The Neuro-Symbolic 🔥Continuum🔥 Between Language, Knowledge and Reasoning

Yejin Choi

Paul G. Allen School of Computer Science & Engineering
University of Washington &
Allen Institute for Artificial Intelligence

FAQ these days...

“[NLP/common sense/…] is
almost solved by ChatGPT
and I have an existential crisis”

Hasty generalization fallacy?

BE

The trophy doesn't fit in the brown suitcase because it is too big. What is too big?

The trophy is too big to fit in the brown suitcase.

BE

The trophy doesn't fit in the brown suitcase because it is too small. What is too small?

If the trophy doesn't fit in the brown suitcase because it is too small, then the trophy itself is too small to fit in the suitcase. The suitcase may be larger in size than the trophy, making it impossible for the trophy to fit inside. It is also possible that the trophy is the

— examples from Benjamin Han

I believe that we will continue to be surprised by the new capabilities of deep neural networks 🍒 🍒 🍒 🍒 🍒 🍒 🍒

And yet, these networks will continue making mistakes on adversarial or edge cases 🍋 🍋 🍋 🍋 🍋 🍋 🍋 🍋 🍋 🍋 🍋

The problem is that we simply do not know the depth or the breadth of the adversarial or edge cases (i.e., the amount of hidden lemons 🍋 🍋 🍋 🍋 🍋 🍋 ? ? ? ? ? ? ?

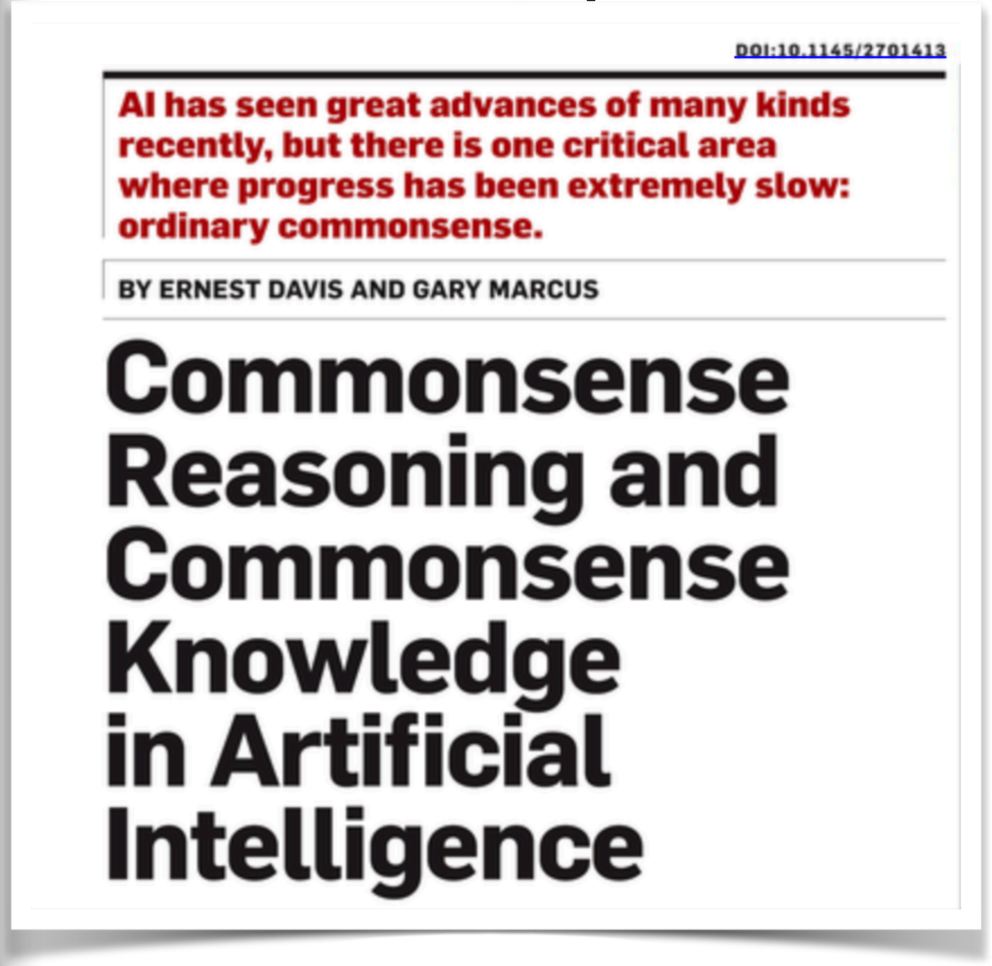
Dark matter is what matters in modern physics

- Only 5% of universe is normal matter. The remaining 95% is dark matter and dark energy.
- Dark matter is completely invisible, yet affects what are visible: the orbits of stars and the trajectory of light

Dark matter of language?

Normal matter: visible text (words, sentences)

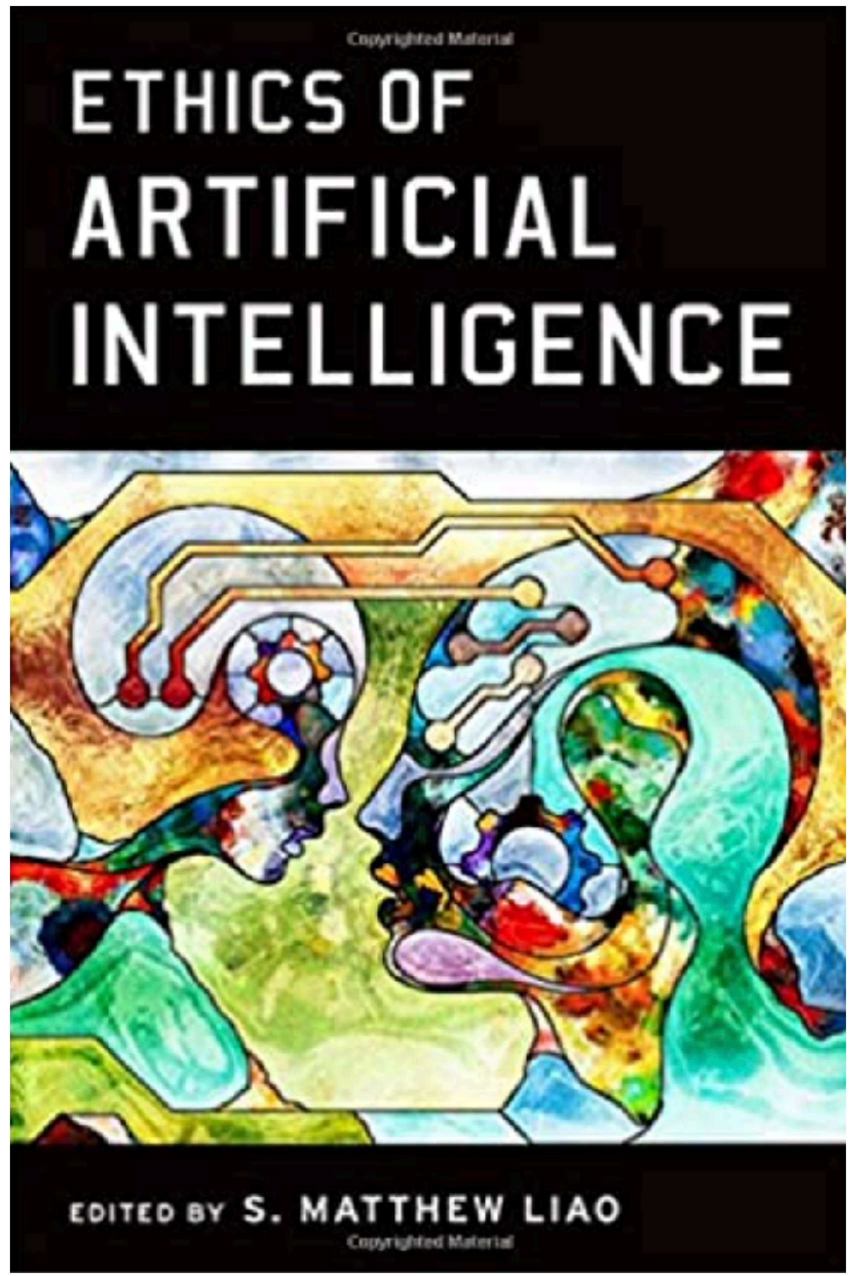
Dark matter: the unspoken rules of how the world works, which influence the way people use and interpret language


DOI:10.1145/2701413

AI has seen great advances of many kinds recently, but there is one critical area where progress has been extremely slow: ordinary commonsense.

BY ERNEST DAVIS AND GARY MARCUS

Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence



Trivial for Humans, Hard for Machines, because…

1. obvious things are never spoken; it's all the implicit stuff
    - how many eyes a horse has?
    - Inhale vs exhale vs murder
2. exceptions are not exceptional, but only expected
    - any rule of thumb can have an endless list of unforeseen exceptions
    - Birds can fly, except…
3. lack of universal truth
    - ambiguous, messy, beyond the realm of conventional logic and math

2082: An ACL Odyssey

https://www.youtube.com/watch?v=lLCEy2mu4Js

Dark Matter
Schrödinger's cat
Wave-particle duality
Spacetime continuum
Mass-energy equivalence

Chapter 1: The ambiguity
Chapter 2: The continuum
Chapter 3: The dark matter

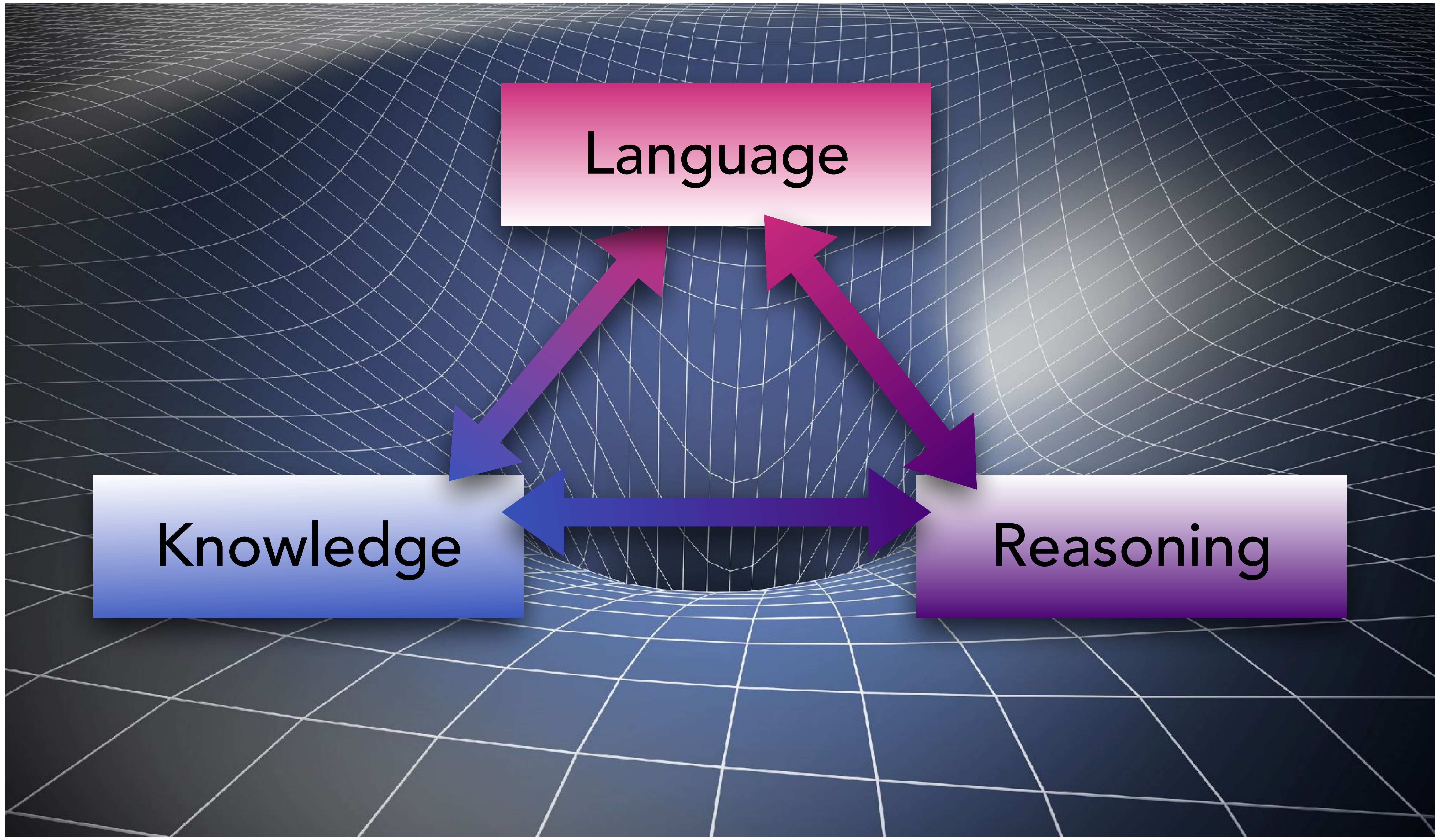
Commonsense
Norms and morals
Ambiguity of language
Language-knowledge-reasoning continuum

keynote at ACL 2022
Charge: to reflect on the past 60 years of NLP research and to project to the future 60 years

The continuum

- Spacetime continuum
- Mass-energy equivalence

The continuum between



Corollary: Formal logic is overrated

Symbolic Knowledge Distillation

From Neural Language Models to Causal Commonsense Models

Peter West

New:
ATOMIC-10x
COMET-distill

Chandra Bhagavatula, Jack Hessel, Jena Hwang, Liwei Jiang, Ronan Le Bras, Ximing Lu, Sean Welleck, Yejin Choi

Maieutic Prompting:

