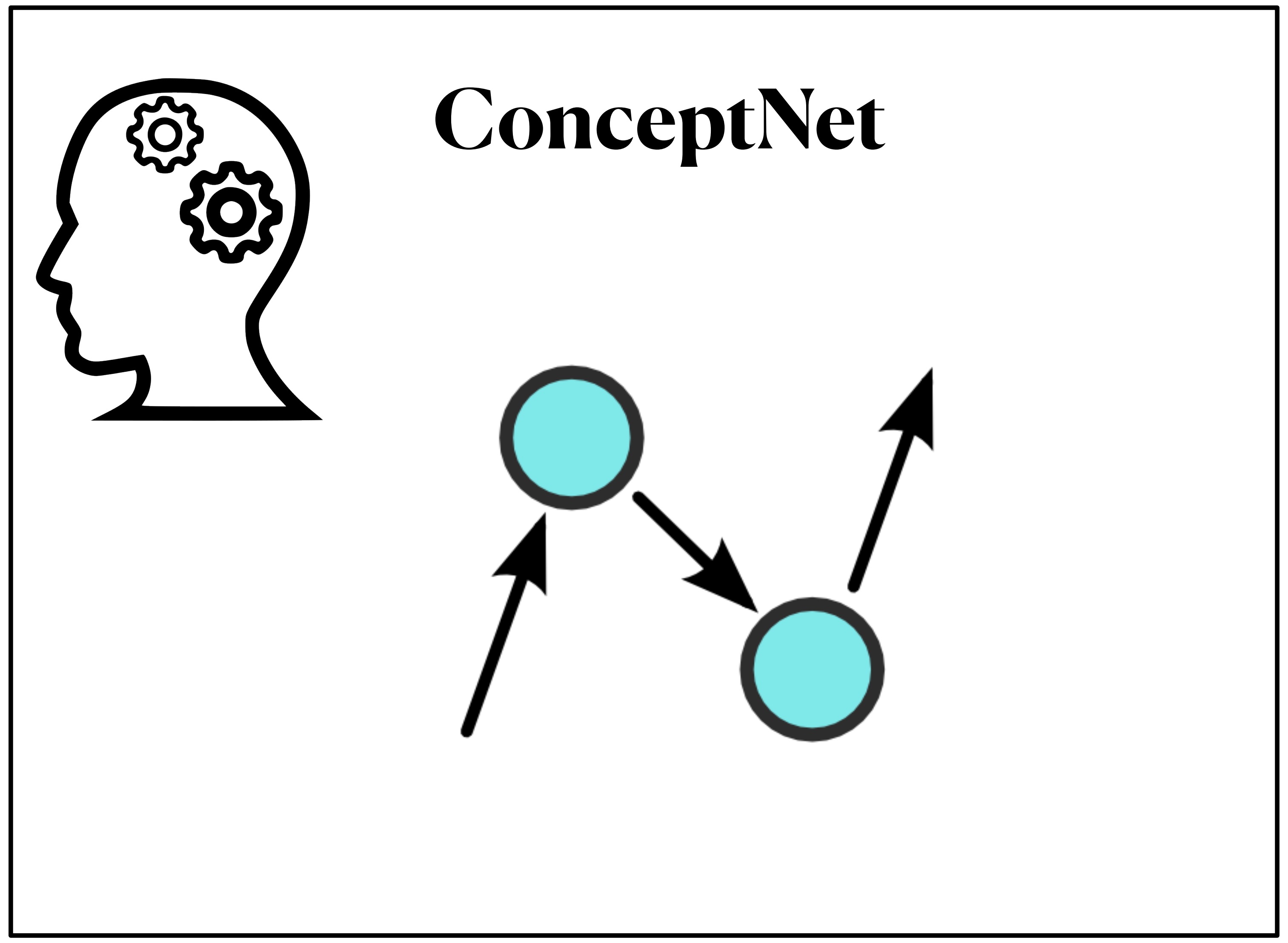


Incorporating Relation Knowledge into Commonsense Reading Comprehension with Multi-task Learning. *Jiangnan Xia, Chen Wu, and Ming Yan*. CIKM 2019.

Multitask Learning

Incorporating Relation Knowledge into Commonsense Reading Comprehension with Multi-task Learning. *Jiangnan Xia, Chen Wu, and Ming Yan*. CIKM 2019.

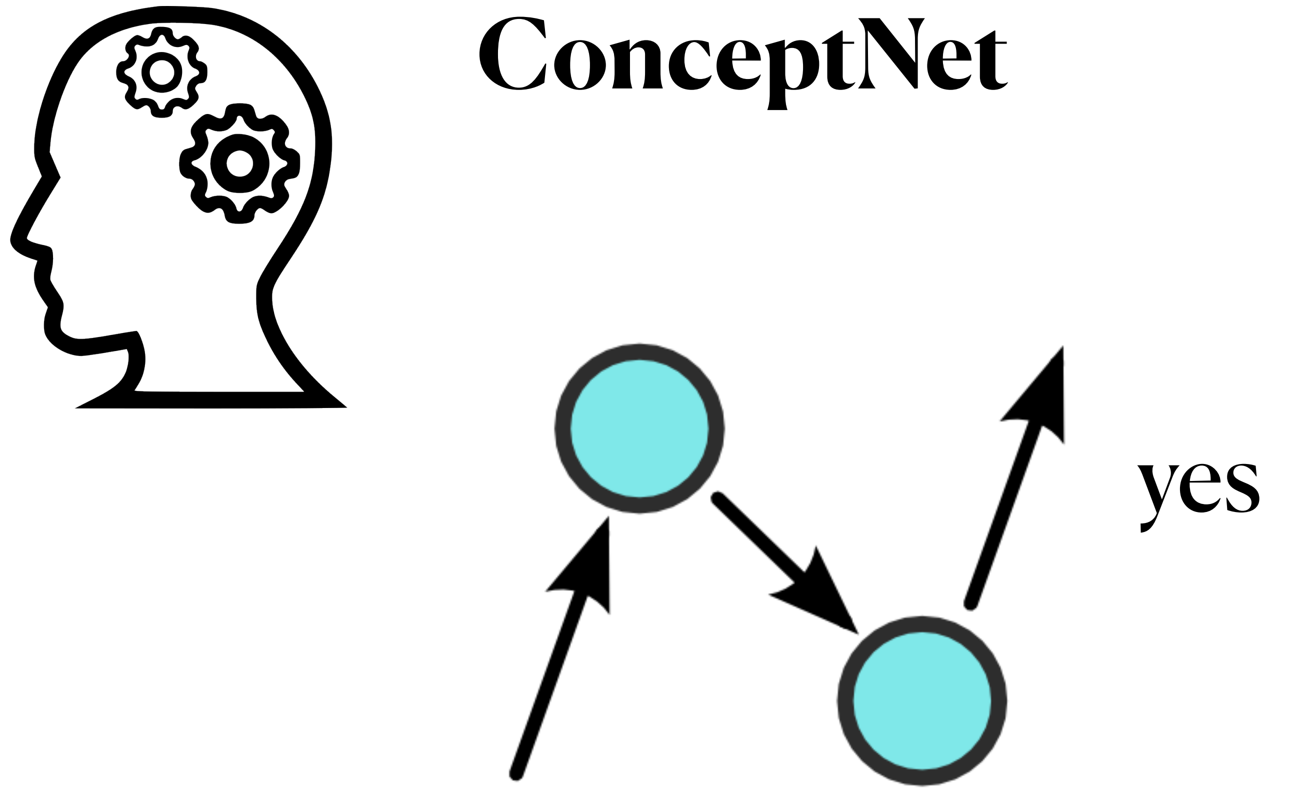
Multitask Learning



Incorporating Relation Knowledge into Commonsense Reading Comprehension with Multi-task Learning. *Jiangnan Xia, Chen Wu, and Ming Yan*. CIKM 2019.

Multitask Learning

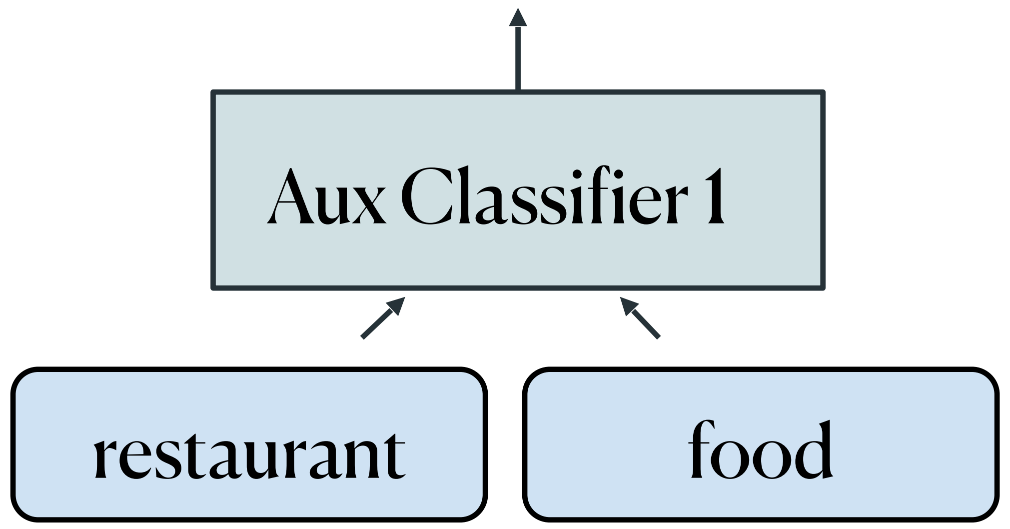
ConceptNet



Auxiliary Task

S_i = restaurant E_j = food

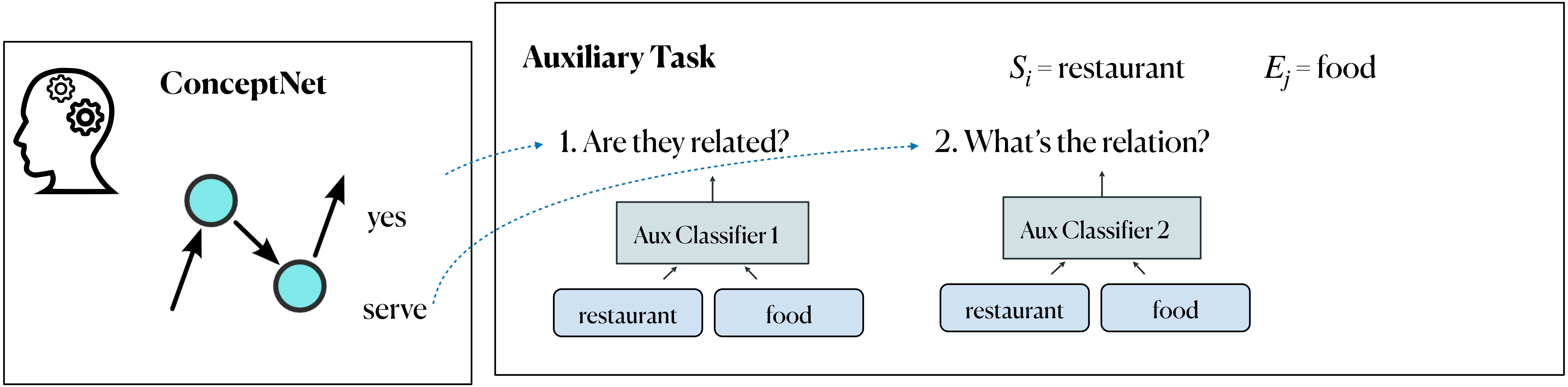
1. Are they related?