Logically Consistent Reasoning with Recursive Explan

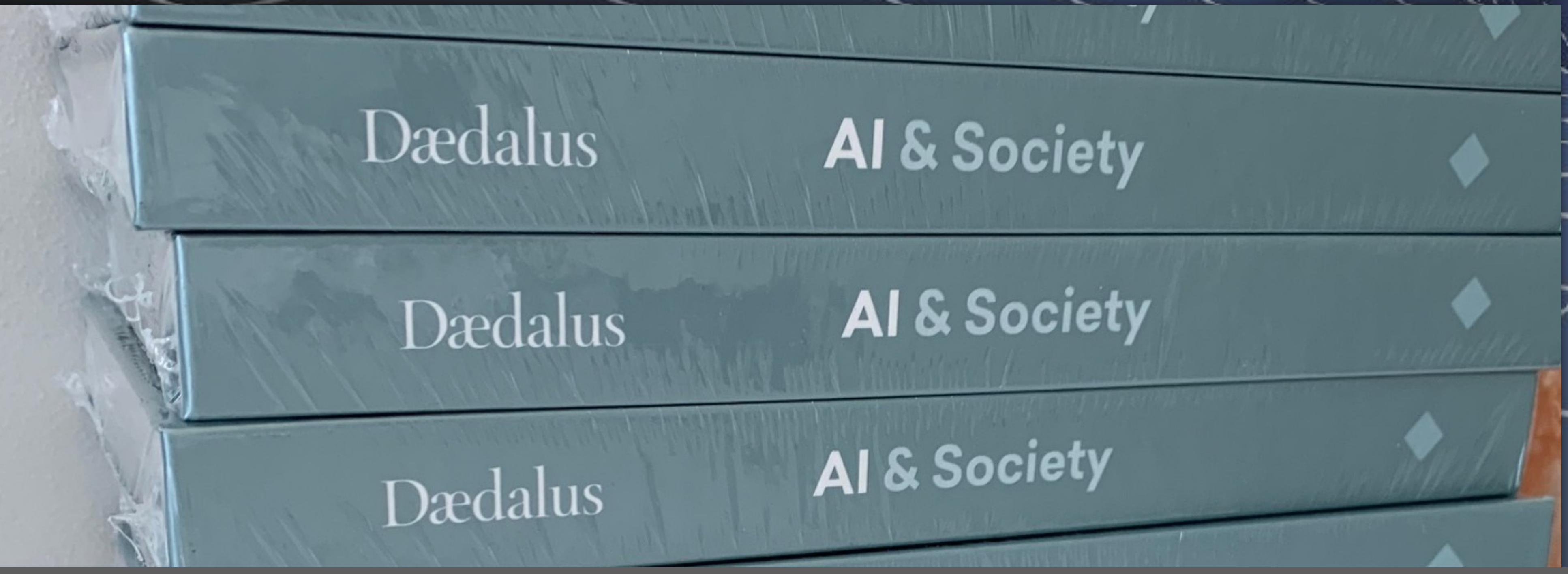
— https://arxiv.org/abs/2205.11822 —

Jaehun Jung, **Lianhui Qin**, **Sean Welleck**, **Faeze Brahman**, **Chandra Bhagavatula**, **Ronan Le Bras**, **Yeji Cho**

Language

Knowledge

Reasoning



"The Curious Case of Commonsense Intelligence"

https://www.amacad.org/publication/curious-case-commonsense-intelligence

PAUL G. ALLEN SCHOOL W

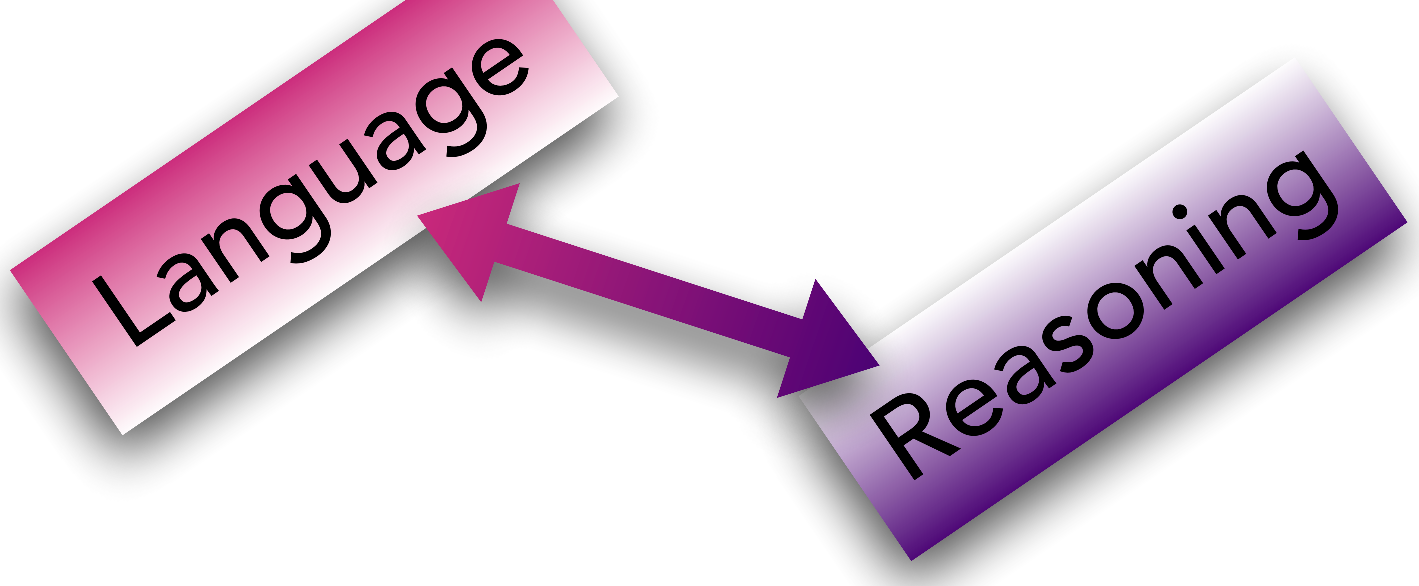
— 🏆 **Best Method Paper Award at NAACL 2022** 🏆 —

NeuroLogic A*esque

Constrained Text Generation with Lookahead Heuristic

Ximing Lu, Sean Welleck, Peter West, Liwei Jiang, Lianhui Qin, Youngjae Yu

Language models are “*sometimes*” amazing!



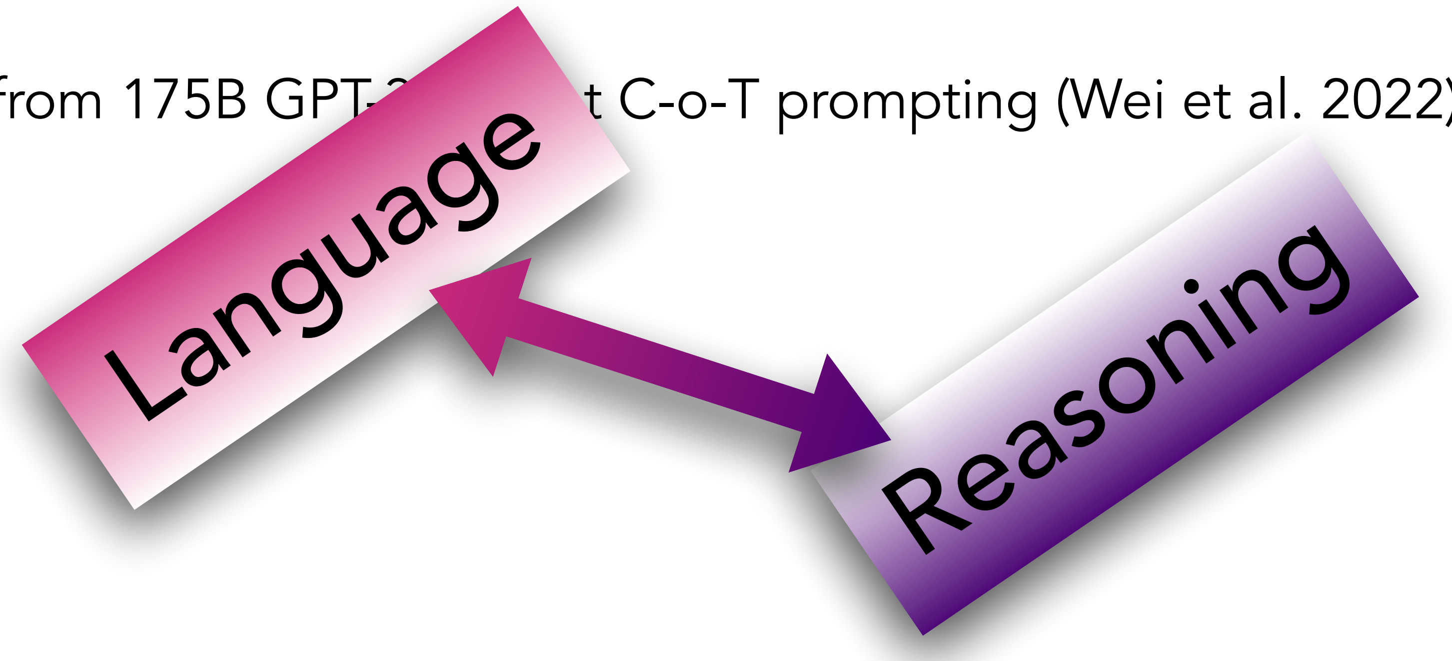
If you travel west far enough from the west coast, you will reach the east coast?

The world is round, so you will reach the east coast. Therefore, the answer is ***True***.

Hmm…

many wonderful ways in which LMs fail to reason…

Actual generation from 175B GPT-3 … t C-o-T prompting (Wei et al. 2022)



Exp. does not logically lead to ans.

Smoke is not the source of fire?

Smoke is a result of fire. Therefore, the statement is ***False***.

???

Invariant to negation

One is a number that comes before zero?

One is … Therefore, the statement is ***True***.

One is <u>not</u> a number that comes before zero?

One is … Therefore, the statement is ***True***.

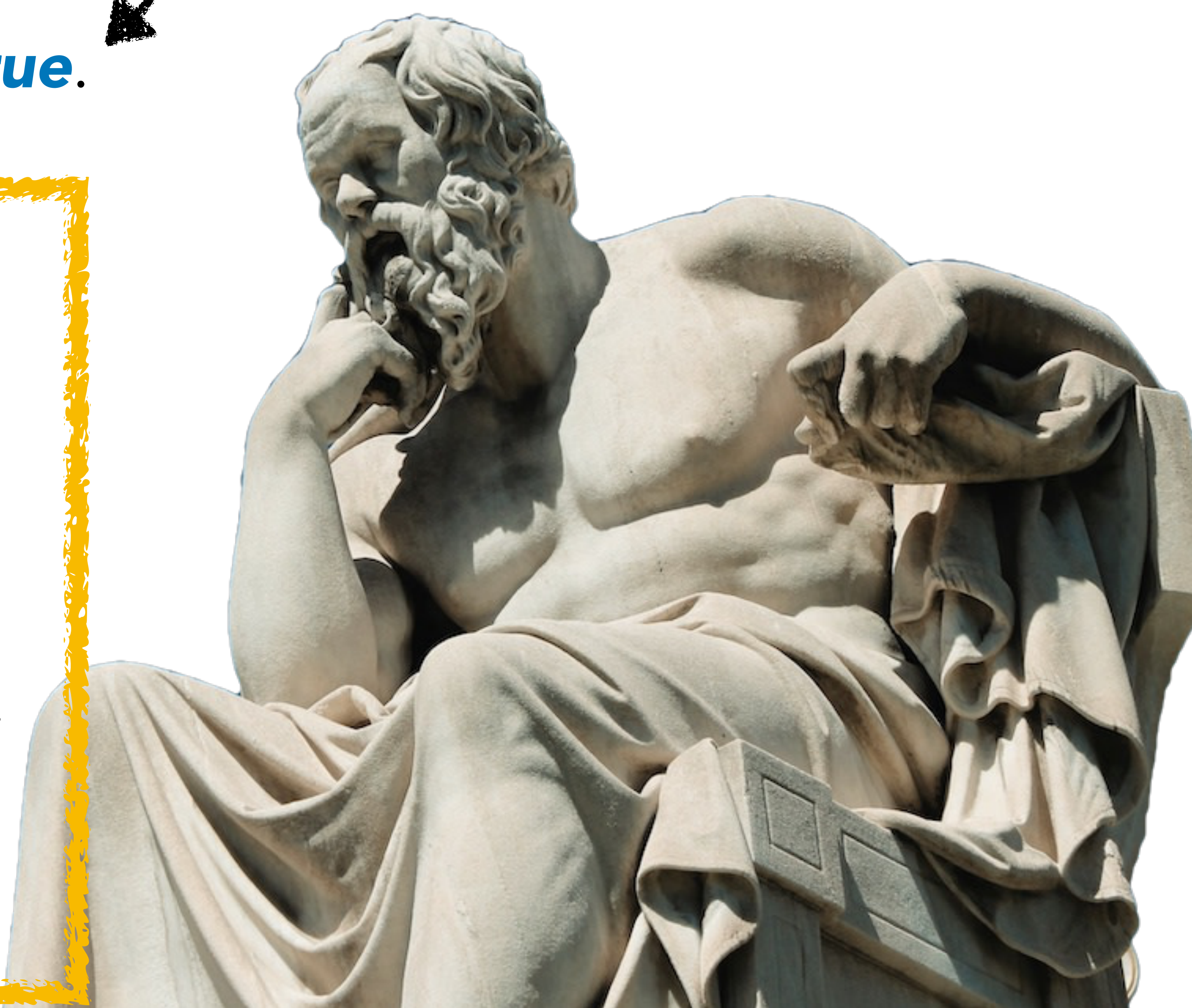
Falsifies its own exp.

Butterflies fly with 3 wings?

Butterflies have 4 wings. Therefore, the statement is False.

Butterflies have 4 wings?

Butterflies have 2 wings on each side of their body. Therefore, the statement is **False**.

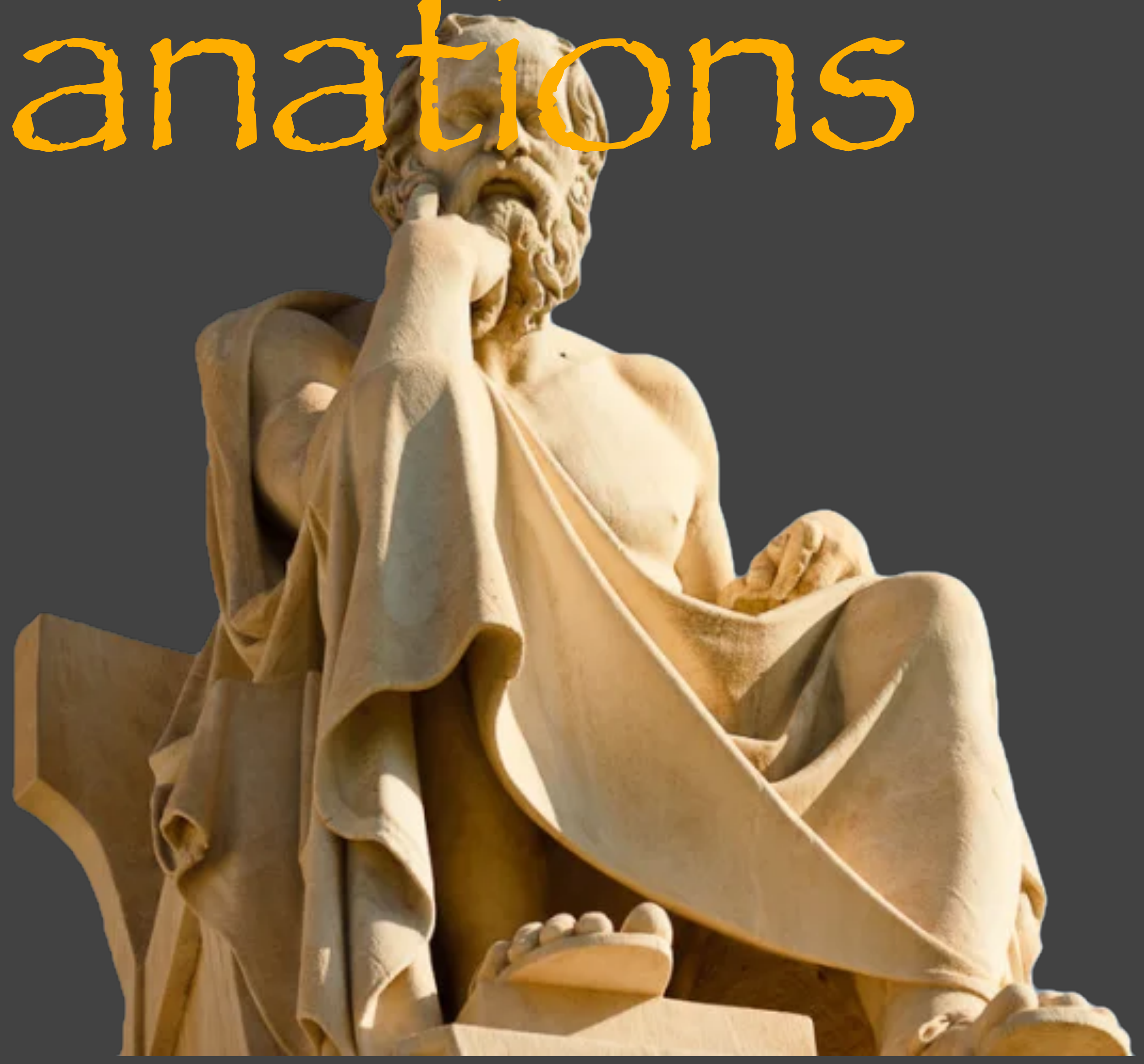


How to make lemonade from GPT-3?
Let's get philosophical

🍋 🍋 🍋

and use Socrates' Maieutic Method

Maeutic Prompting: Logically Consistent Reasoning with Recursive Explanations

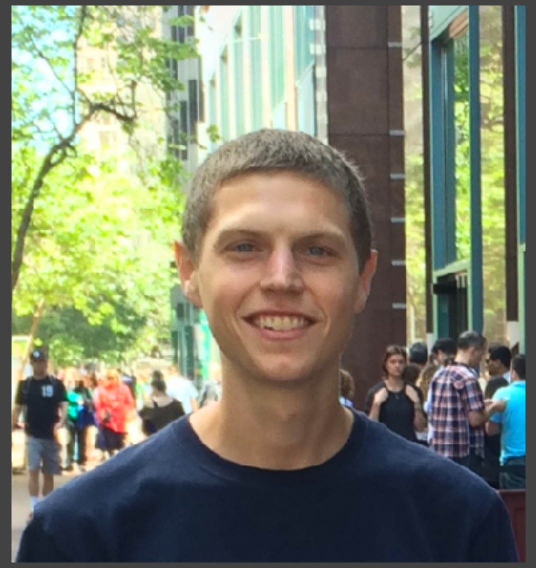


— EMNLP 2022 —

Jaehun Jung

Lianhui Qin

Sean Welleck



Faeze Brahman

Chandra Bhagavatula

Ronan Le Bras

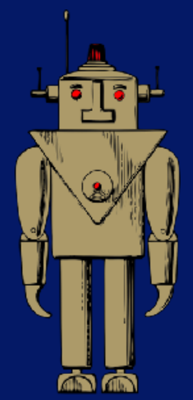
Yejin Choi

Maieutic Tree *G*

Q

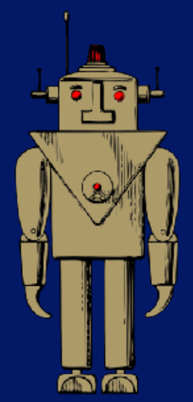
"If you travel west far enough from the west coast, you will reach the east coast?"

Q : If you travel west far enough from the west coast, you will reach the east coast? True, because



E_T : The Earth is round and if you travel in any direction long enough, you will eventually return to where you started.

Q : If you travel west far enough from the west coast, you will reach the east coast? False, because



E_F : You cannot reach the east coast by going west.

Maieutic Tree ***G***

Q

Q : If you travel west far enough from the west coast, you will reach the east coast? True, because

E_T : `The Earth is round and if you travel in any direction long enough, you will eventually return to where you started.`

Q : If you travel west far enough from the west coast, you will reach the east coast? False, because

E_F : `You cannot reach the east coast by going west.`

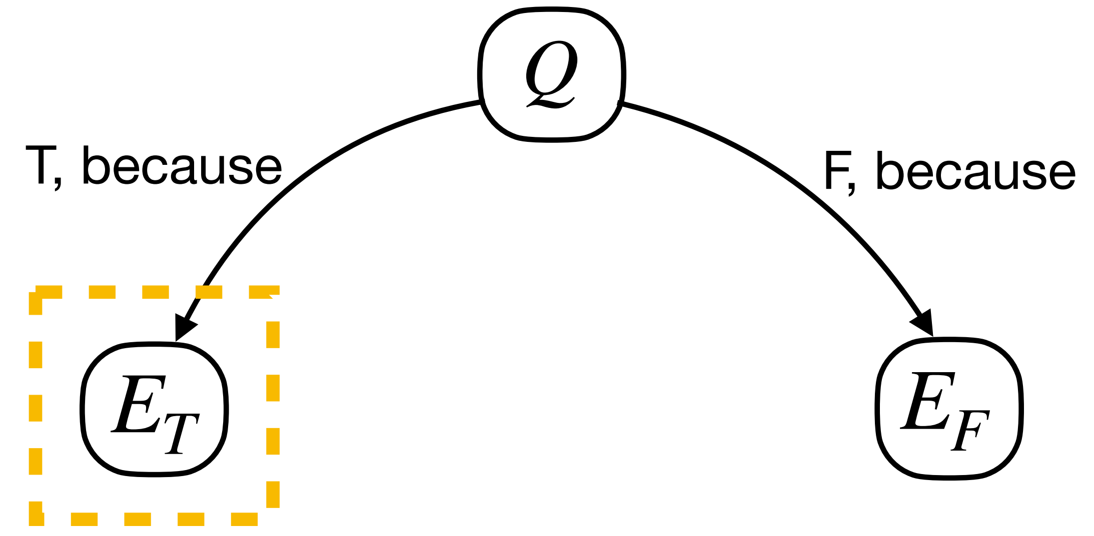
E_T : The Earth is round and if you travel in any direction long enough, you will eventually return to where you started?

`True.`

$\neg E_T$: The Earth is round and if you travel in any direction long enough, you will <u>not</u> return to where you started?

`False.`

Maieutic Tree *G*



is logically integral to E_T !

E_T : The Earth is round and if you travel in any direction long enough, you will eventually return to where you started?

True.

$\neg E_T$: The Earth is round and if you travel in any direction long enough, you will <u>not</u> return to where you started?

False.

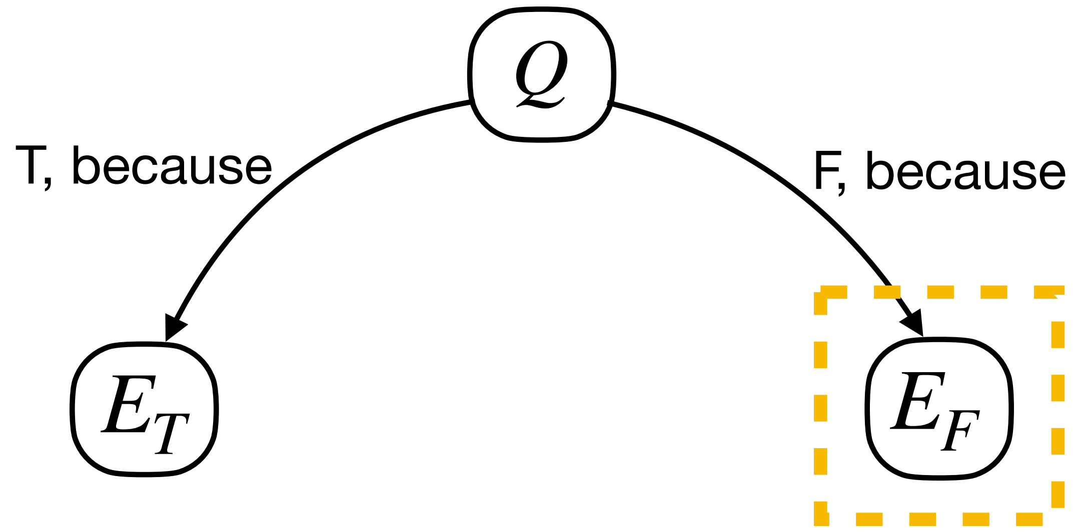
E_F : You cannot reach the east coast by going west?

True.

$\neg E_F$: You <u>can</u> reach the east coast by going west?

True.

Maieutic Tree *G*

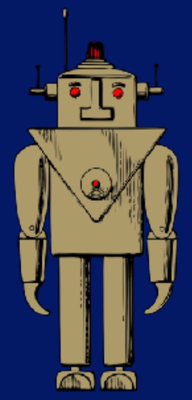


is not logically integral to E_F ;
Generate further with
E_F as a question!

E_F : You cannot reach the east coast by going west?

True, because

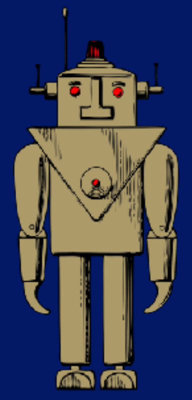
E_{FT} : You can reach the east coast by going west by traveling around the world.



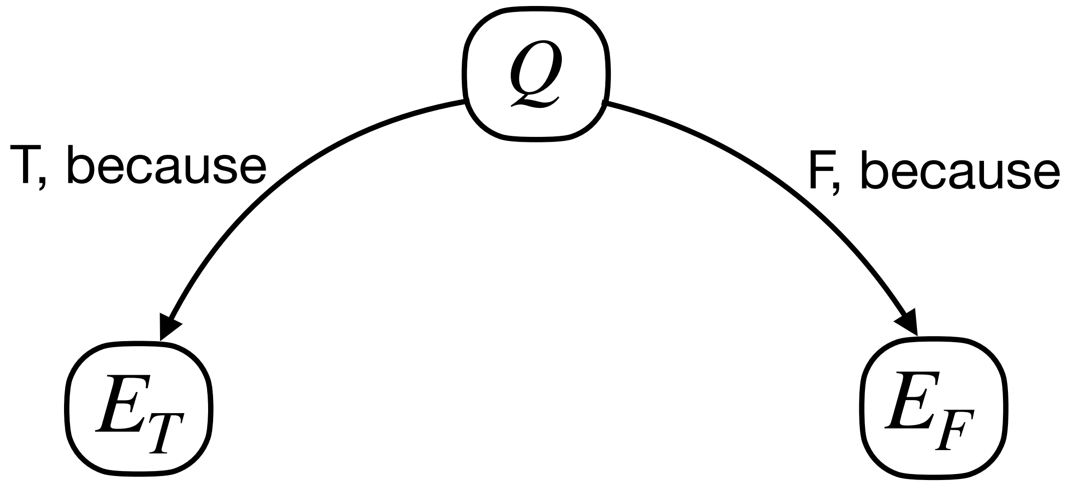
E_F : You cannot reach the east coast by going west?

False, because

E_{FF} : If you travel in a specific straight line, you will eventually reach the other side.



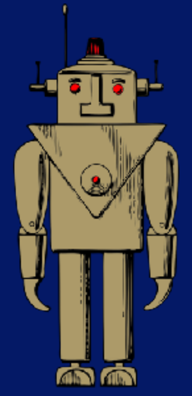
Maieutic Tree *G*



E_F : You cannot reach the east coast by going west?

True, because

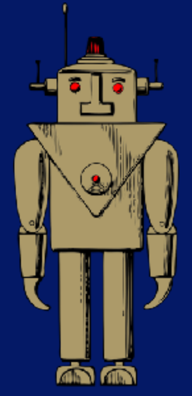
E_{FT} : You can reach the east coast by going west by traveling around the world.



E_F : You cannot reach the east coast by going west?

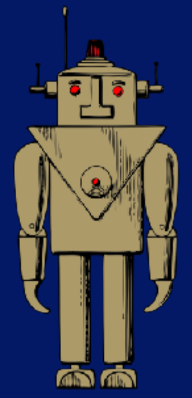
False, because

E_{FF} : If you travel in a specific straight line, you will eventually reach the other side.



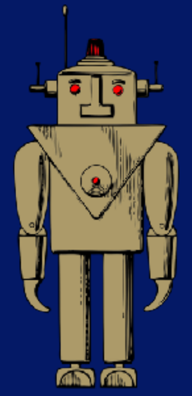
E_{FT} : You can reach the east coast by going west by traveling around the world?

True.

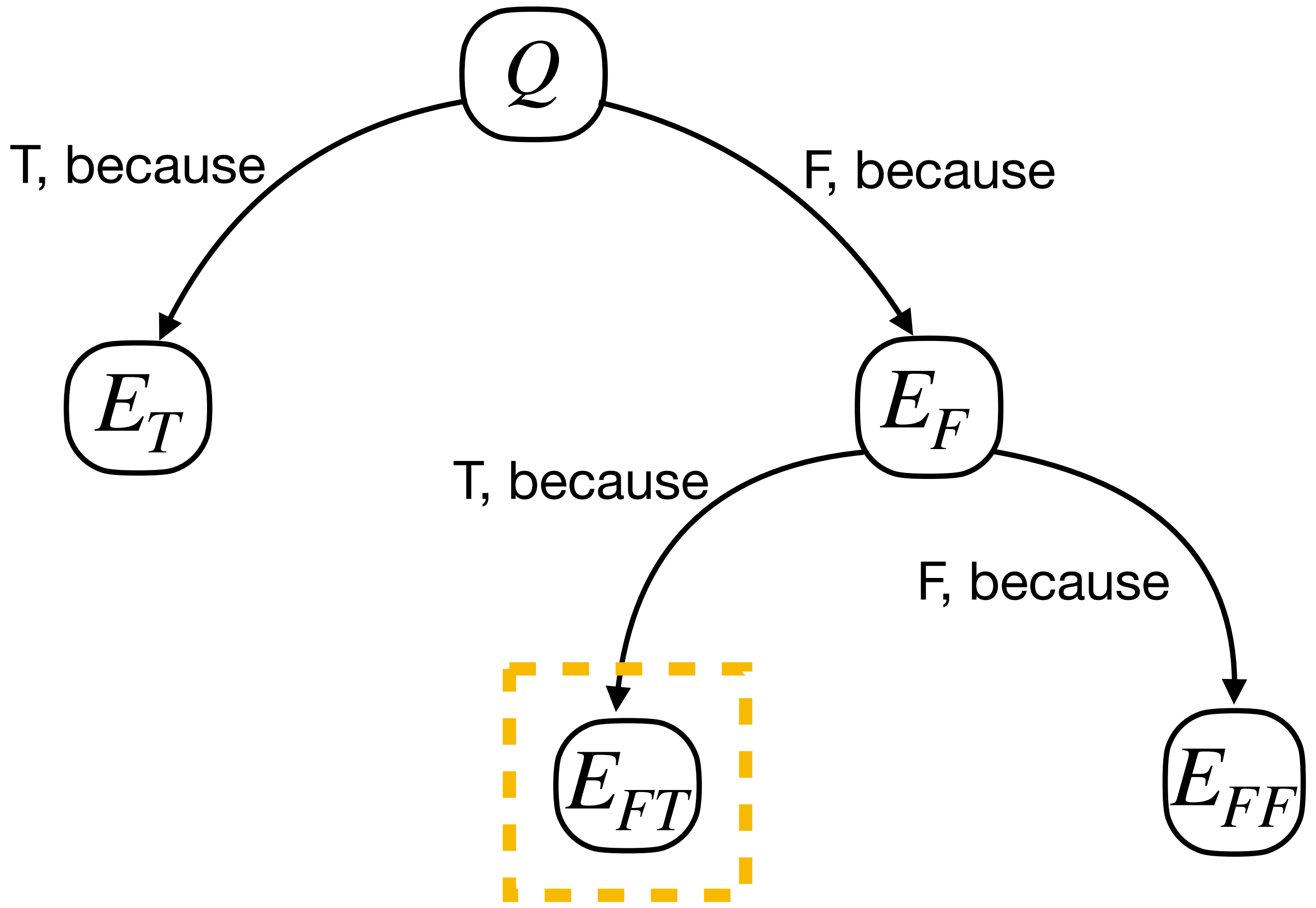


$\neg E_{FT}$: You cannot reach the east coast by going west by traveling around the world?