Incorporating Relation Knowledge into Commonsense Reading Comprehension with Multi-task Learning. *Jiangnan Xia, Chen Wu, and Ming Yan*. CIKM 2019.

Incorporating External Knowledge into Neural Models

Multitask Learning



Incorporating Relation Knowledge into Commonsense Reading Comprehension with Multi-task Learning. *Jiangnan Xia, Chen Wu, and Ming Yan*. CIKM 2019.

Limitations of Neurosymbolic Methods

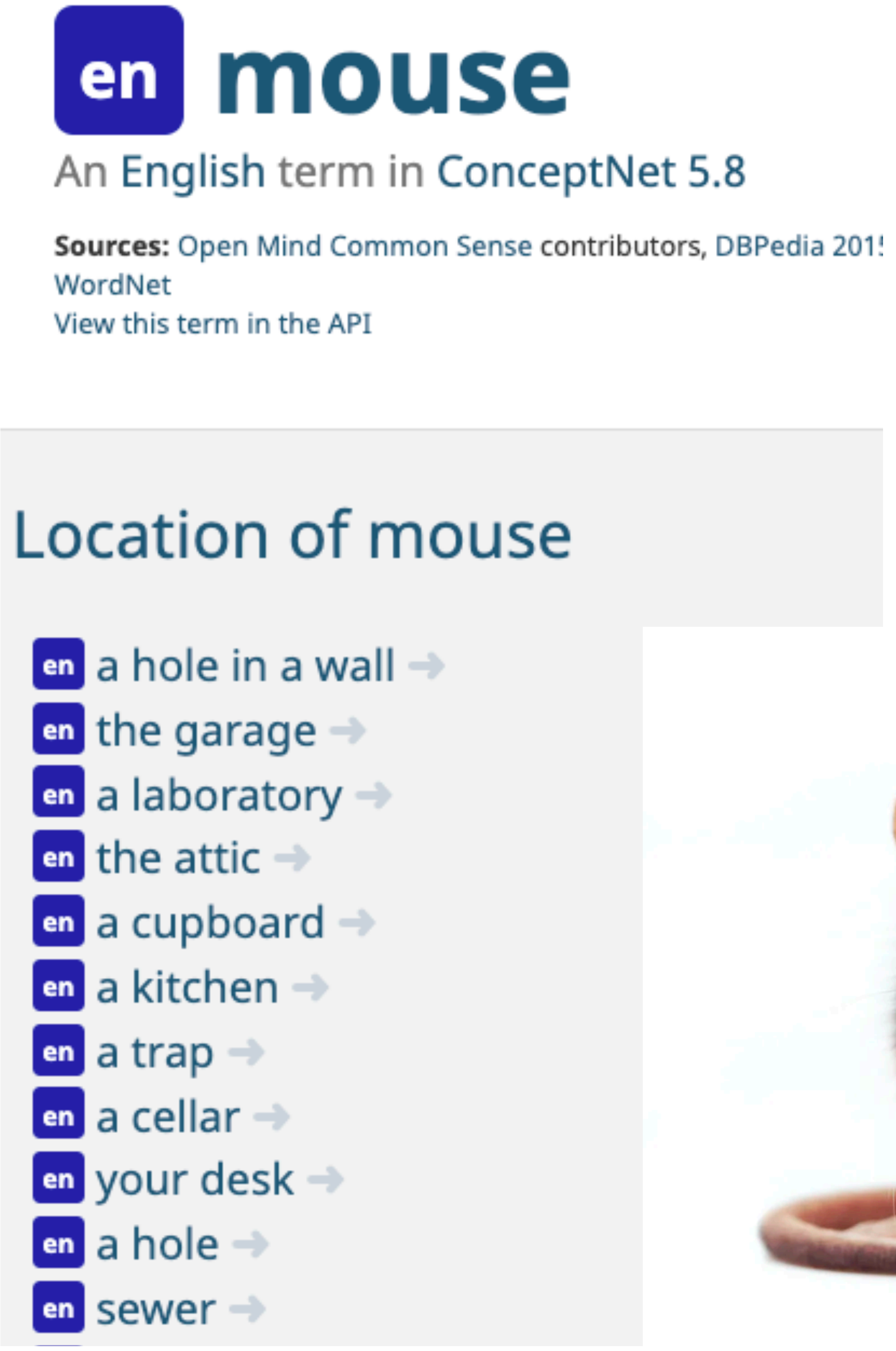
Limitations of Neurosymbolic Methods

- Knowledge graphs have **limited coverage**

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

Limitations of Neurosymbolic Methods

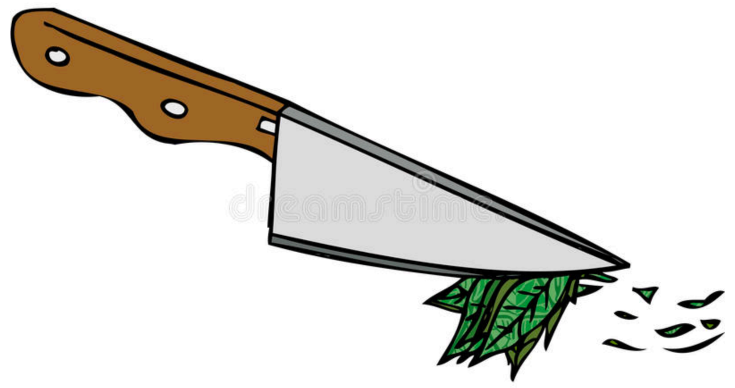
- Knowledge graphs have **limited coverage**
- Inferences may be correct only in certain **contexts**



Limitations of Neurosymbolic Methods

- Knowledge graphs have **limited coverage**
- Inferences may be correct only in certain **contexts**
- Long KB paths have **limited precision**

Kitchen $\xleftarrow{\text{location}}$ **Knife** $\xrightarrow{\text{capable of}}$ **Kill**



Limitations of Neurosymbolic Methods

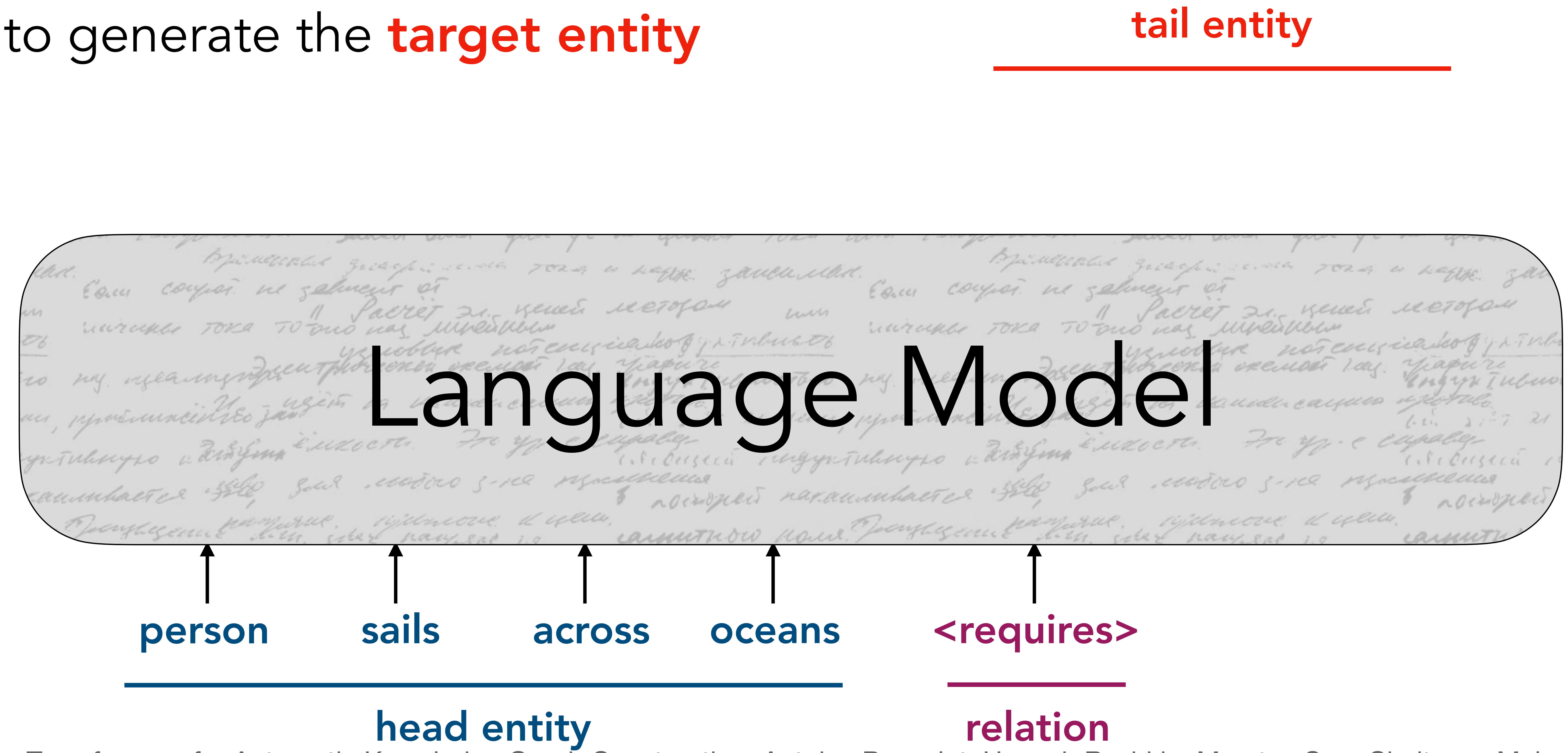
- Knowledge graphs have **limited coverage**
- Inferences may be correct only in certain **contexts**
- Long KB paths have **limited precision**
- Tradeoff: embedding knowledge (**better generalization**) vs. hard constraints (**more accurate**)