False.

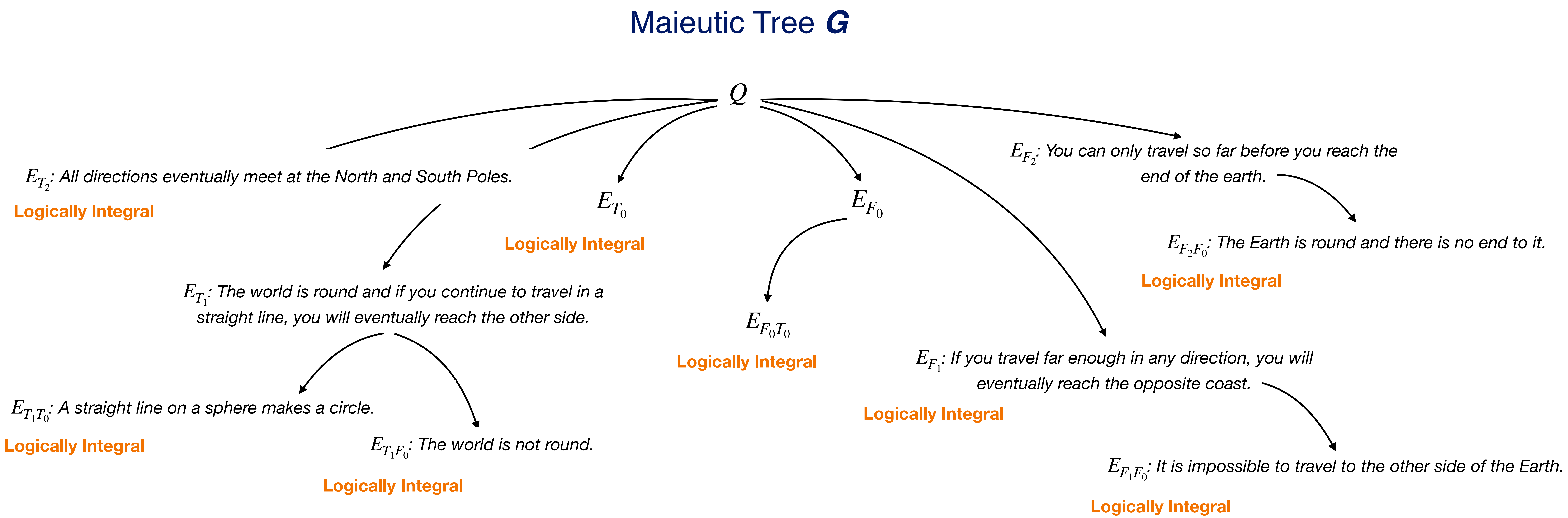


Maieutic Tree *G*



is logically integral to E_{FT} !

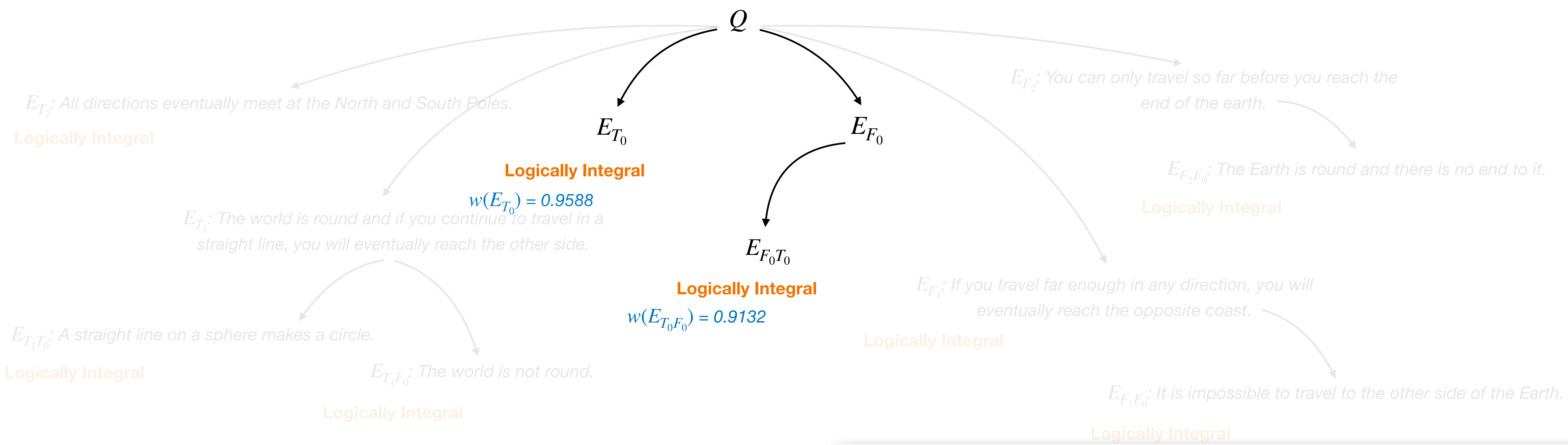
Maieutic Prompting - Abductive Generation



Actual maieutic tree for Q: "If you travel west far enough from the west coast, you will reach the east coast?"
Generated using 175B GPT-3 Davinci (Max Depth: 2 / Width: 3 per truth value)

Maieutic Prompting - Inference

Maieutic Tree ***G***



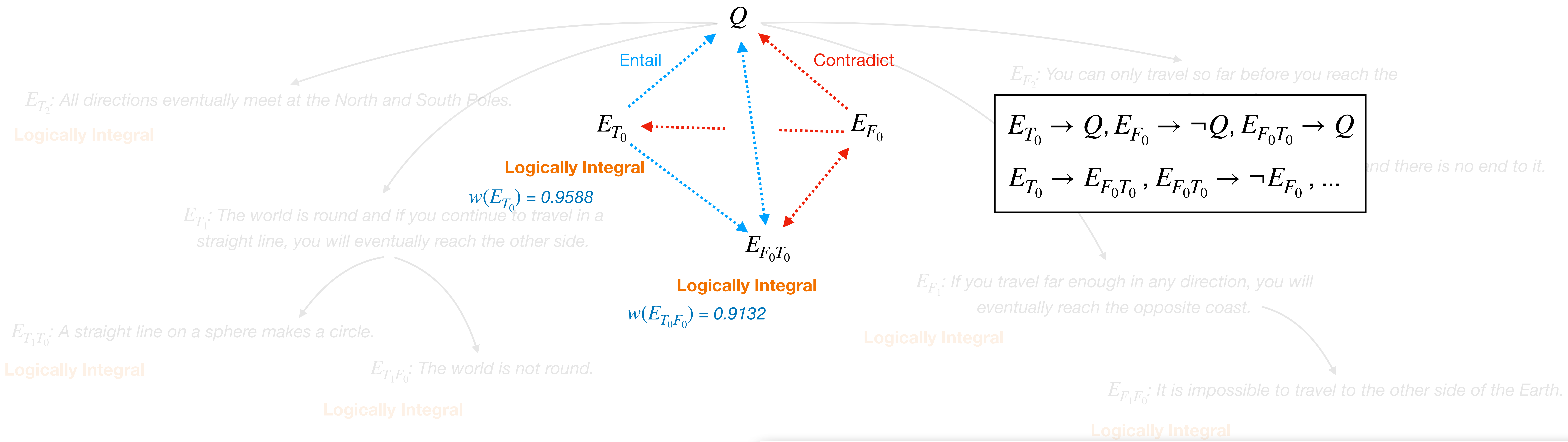
Unary logical constraints with *Belief:*
How strongly does GPT-3 believe in E (or not E)?

For each leaf node E in $\mathcal{G}$, we define strength of LM's *belief* on either E or $\neg E$ as following:

$$\textit{Belief} \begin{cases} w(E) := \dfrac{p_{LM}(\textit{True}|E,\ldots)}{p_{LM}(\textit{True}|E,\ldots) + p_{LM}(\textit{True}|\neg E,\ldots)} & \text{if } E \text{ is } \textit{True} \\ w(\neg E) := \dfrac{p_{LM}(\textit{True}|\neg E,\ldots)}{p_{LM}(\textit{True}|\neg E,\ldots) + p_{LM}(\textit{True}|E,\ldots)} & \text{if } E \text{ is } \textit{False} \end{cases}$$

Maieutic Prompting - Inference

Maieutic Tree *G*

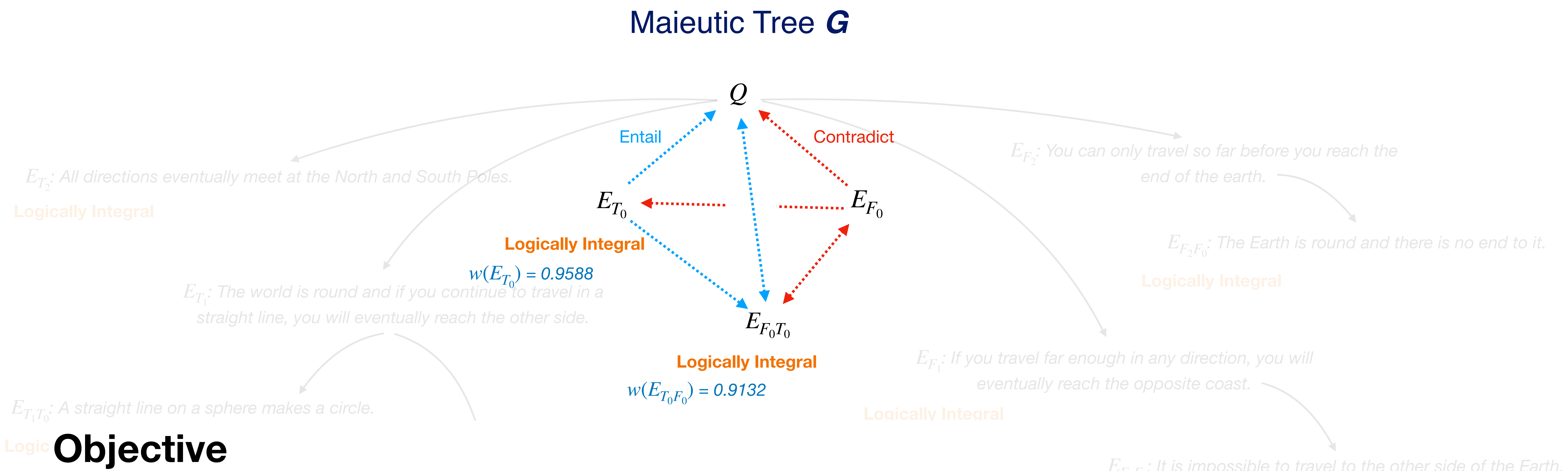


Binary logical constraints with *Consistency:*
Do the two Es support or contradict each other?

For all pairs of nodes (E_1, E_2) in $\mathcal{G}$, we define the logical consistency between the propositions using NLI labels, with weights fixed to 1:

$$\textit{Consistency} \quad \begin{cases} w(E_1 \rightarrow E_2) = 1 & \text{if } \texttt{Entail}(E_1, E_2) \\ w(E_1 \rightarrow \neg E_2) = 1 & \text{if } \texttt{Contradict}(E_1, E_2) \end{cases}$$

Maieutic Prompting - Inference



Objective

$$\underset{\forall i\, E_i, Q \in \{T,F\}}{\arg\max} \sum_{c \in \mathcal{C}_{blf} \cup \mathcal{C}_{con}} w(c) \cdot \mathbb{1}_{\{c=\text{True}\}}$$

Weighted Max-SAT solver (Morgado, 2001) **can deterministically find the best assignment!**

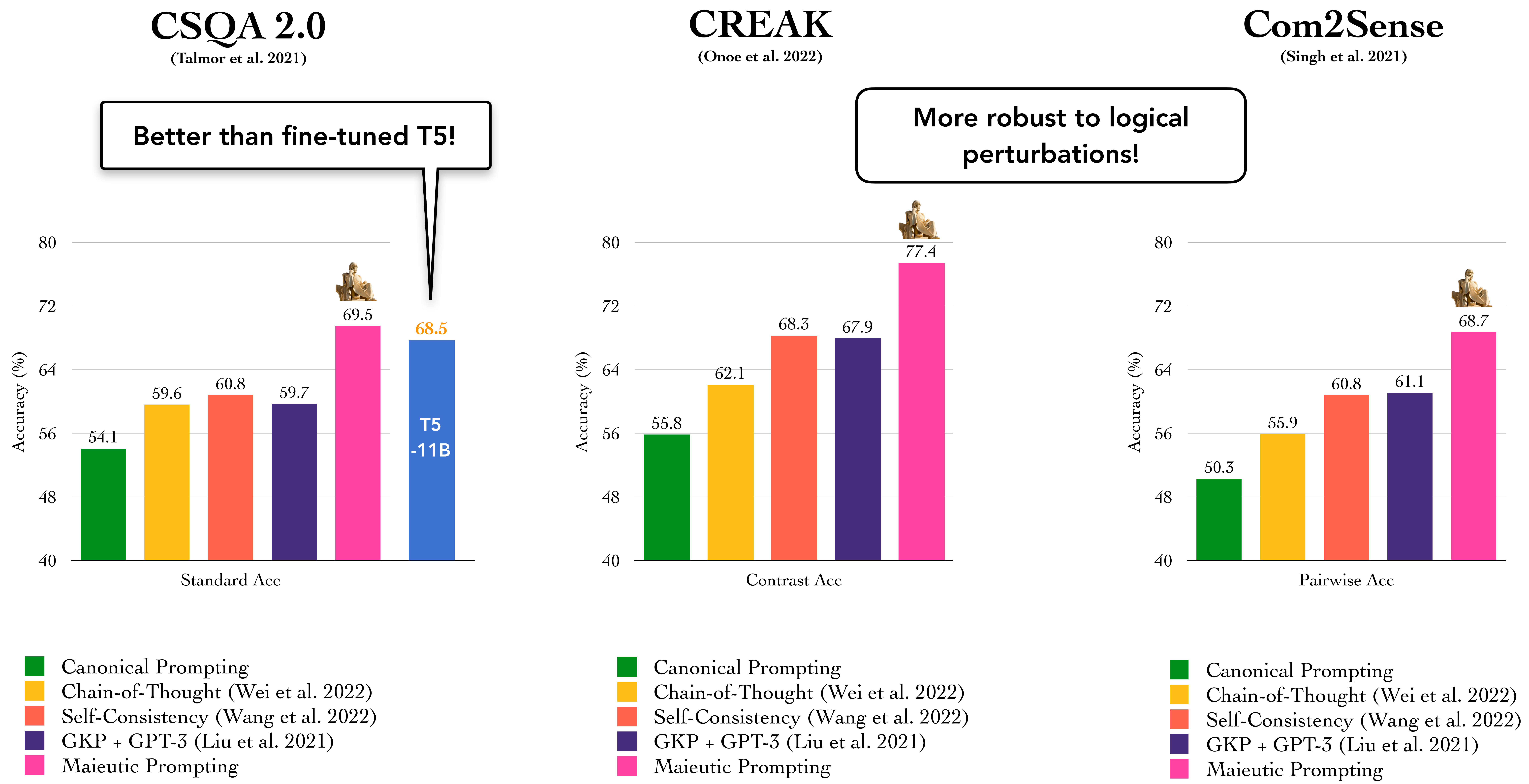
Max-SAT output:

Q: ***True***

E_{T_0}: ***True***

$E_{F_0T_0}$: ***True***

E_{F_0}: ***False***


CSQA 2.0
(Talmor et al. 2021)
Better than fine-tuned T5!
80
72
64
56
48
40
Accuracy (%)
54.1
59.6
60.8
59.7
69.5
68.5
T5
-11B
Standard Acc
CREAK
(Onoe et al. 2022)
More robust to logical perturbations!
Accuracy (%)
55.8
62.1
68.3
67.9
77.4
Contrast Acc
Com2Sense
(Singh et al. 2021)
Accuracy (%)
50.3
55.9
60.8
61.1
68.7
Pairwise Acc
Canonical Prompting
Chain-of-Thought (Wei et al. 2022)
Self-Consistency (Wang et al. 2022)
GKP + GPT-3 (Liu et al. 2021)
Maieutic Prompting

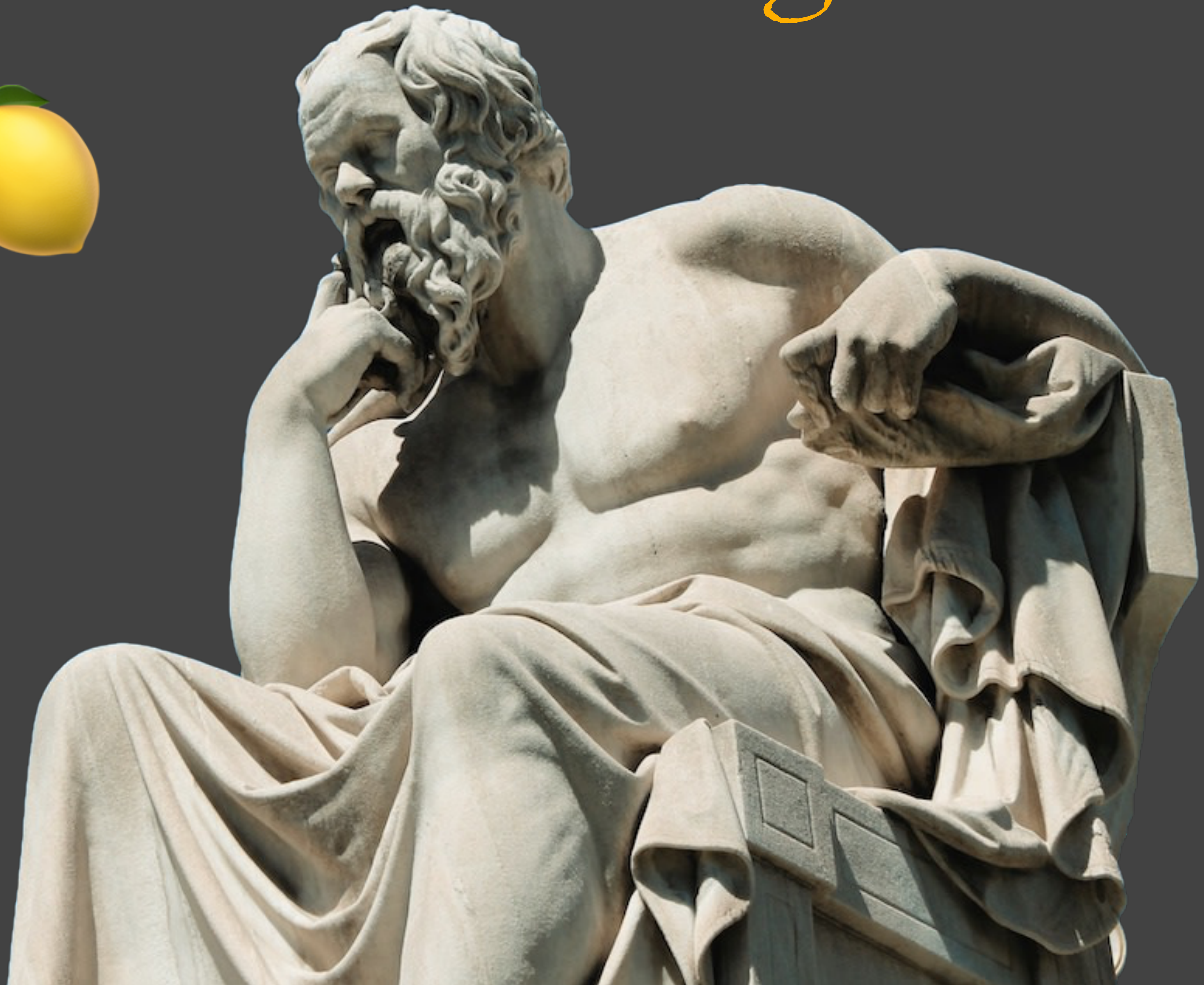
Take away:

Socrates' Maieutic Method

not only enhances flawed human reasoning,

a computational interpretation of it can

dramatically enhance flawed GPT-3's reasoning



PAUL G. ALLEN SCHOOL

W

— 🏆 Best Method Paper Award at NAACL 2022 🏆 —

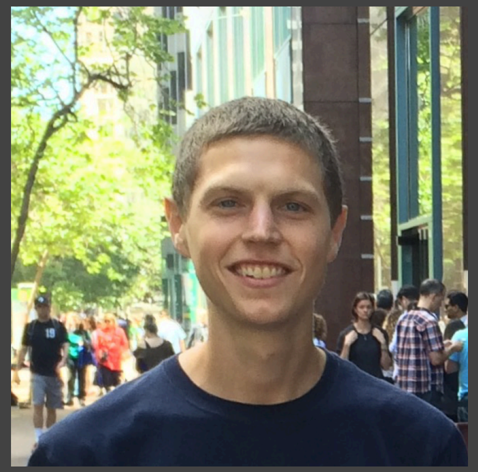
AI2

NeuroLogic A*esque

Constrained Text Generation with Lookahead Heuristic

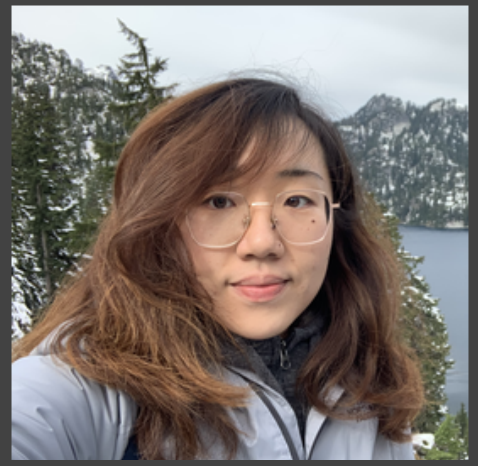
Ximing Lu

Sean Welleck

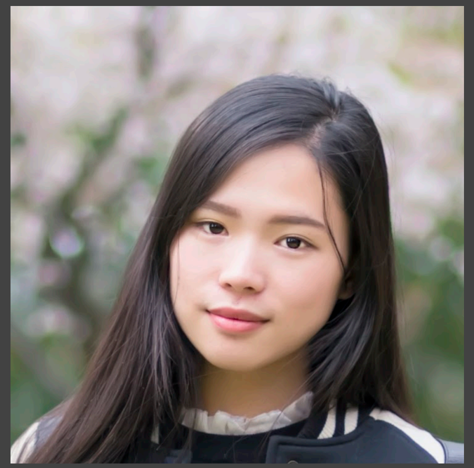


Peter West

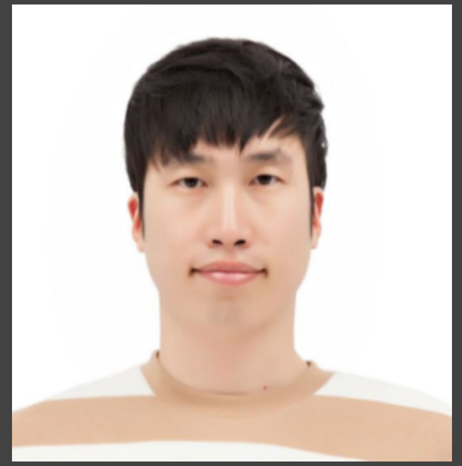
Liwei Jiang



Lianhui Qin

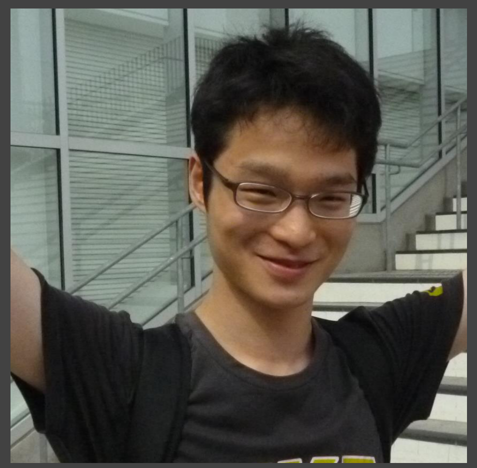


Youngjae Yu



Daniel Khashabi

Jungo Kasai



Ronan Le Bras

Rowan Zellers

Noah Smith

Yejin Choi

NeuroLogic Decoding

(Un)supervised Neural Text Generation with **Predicate Logic Constraints**

—NAACL 2021—

Ximing Lu

Peter West

Rowan Zellers

Ronan LeBras

Chandra Bhagavatula

Yejin Choi

NeuroLogic Decoding

(Un)supervised Neural Text Generation with **Predicate Logic Constraints**

—NAACL 2021—

PAUL G. ALLEN SCHOOL W

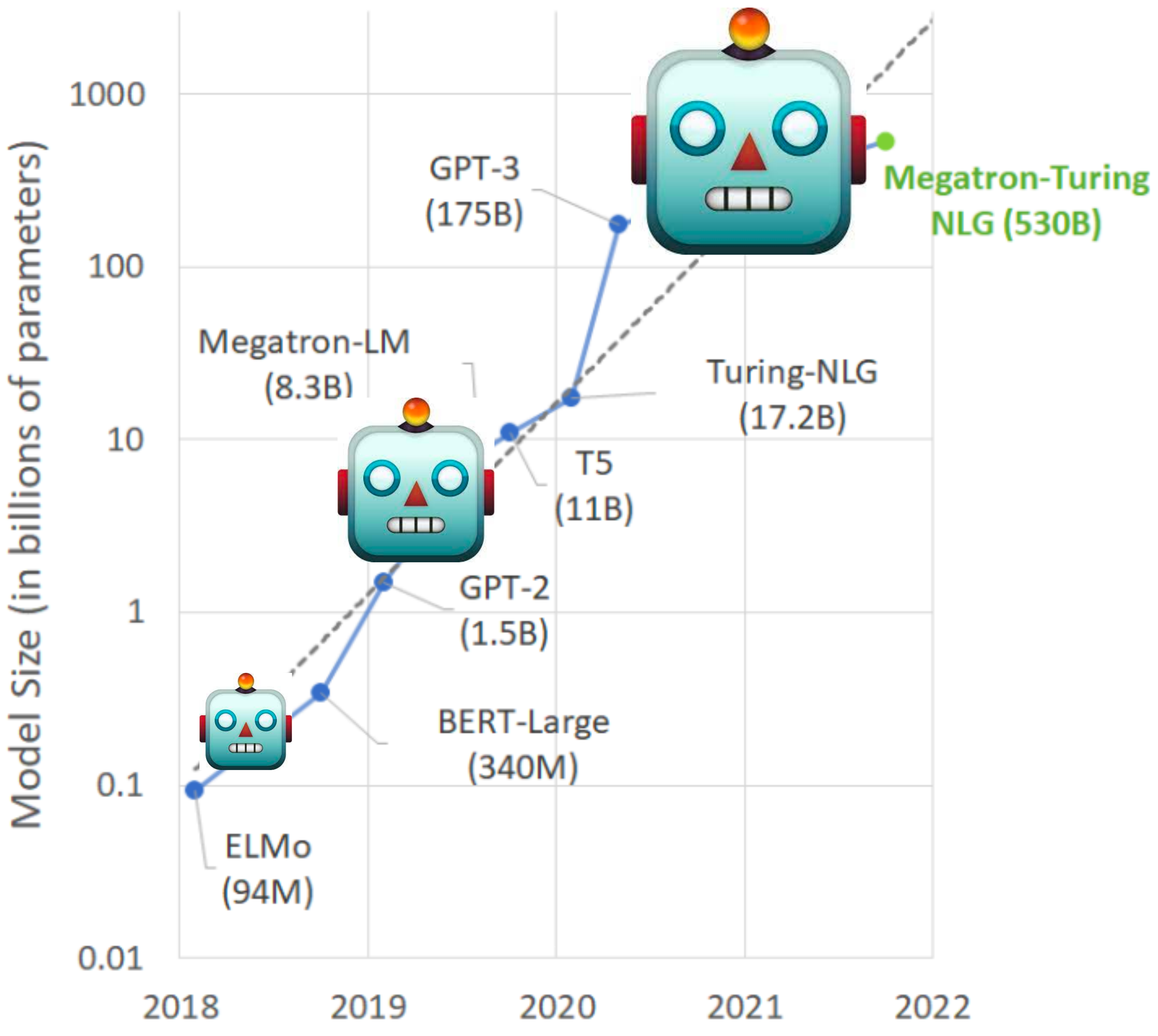
— 🏆 Best Method Paper Award at NAACL 2022 🏆 —