Limitations of Neurosymbolic Methods

- Knowledge graphs have **limited coverage**
- Inferences may be correct only in certain **contexts**
- Long KB paths have **limited precision**
- Tradeoff: embedding knowledge (**better generalization**) vs. hard constraints (**more accurate**)

COMET

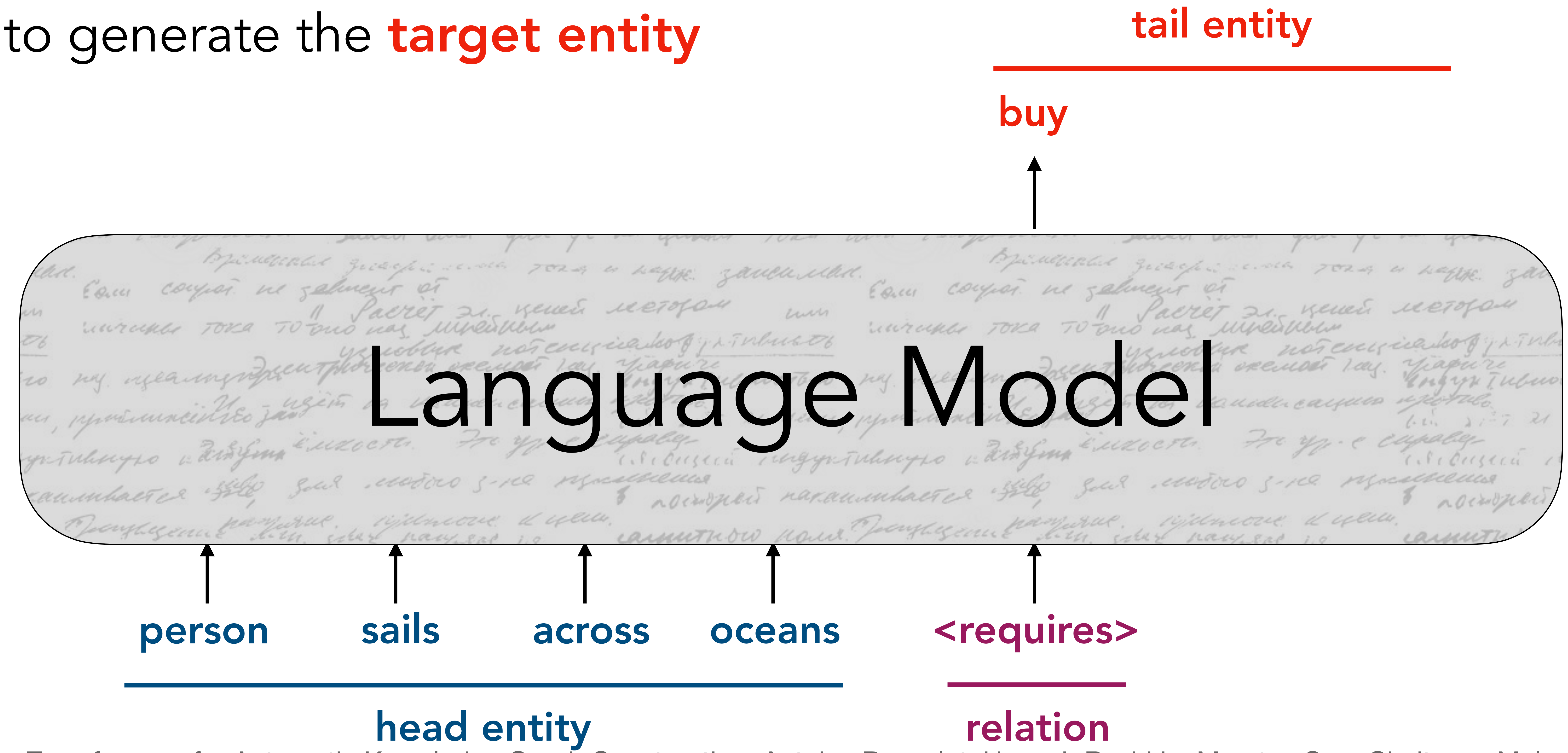
Given a **seed entity** and a **relation**,
learn to generate the **target entity**



COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. ACL 2020

COMET

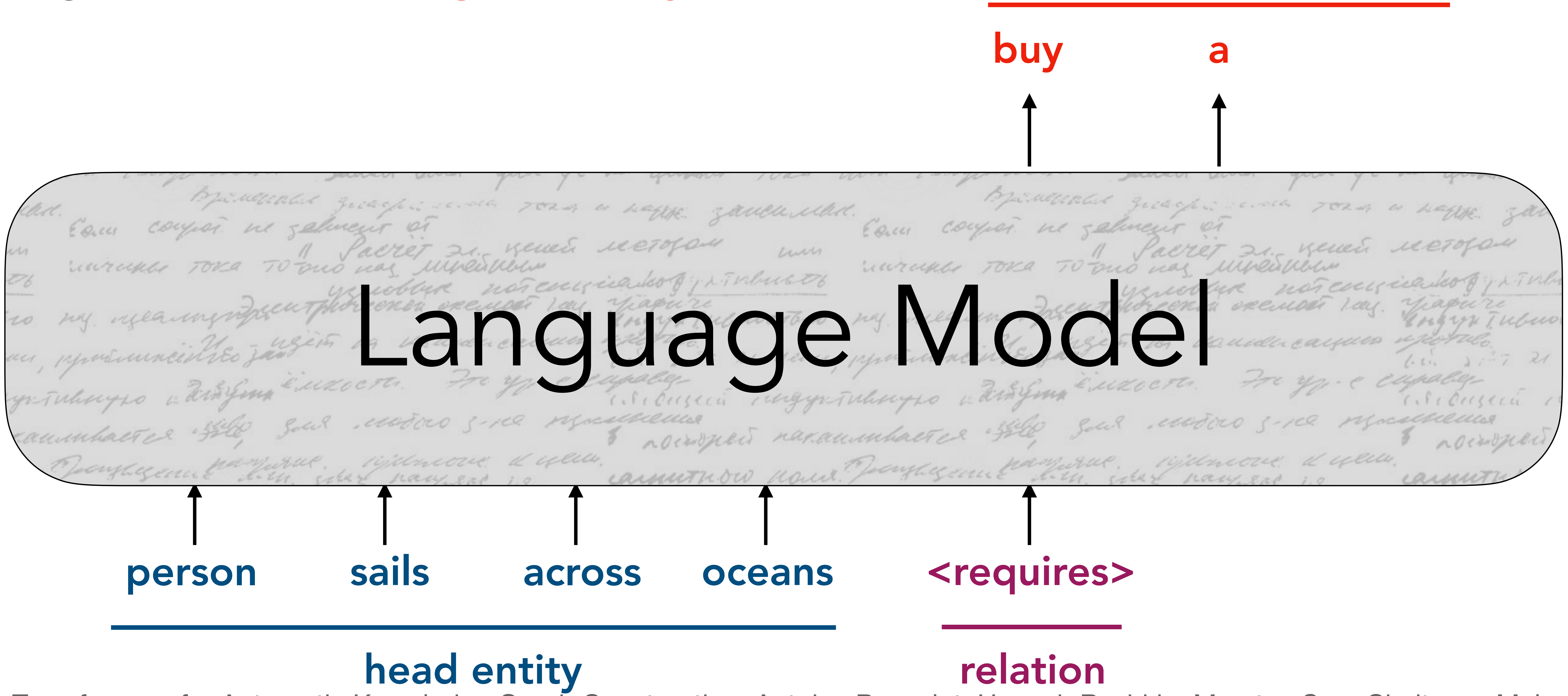
Given a **seed entity** and a **relation**,
learn to generate the **target entity**



COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. ACL 2020

COMET

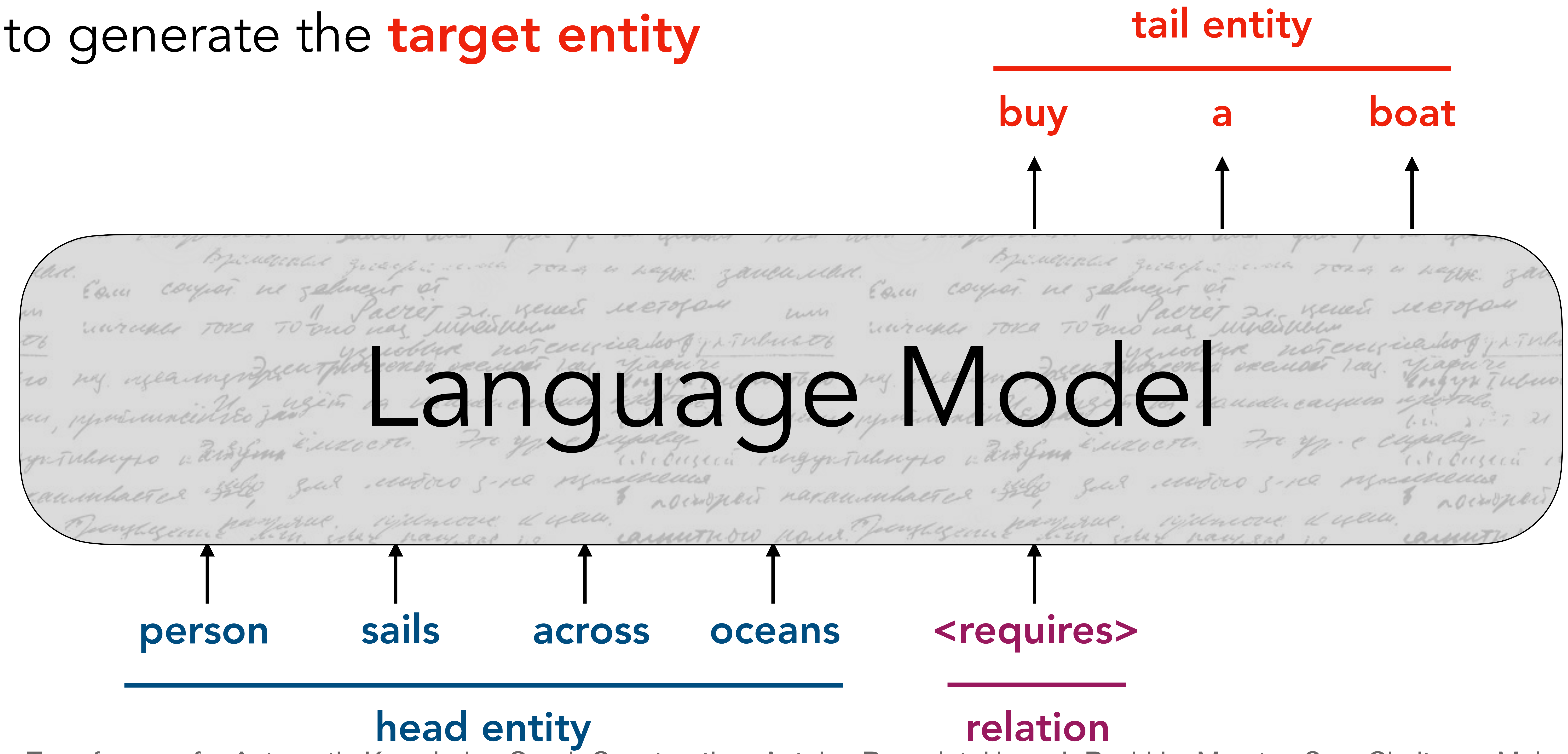
Given a **seed entity** and a **relation**,
learn to generate the **target entity**



COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. ACL 2020

COMET

Given a **seed entity** and a **relation**,
learn to generate the **target entity**

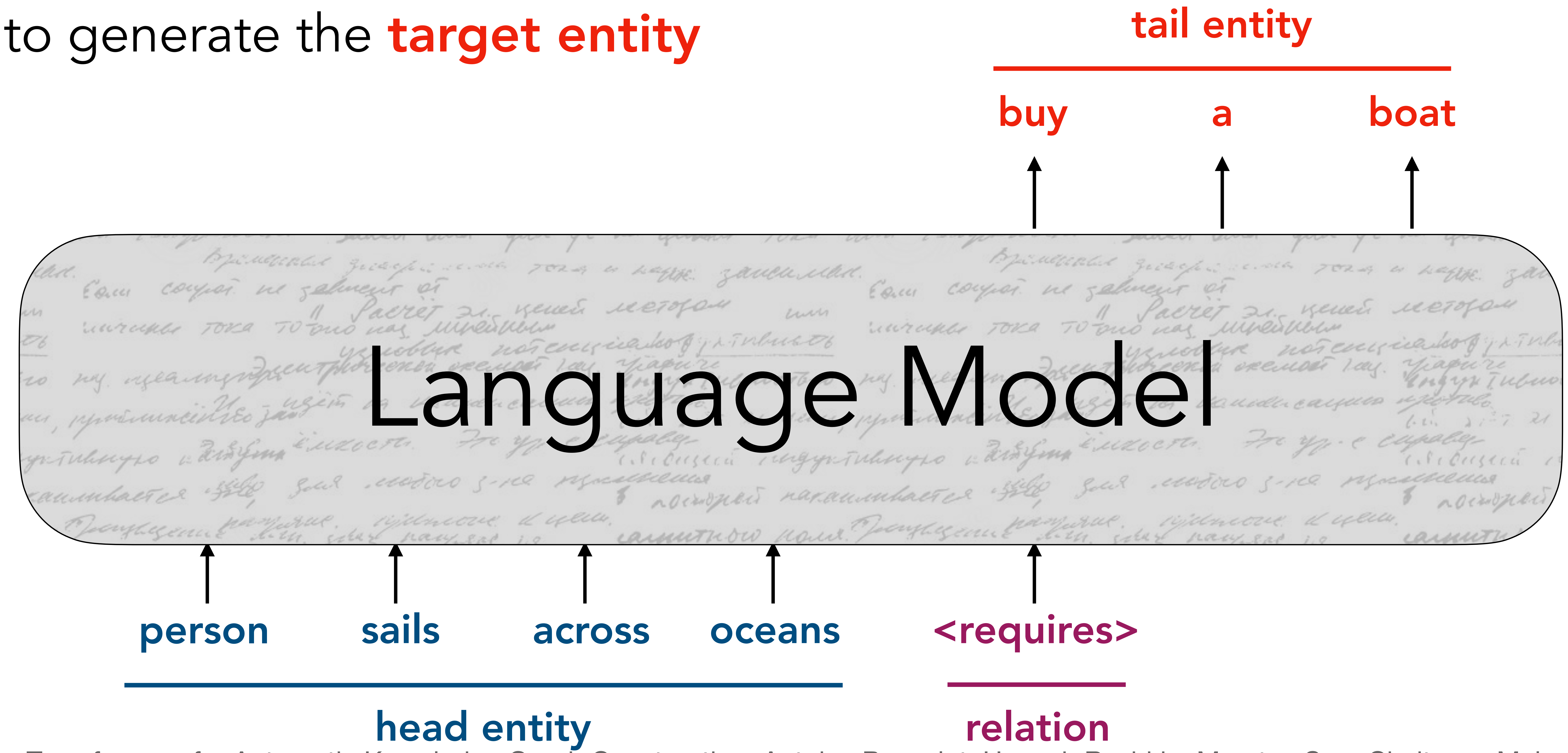


COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. ACL 2020

COMET

Given a **seed entity** and a **relation**, learn to generate the **target entity**

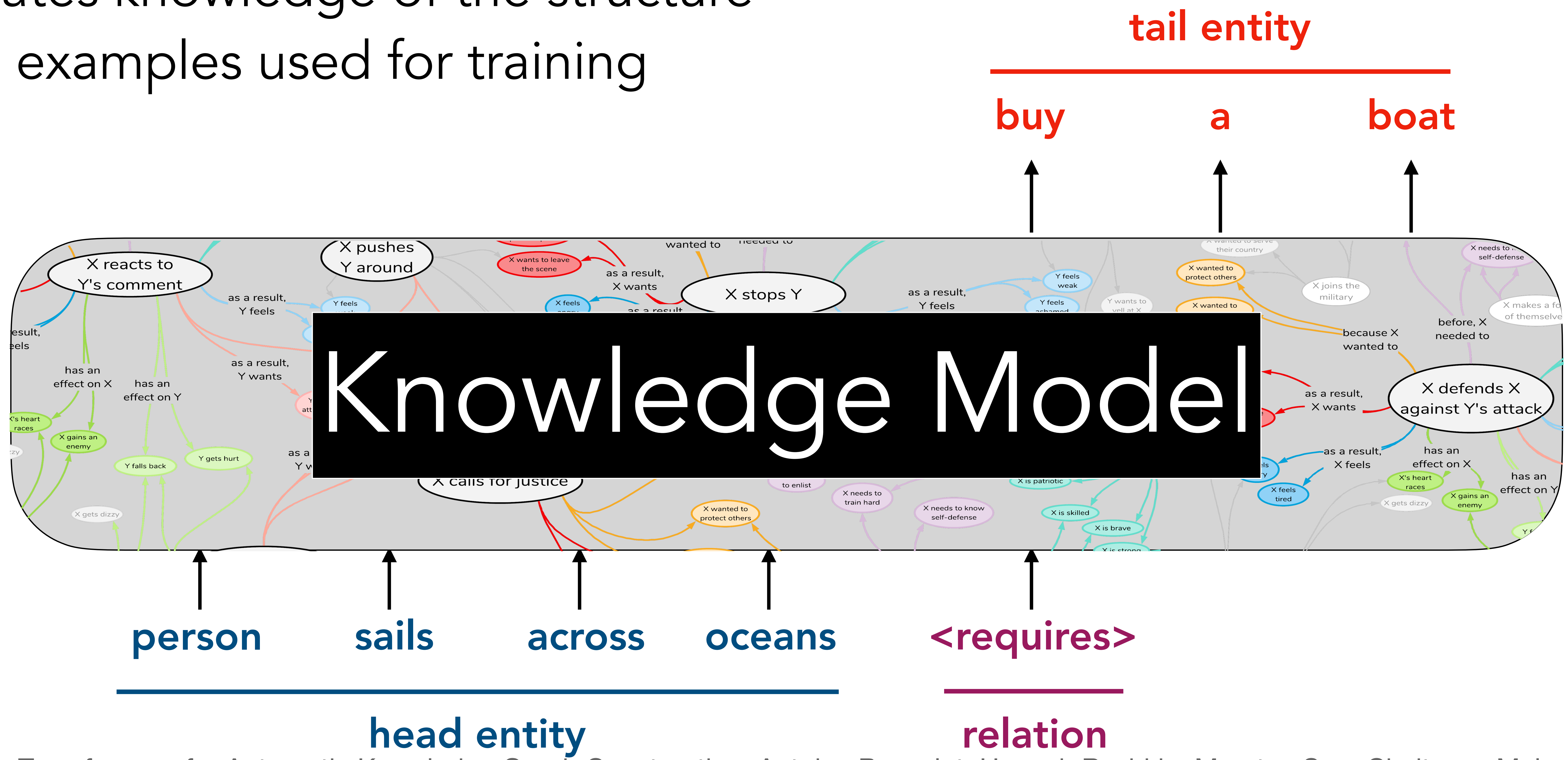
$$\mathscr{L} = -\sum \log P(\textbf{target words} \,|\, \textbf{seed words, relation})$$



COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. ACL 2020

COMET

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



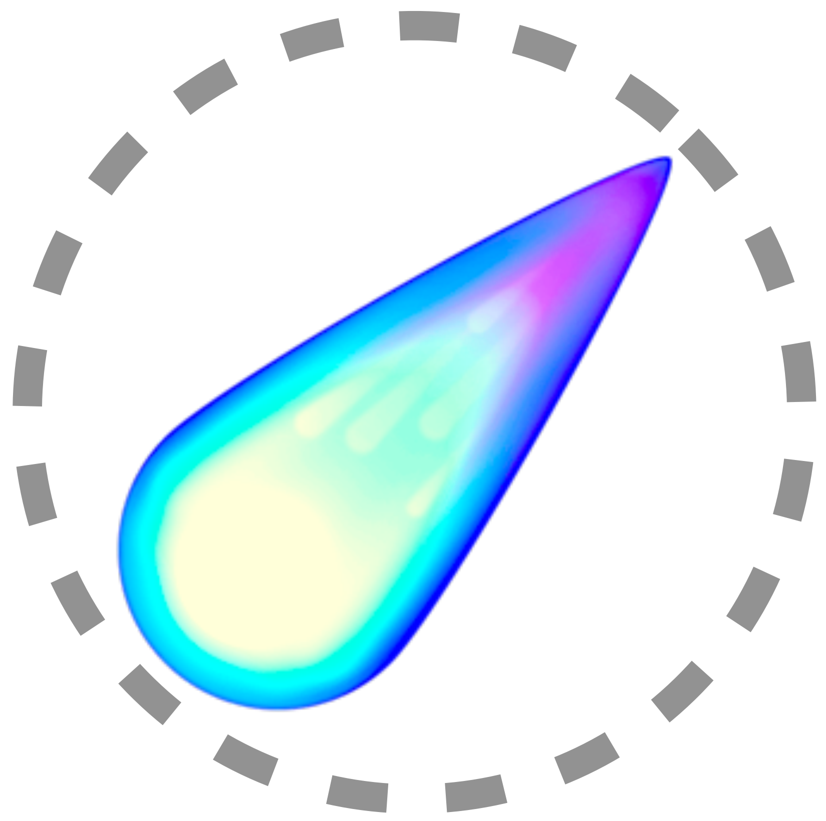
COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. ACL 2020

Generate commonsense knowledge for **any** input concept

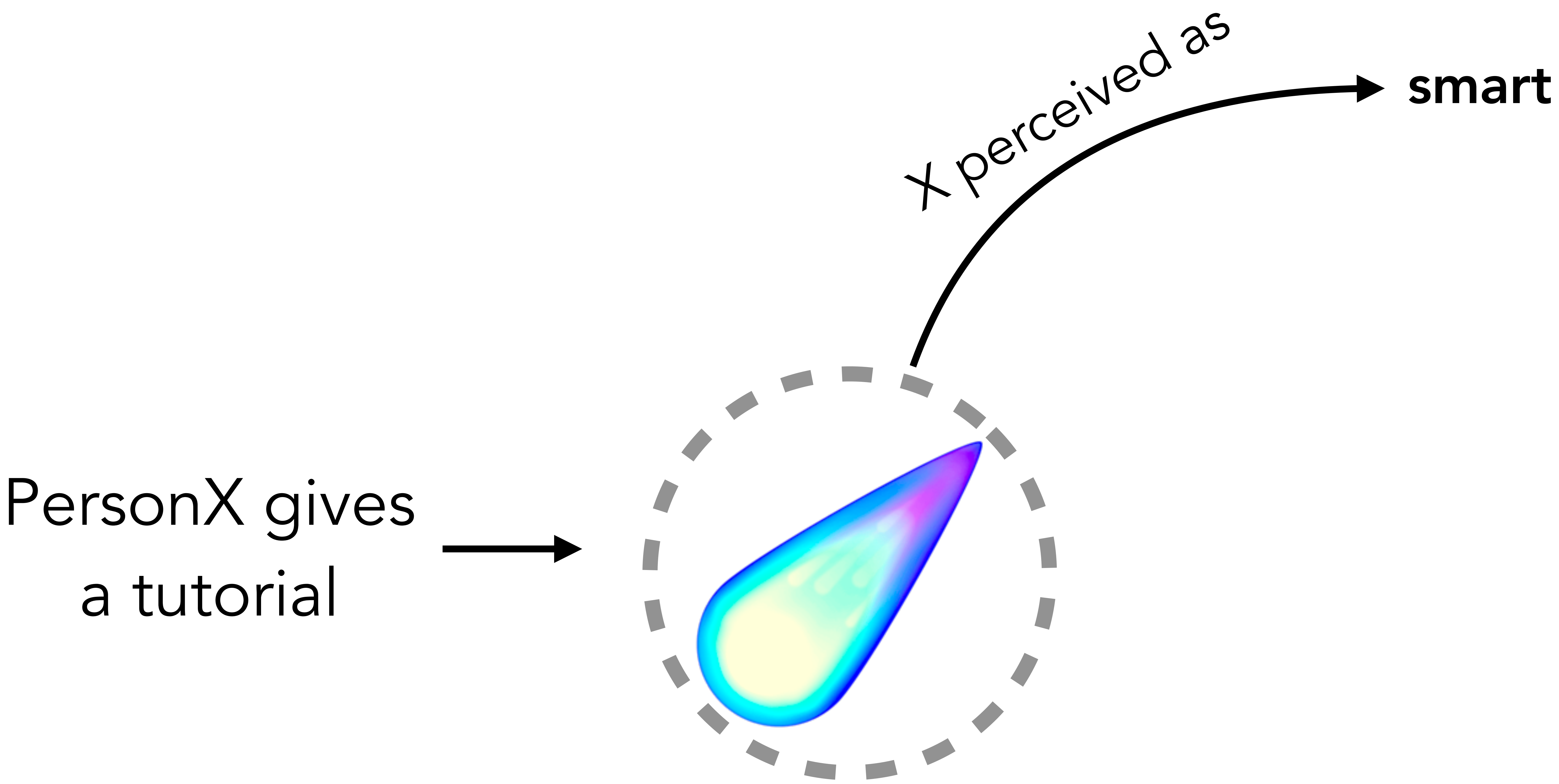

Generate commonsense knowledge for **any** input concept
COMmons**E**nse **T**ransformers
X reacts to Y's comment
X joins the military
X catches a robber in action
X pushes back
X resists Y's attack
X defends X against Y's attack