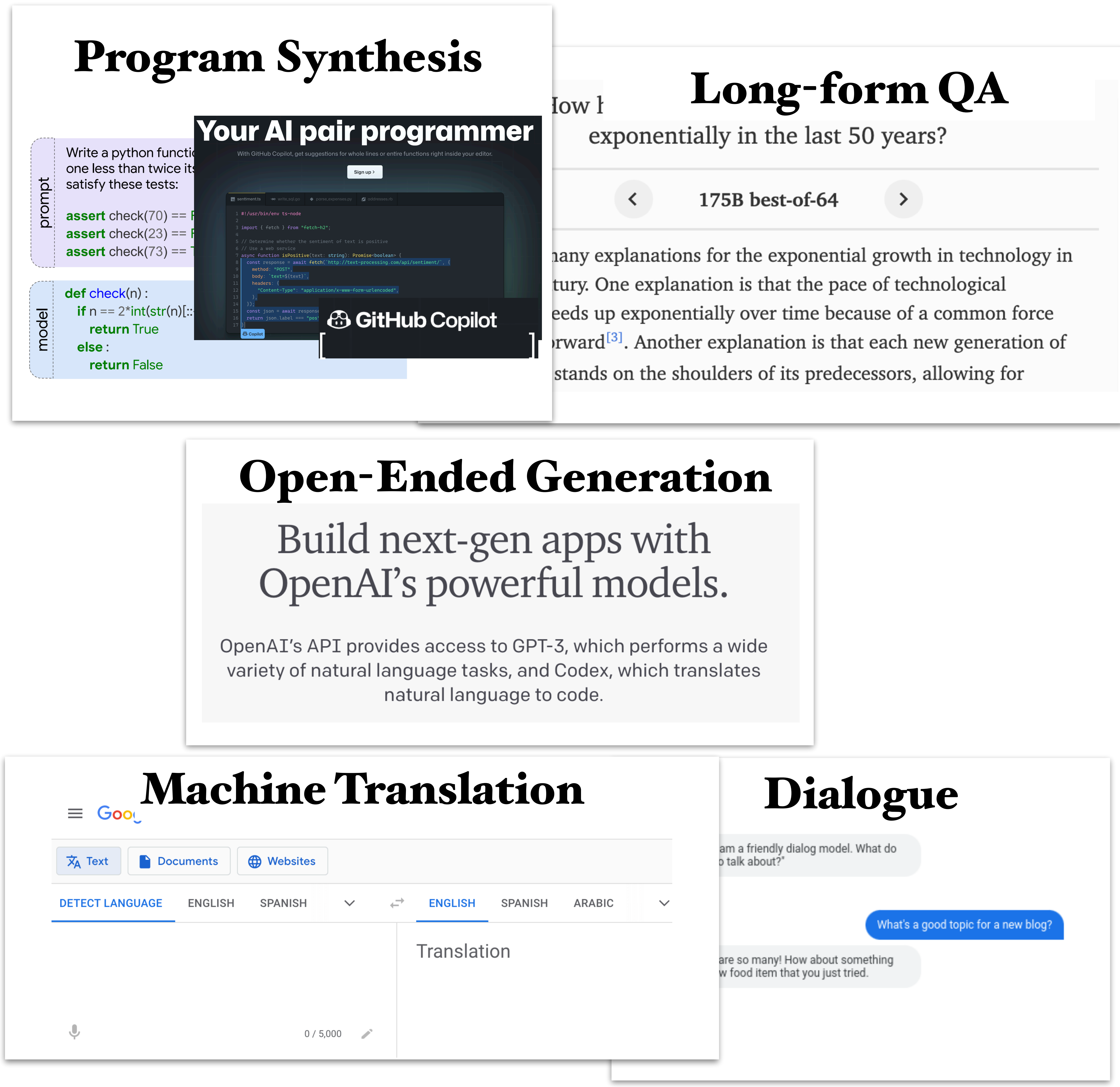
AI2

NeuroLogic A*esque

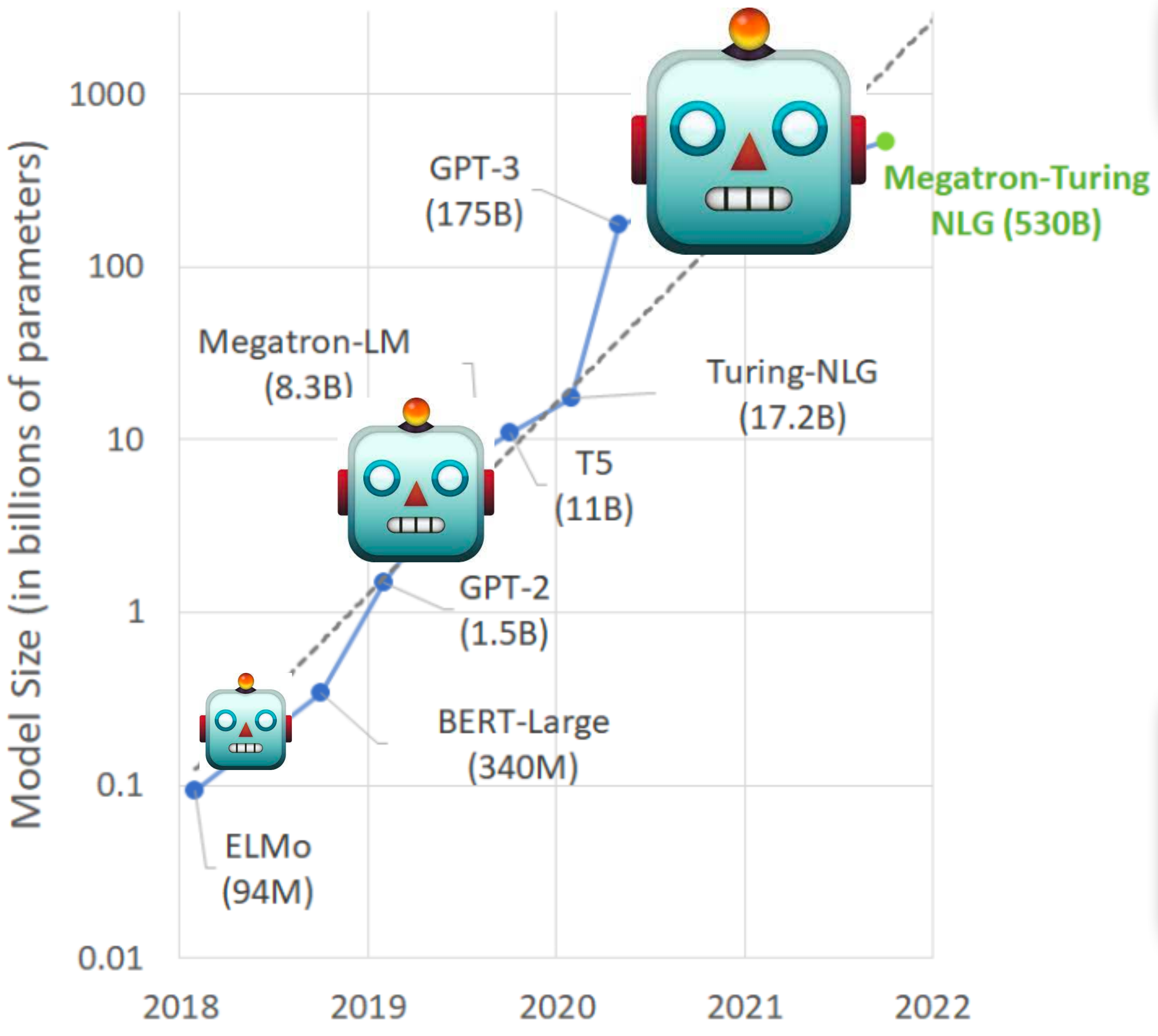
Constrained Text Generation with Lookahead Heuristic

Neural Language Models



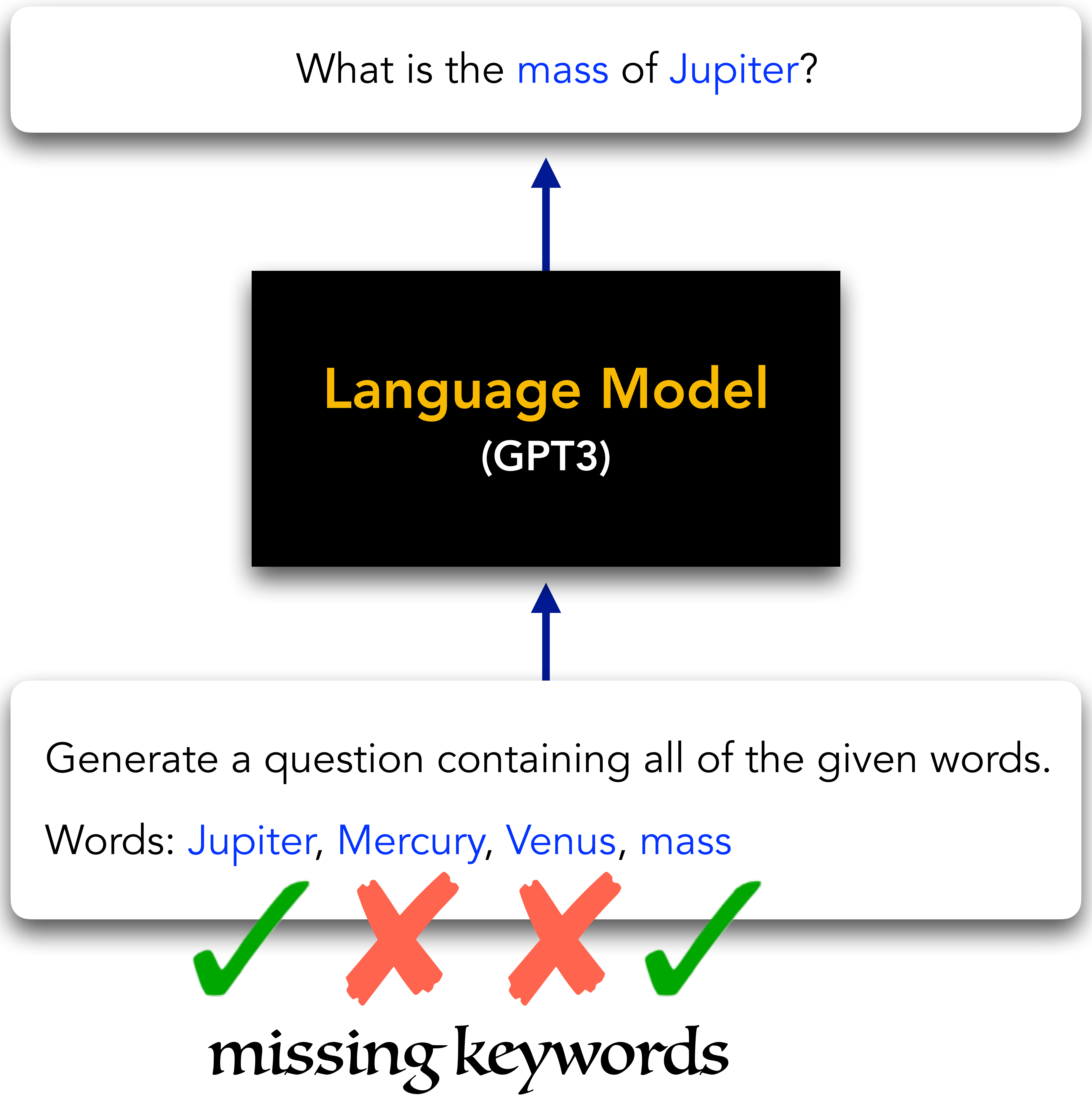


Neural Language Models

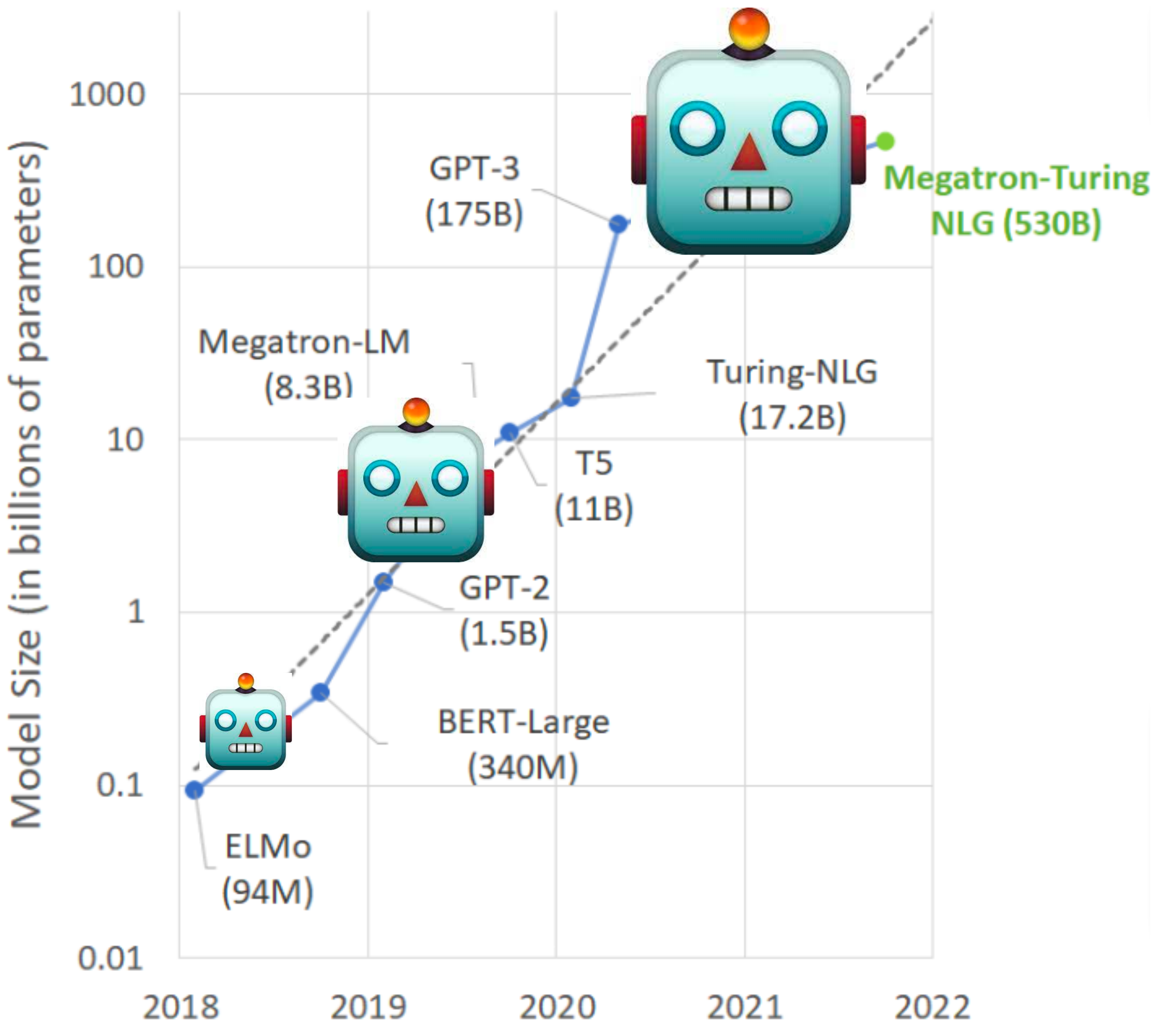


CommonGen

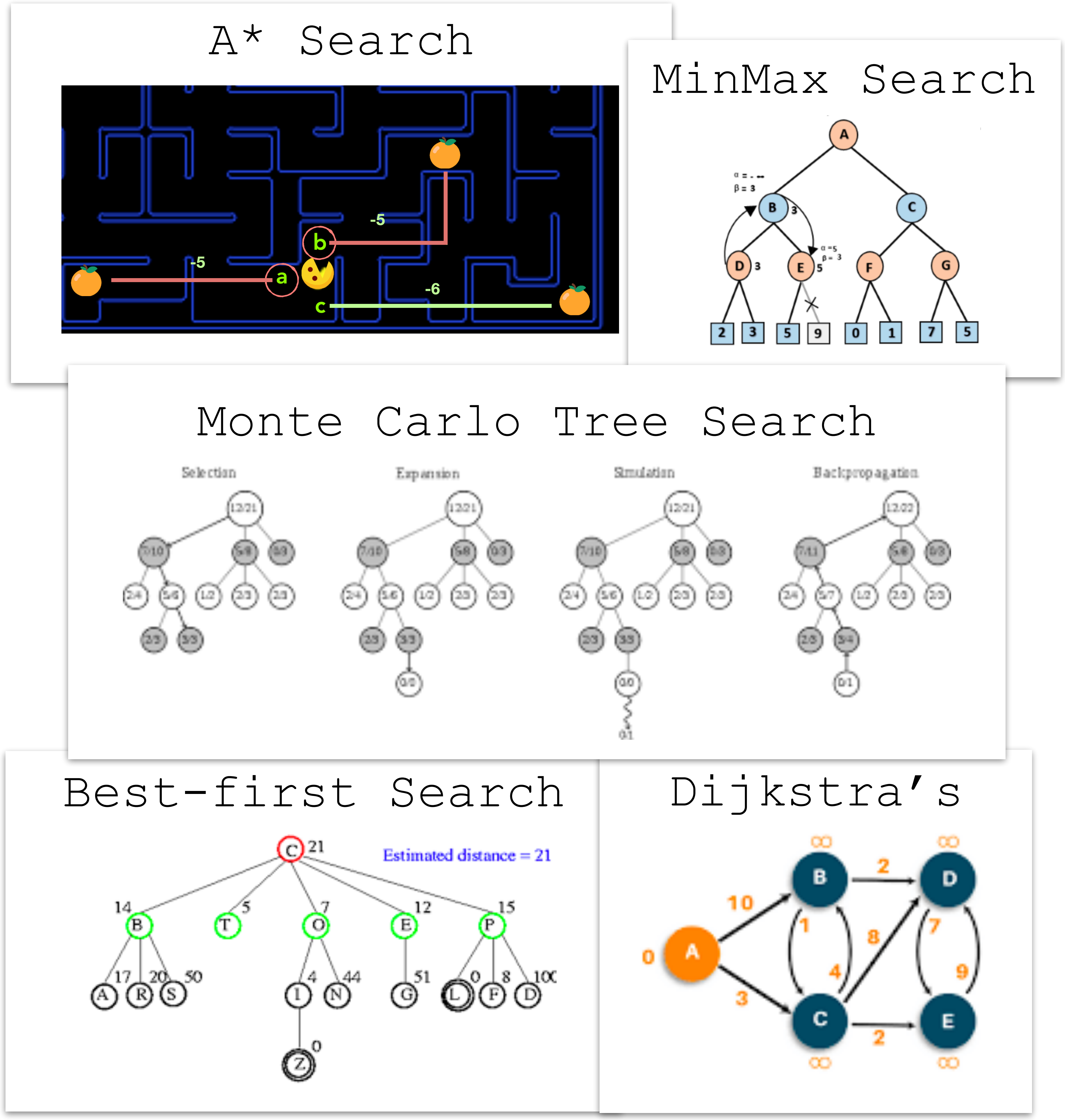
(Liu et al 2020)