COMET - ATOMIC

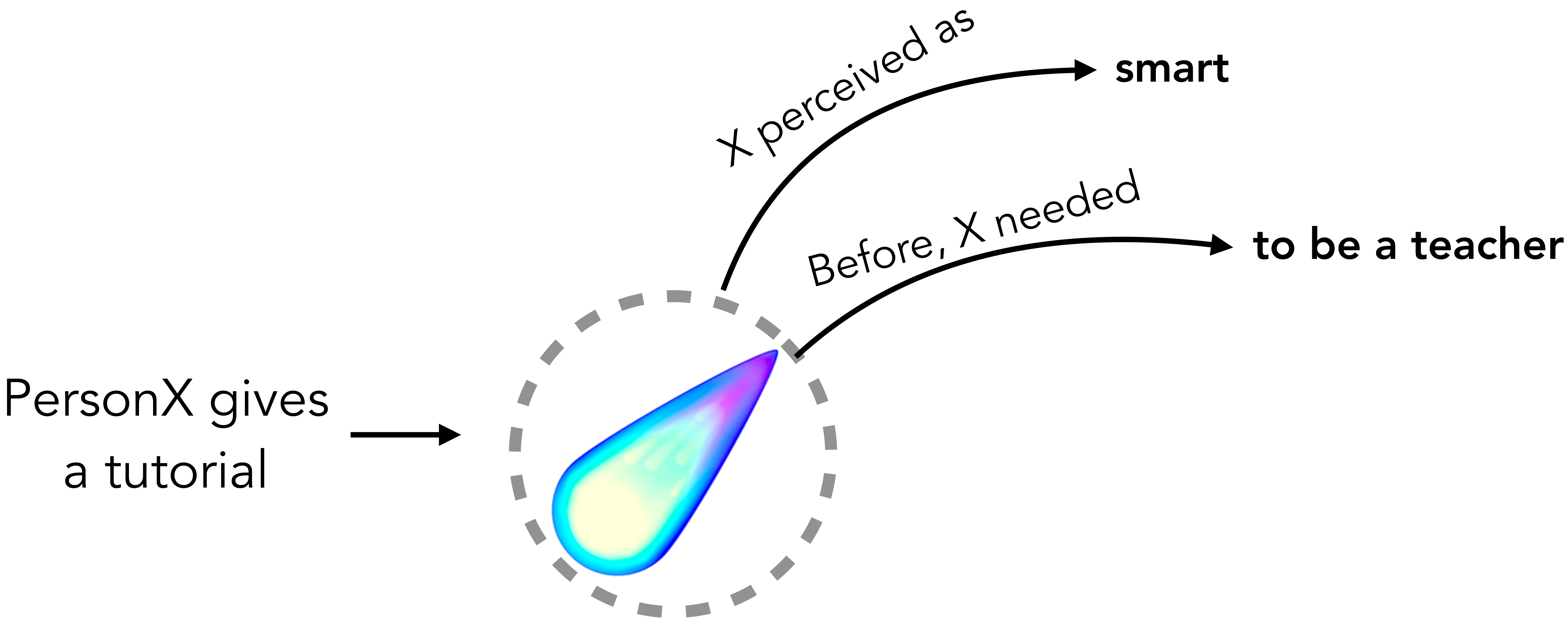
PersonX gives
a tutorial



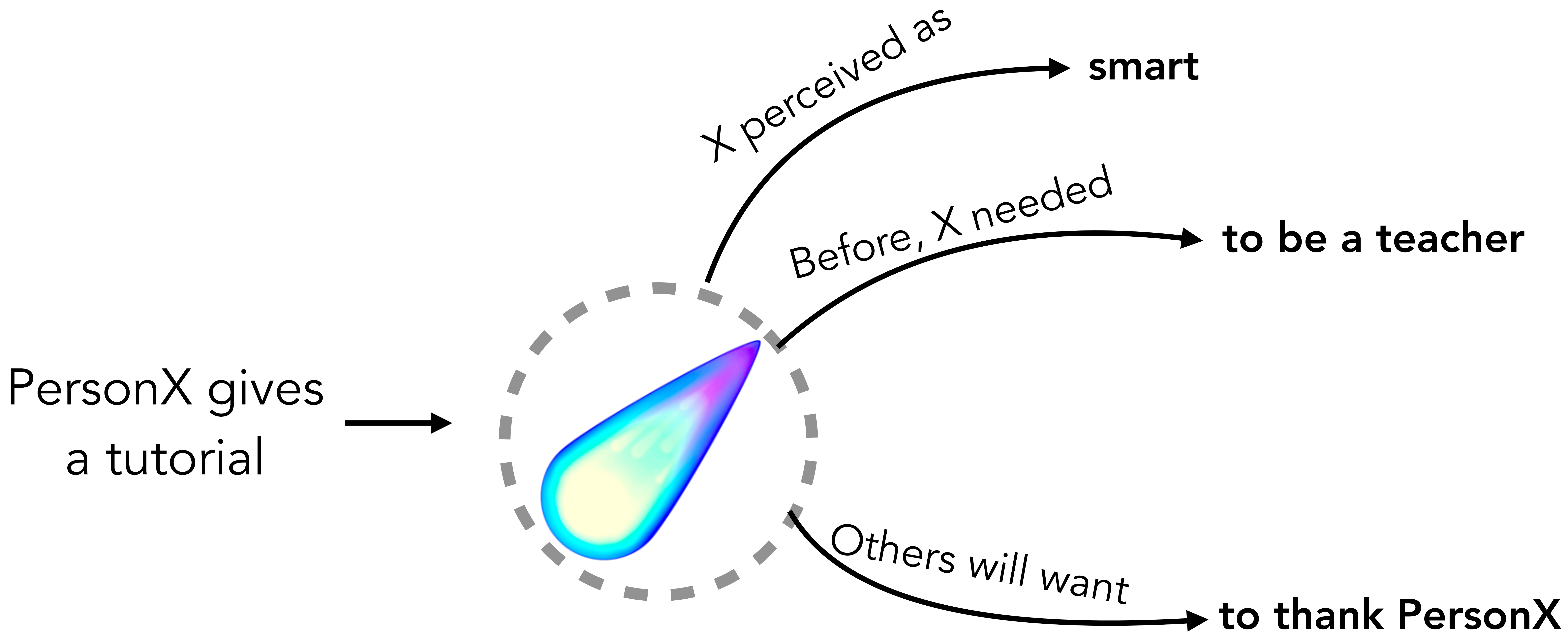
COMET - ATOMIC



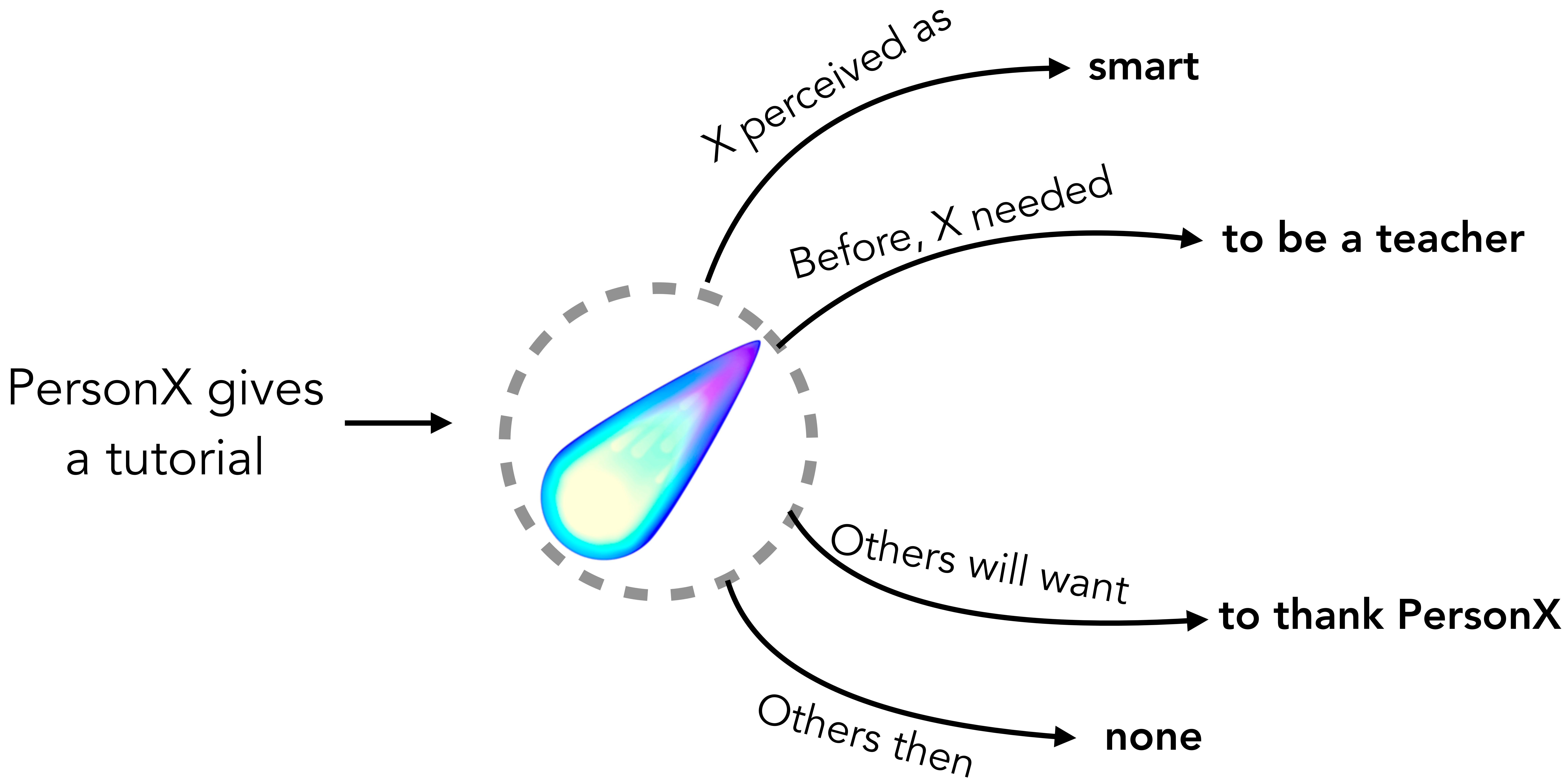
COMET - ATOMIC



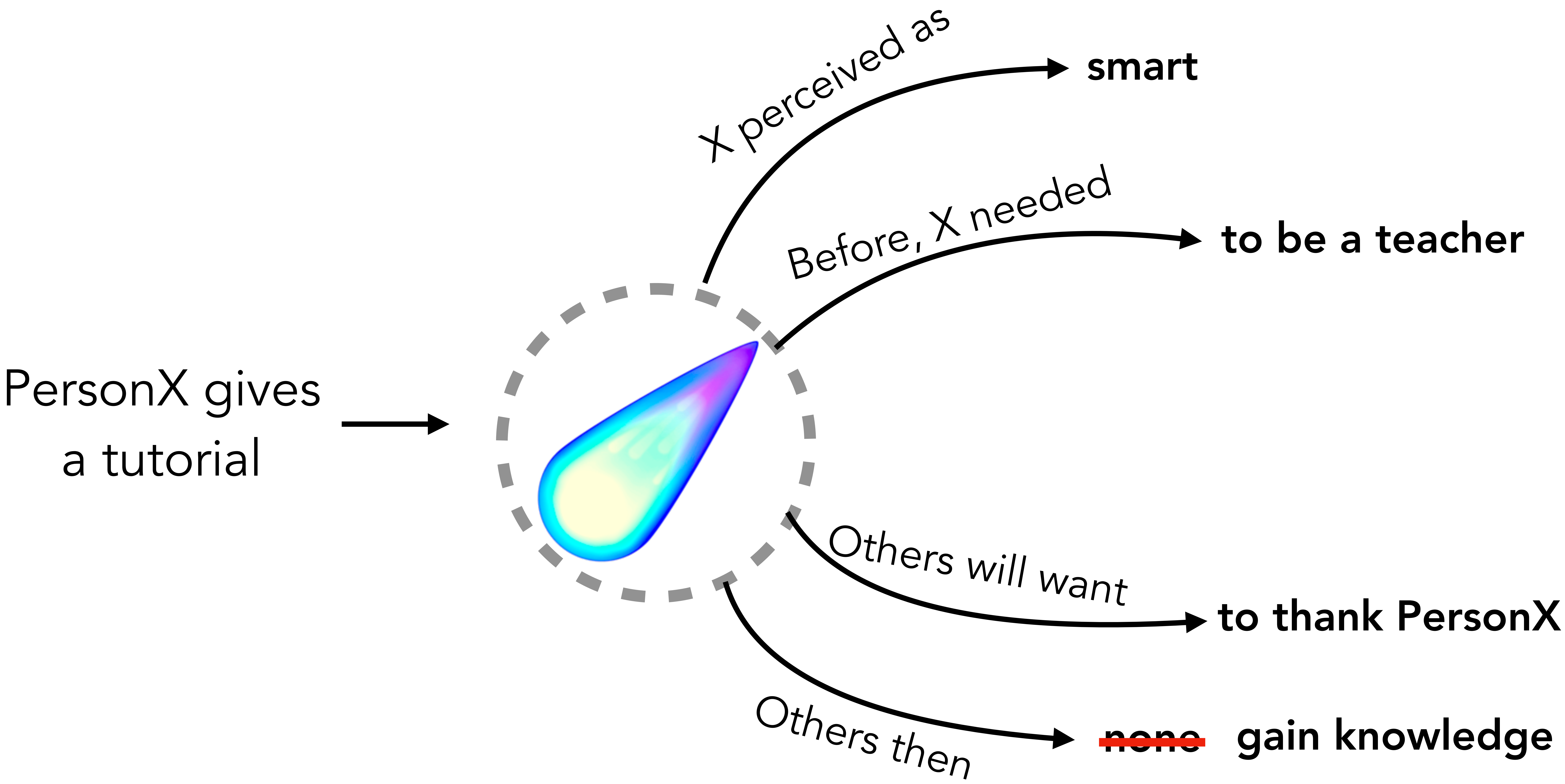
COMET - ATOMIC



COMET - ATOMIC

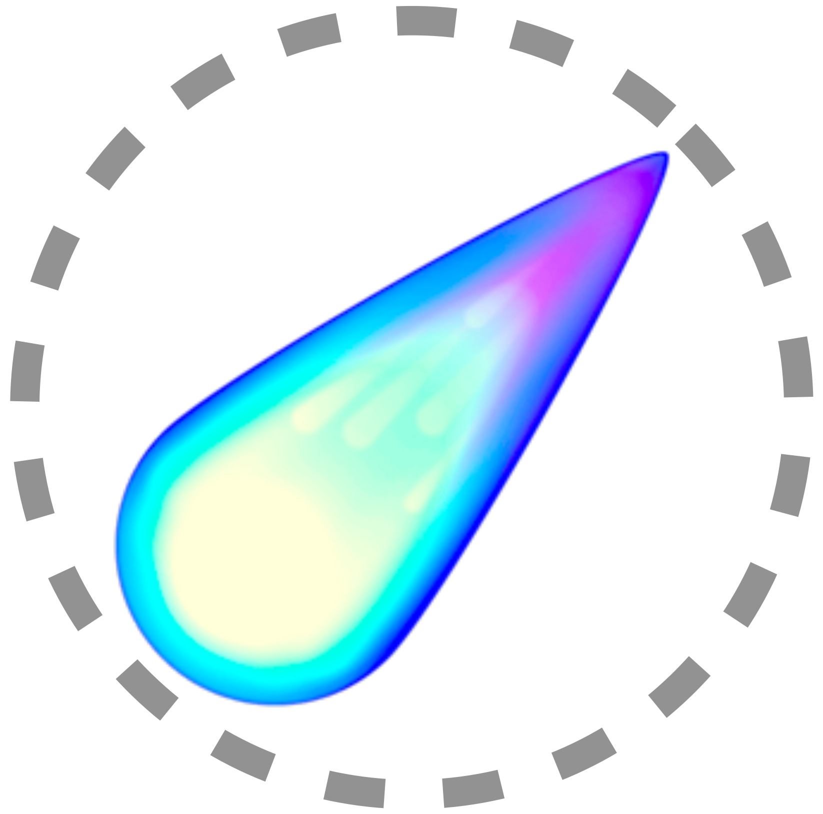


COMET - ATOMIC

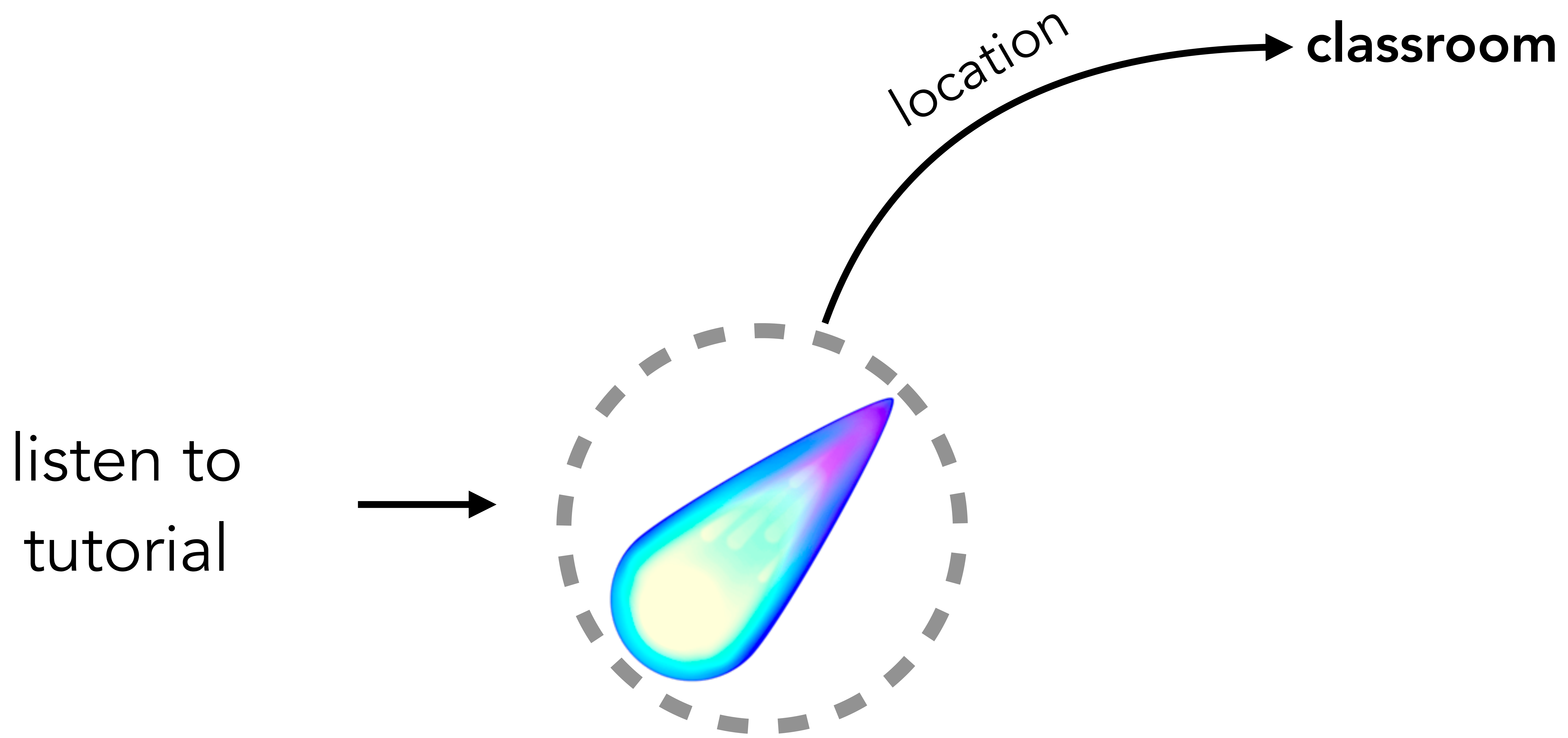


COMET - ConceptNet

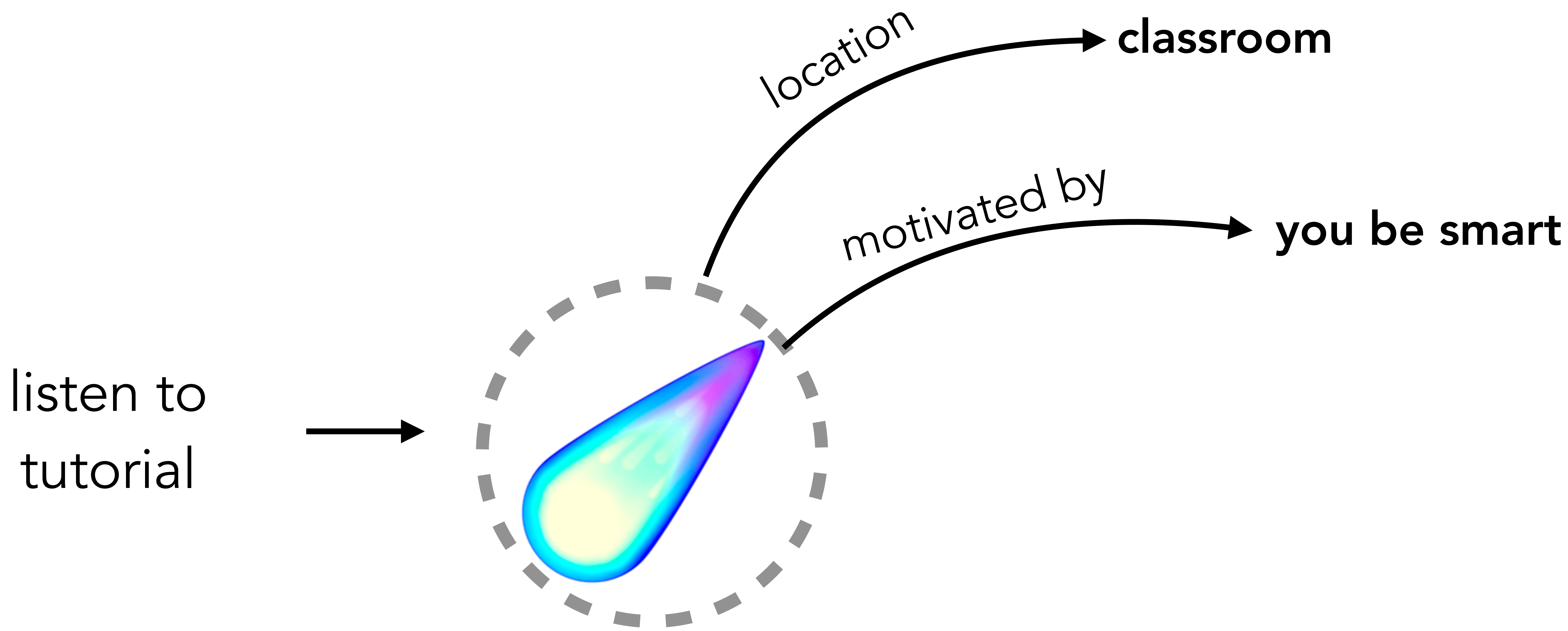
listen to
tutorial



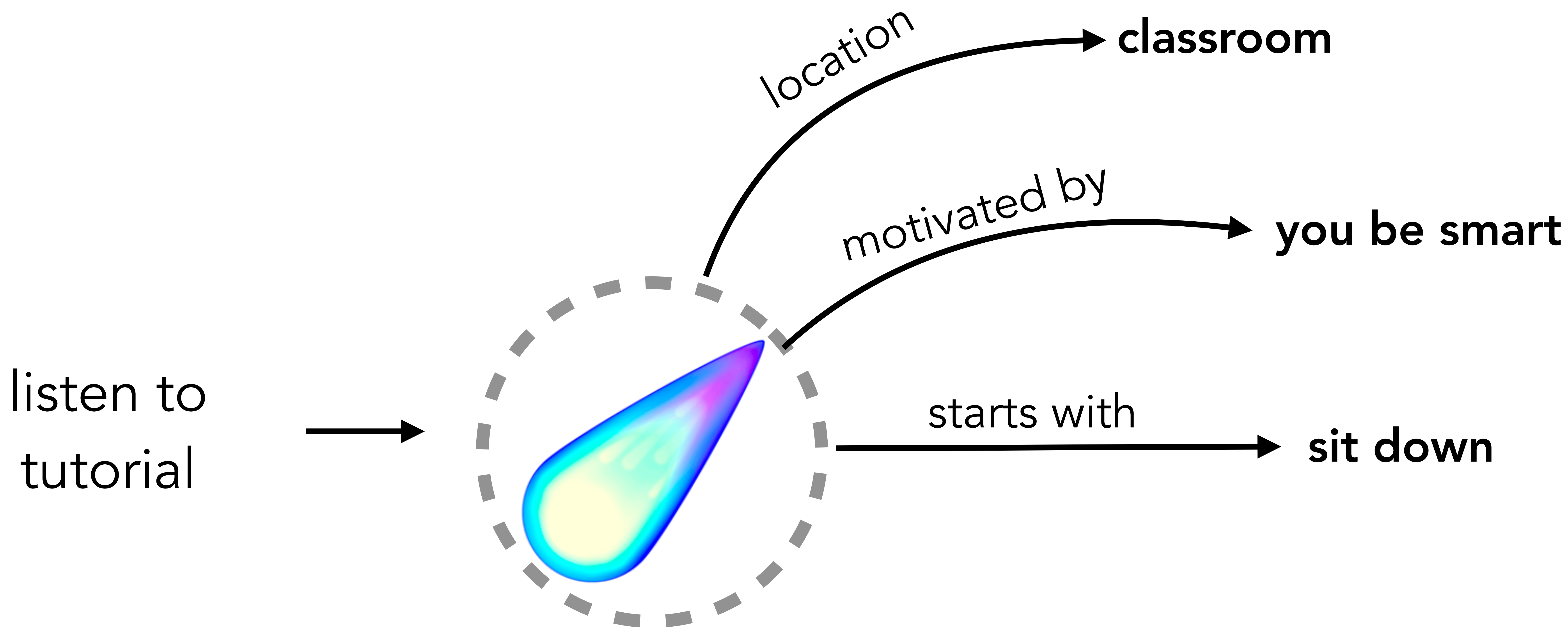
COMET - ConceptNet



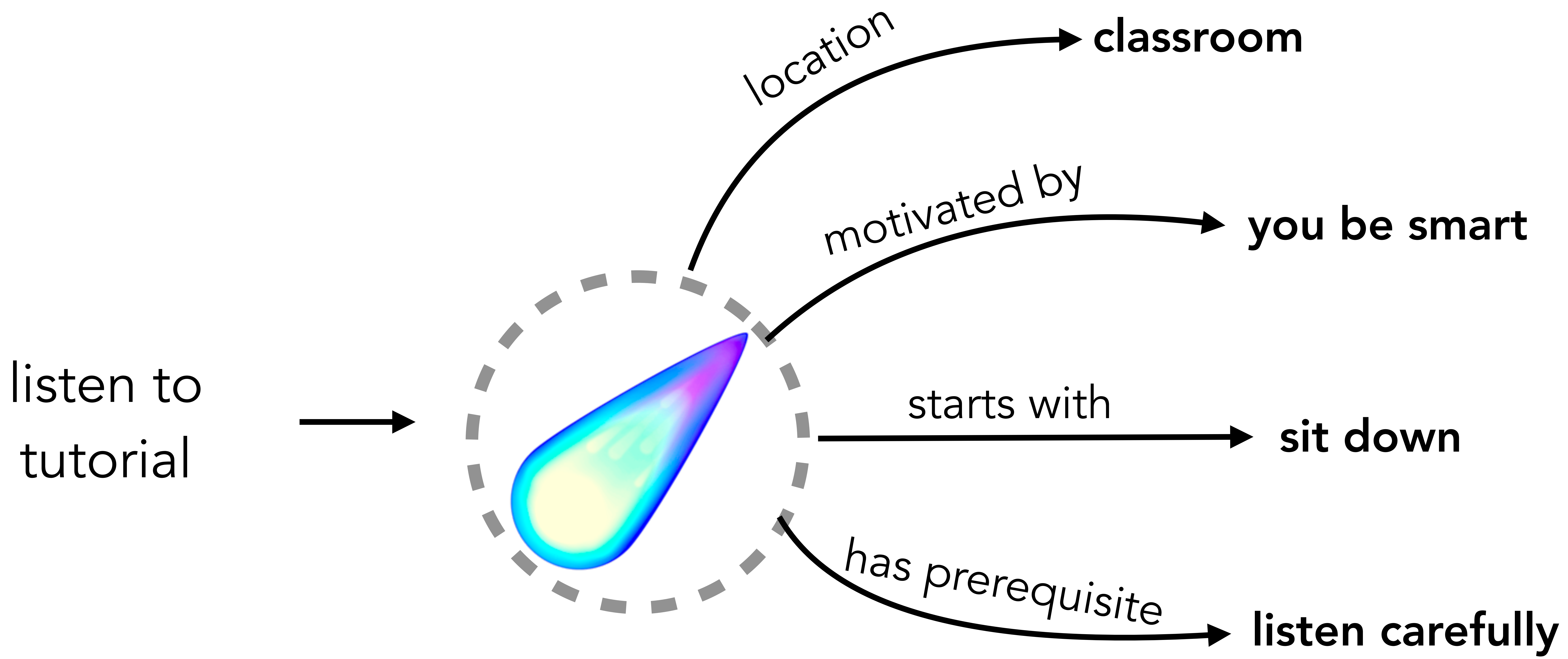
COMET - ConceptNet



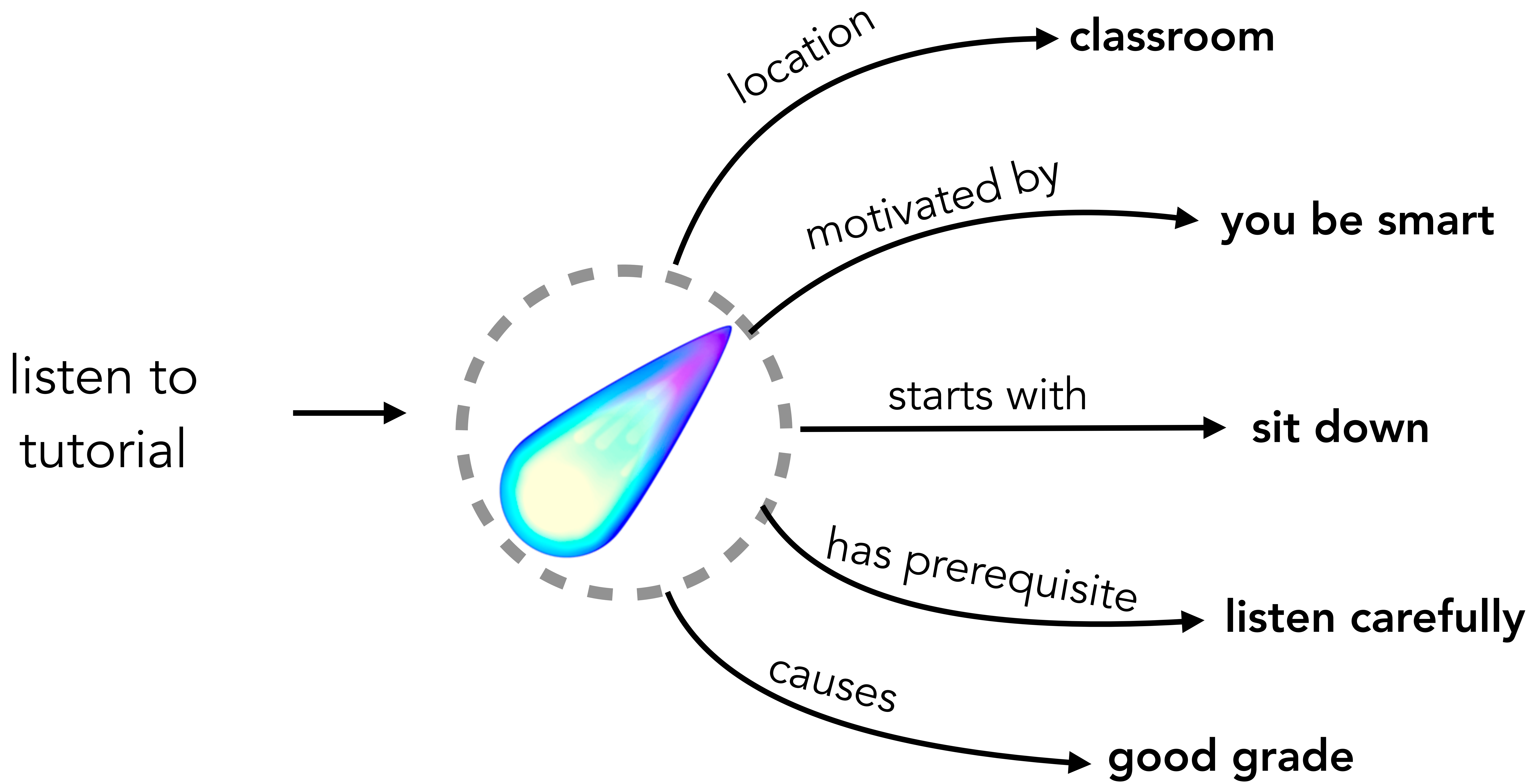
COMET - ConceptNet



COMET - ConceptNet



COMET - ConceptNet



Recap

Recap

Benchmarks:

- [x] Measure progress
- [x] Cover different types of knowledge & reasoning
- [ ] Tradeoff:
  easy to evaluate vs.
  hard to game

Recap

Benchmarks:

- [x] Measure progress
- [x] Cover different types of knowledge & reasoning
- [ ] Tradeoff: easy to evaluate vs. hard to game

Symbolic Knowledge:

- [x] Use for completing missing / unstated knowledge
- [ ] Insufficient coverage
- [ ] How to collect?
- [ ] How to incorporate into models?

Recap

Benchmarks:

- [x] Measure progress
- [x] Cover different types of knowledge & reasoning
- [ ] Tradeoff:
  easy to evaluate vs.
  hard to game

Symbolic Knowledge:

- [x] Use for completing missing / unstated knowledge
- [ ] Insufficient coverage
- [ ] How to collect?
- [ ] How to incorporate into models?

Neural Representations:

- [x] Generalization
- [x] Easy to train/use
- [ ] Inaccurate

Recap

Benchmarks:

- [x] Measure progress
- [x] Cover different types of knowledge & reasoning
- [ ] Tradeoff:
  easy to evaluate vs. hard to game

Symbolic Knowledge:

- [x] Use for completing missing / unstated knowledge
- [ ] Insufficient coverage
- [ ] How to collect?
- [ ] How to incorporate into models?

Neural Representations:

- [x] Generalization
- [x] Easy to train/use
- [ ] Inaccurate

Reasoning Engine:

- [ ] We're only scratching the surface... and we didn't really talk about "reasoning"!

Recap

Benchmarks:
- ☑ Measure progress
- ☑ Cover different types of knowledge & reasoning
- ☐ Tradeoff: easy to evaluate vs. hard to game

Symbolic Knowledge:
- ☑ Use for completing missing / unstated knowledge
- ☐ Insufficient coverage
- ☐ How to collect?
- ☐ How to incorporate into models?

Neural Representations:
- ☑ Generalization
- ☑ Easy to train/use
- ☐ Inaccurate

Reasoning Engine:
- ☐ We're only scratching the surface... and we didn't really talk about "reasoning"!

Thank You!

@VeredShwartz vshwartz@cs.ubc.ca