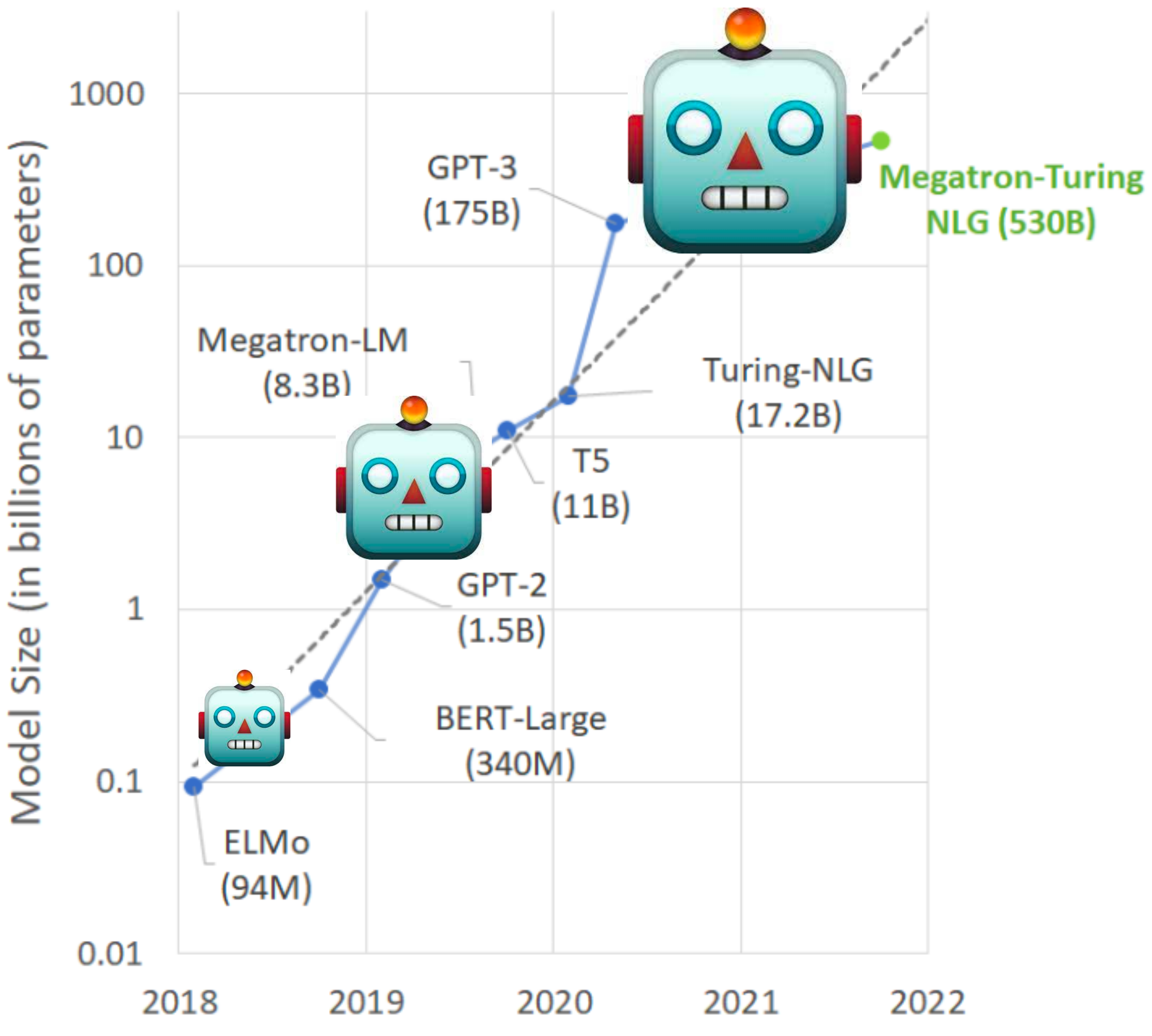
Neural Language Models



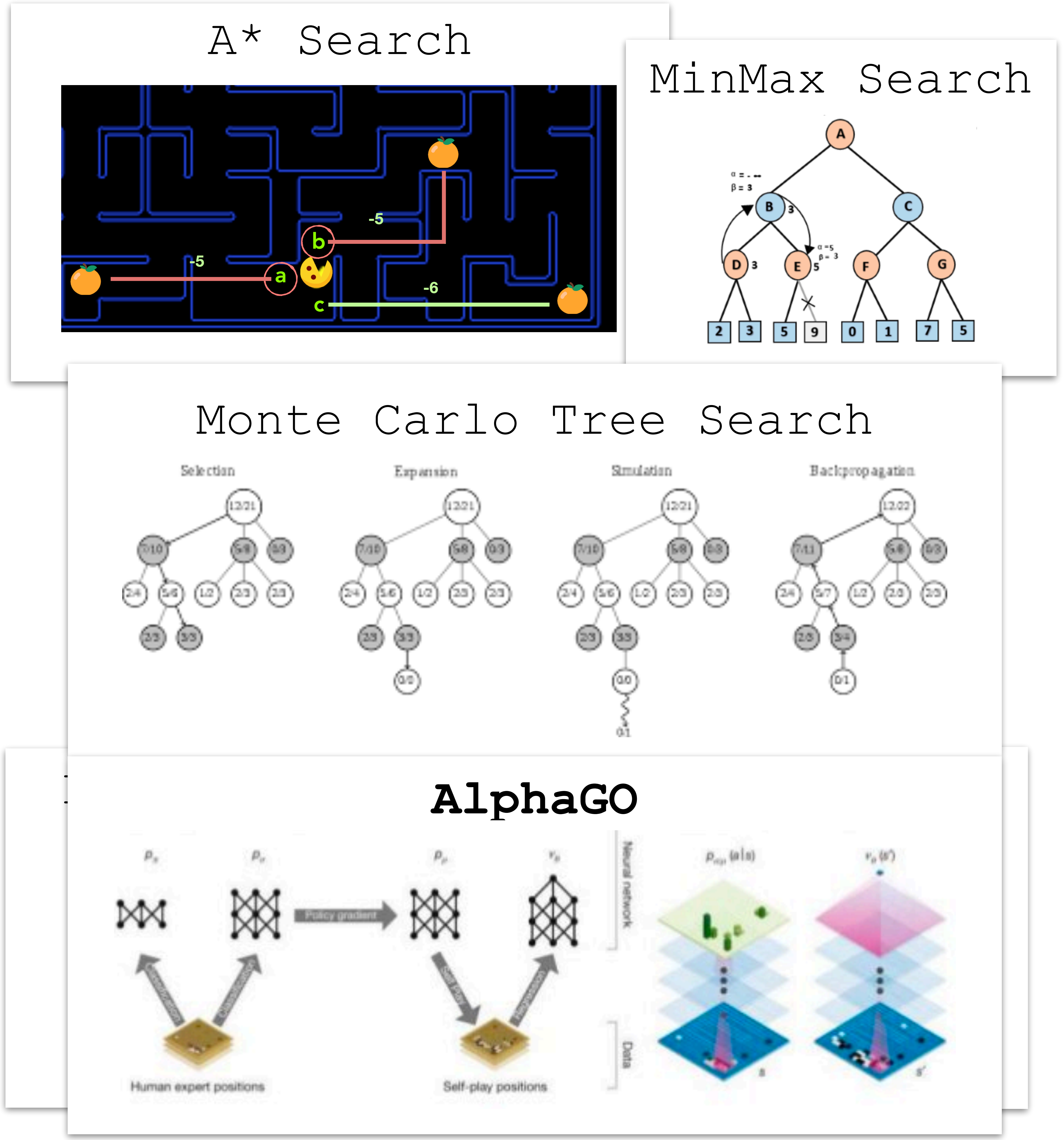
Search Algorithms in Classical AI



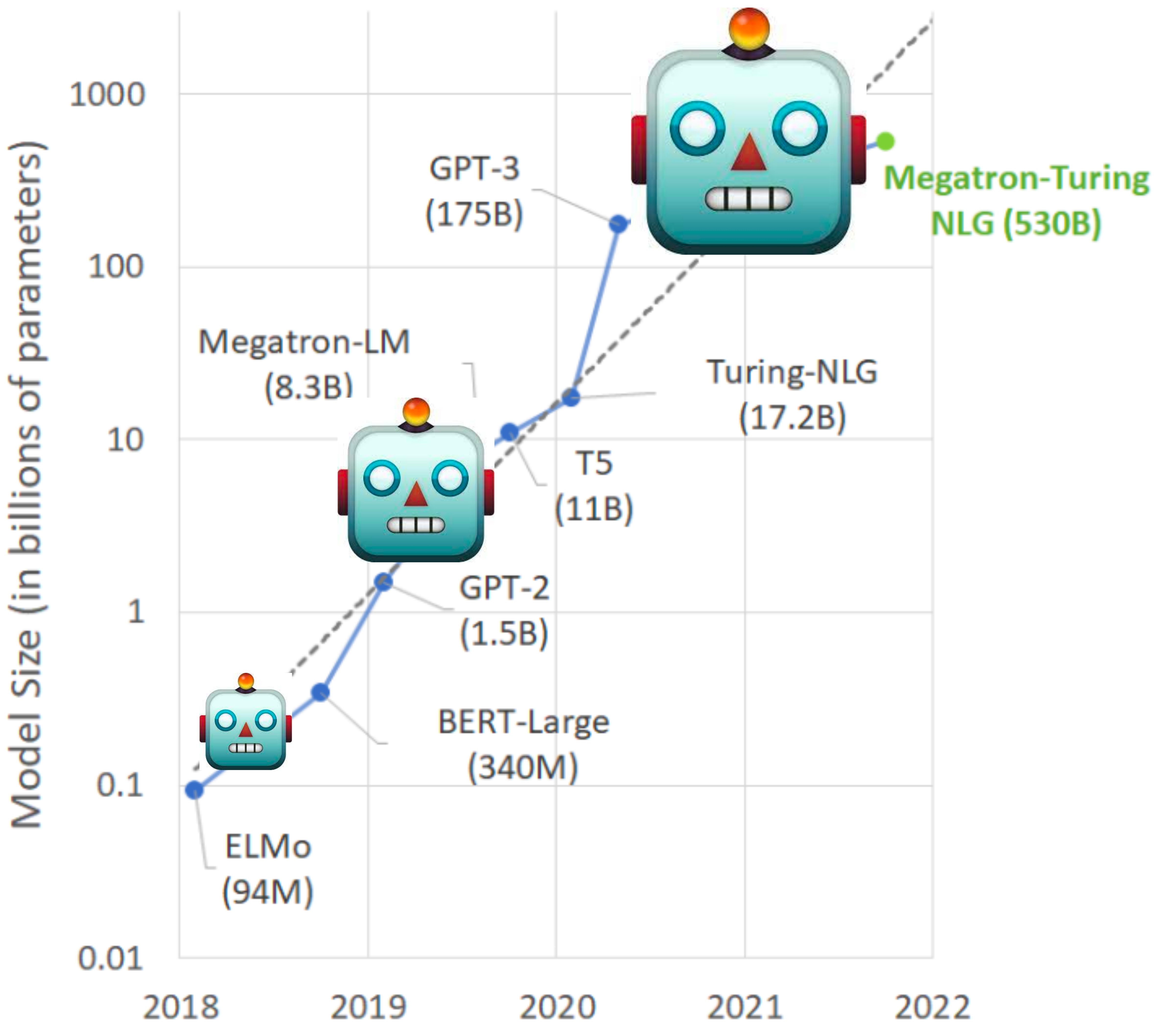
Neural Language Models



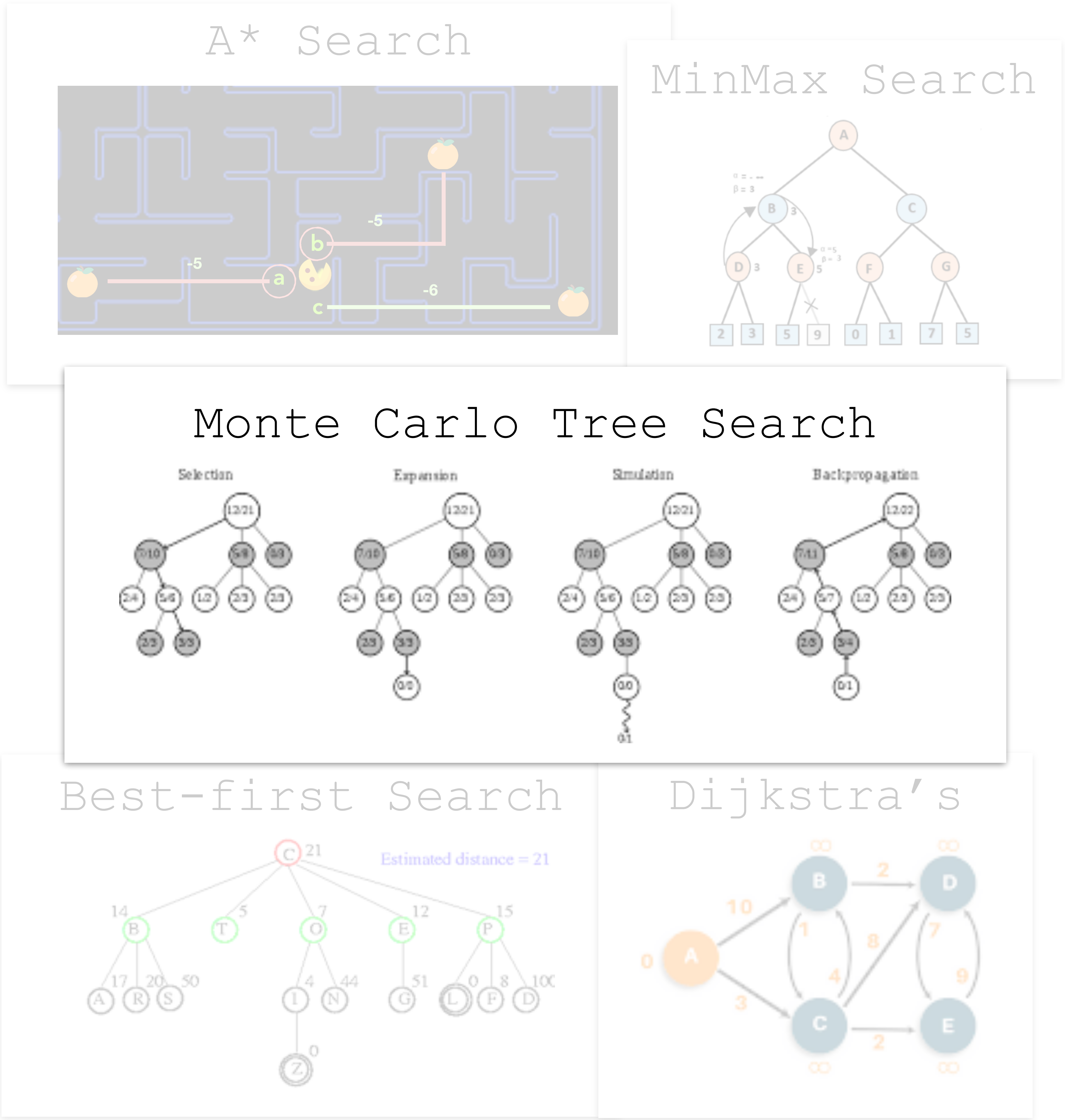
Search Algorithms in Classical AI

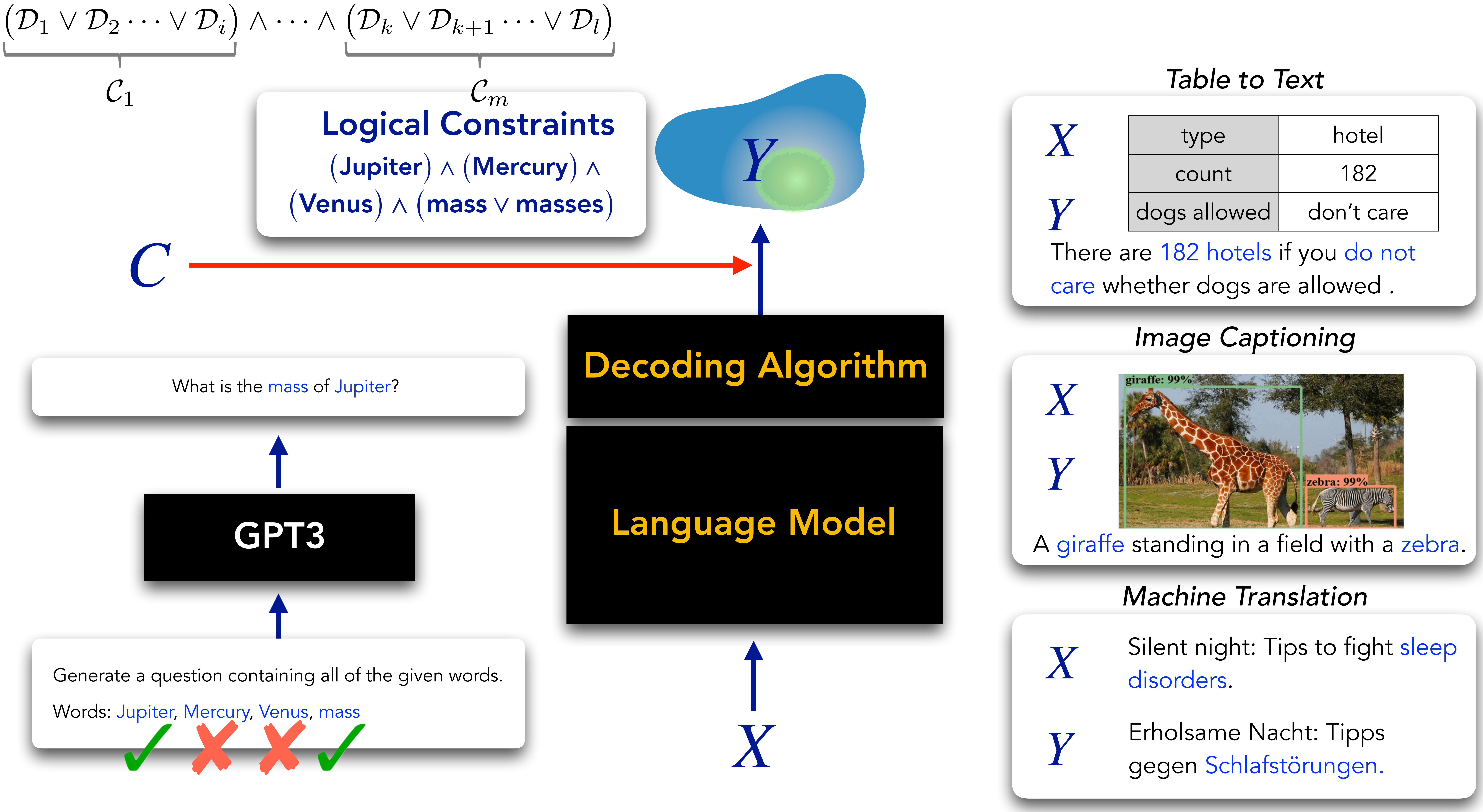


Neural Language Models



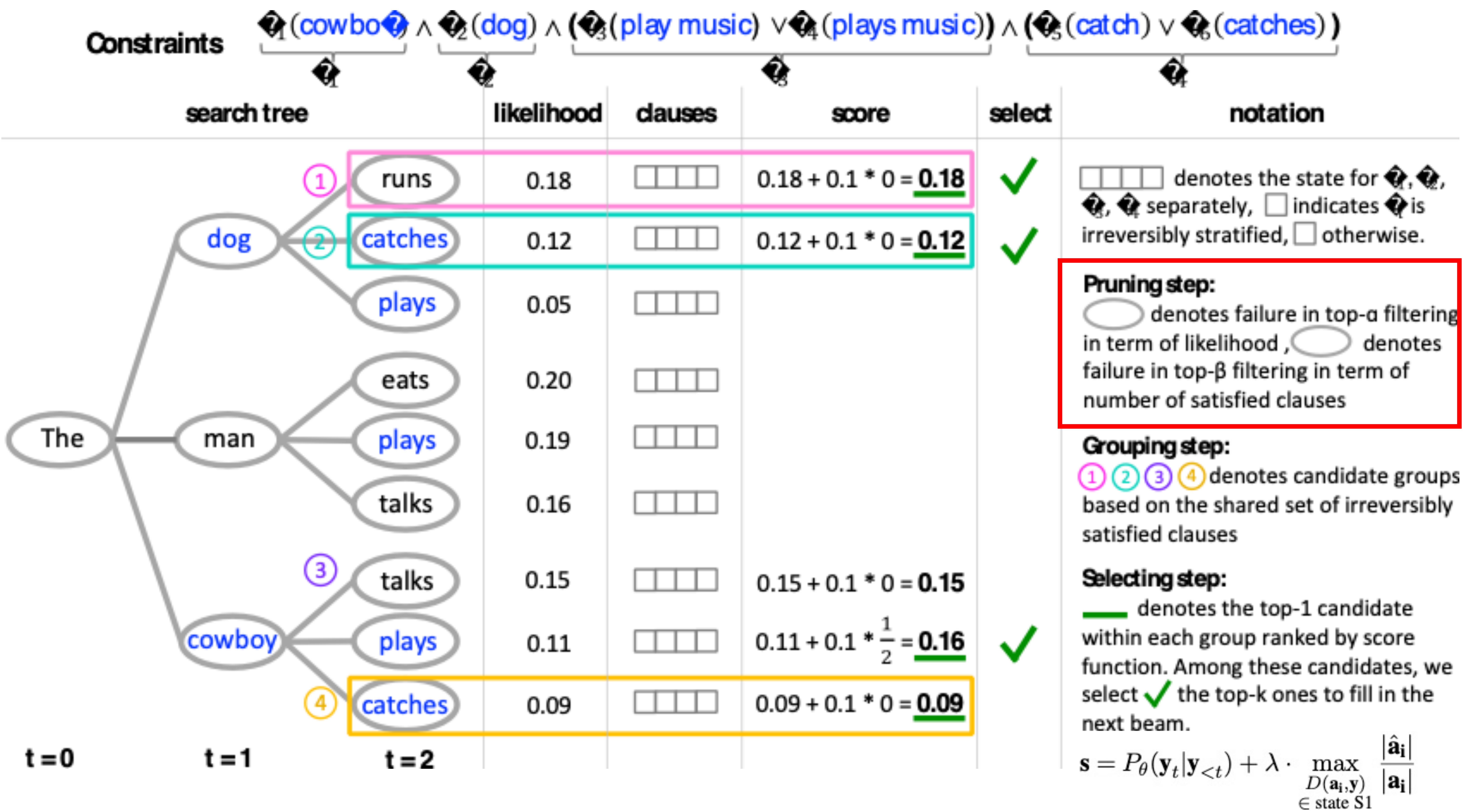
Search Algorithms in Classical AI





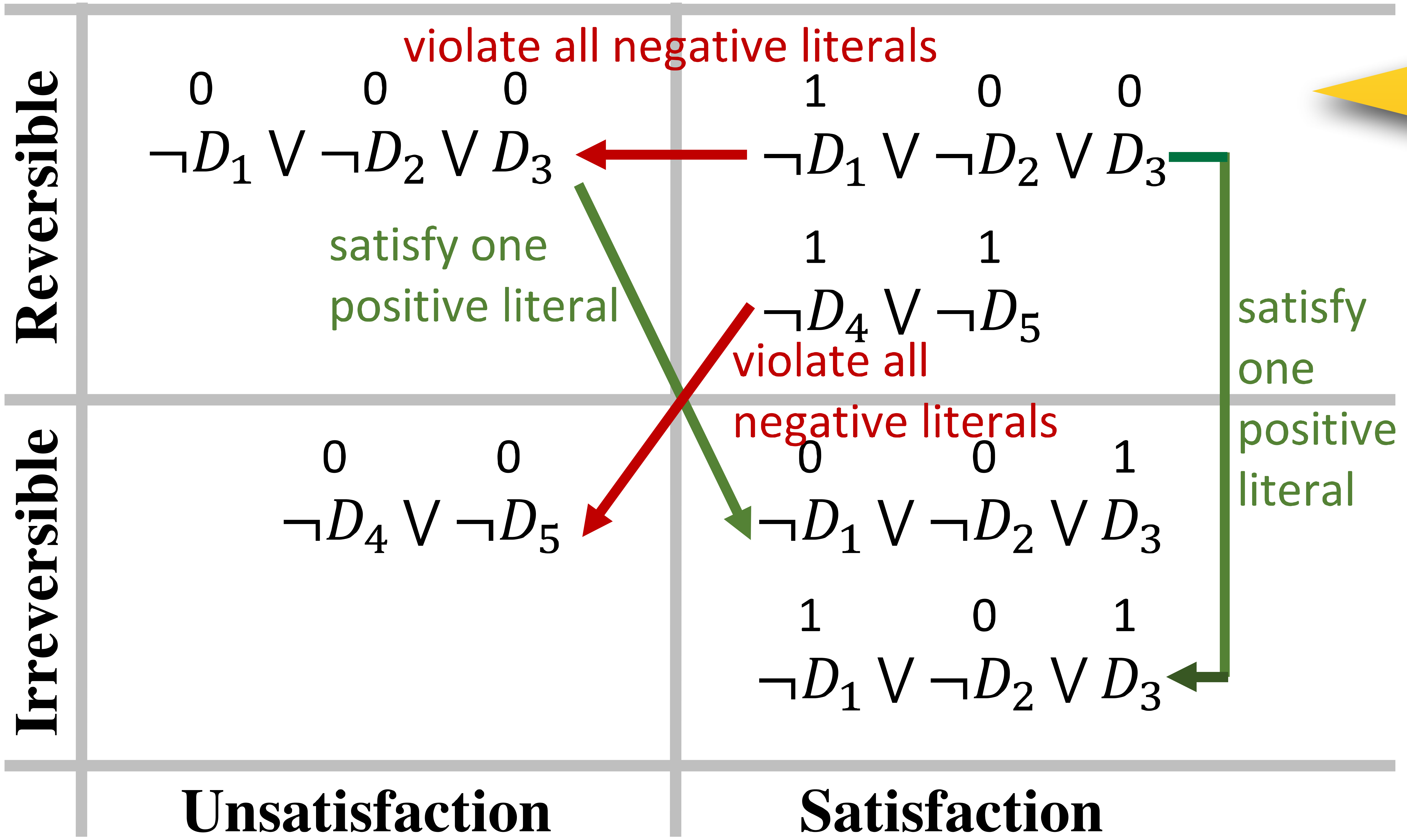
$(\mathcal{D}_1 \vee \mathcal{D}_2 \cdots \vee \mathcal{D}_i) \wedge \cdots \wedge (\mathcal{D}_k \vee \mathcal{D}_{k+1} \cdots \vee \mathcal{D}_l)$
$\mathcal{C}_1$
$\mathcal{C}_m$
Logical Constraints
(Jupiter) ∧ (Mercury) ∧
(Venus) ∧ (mass ∨ masses)
Y
C
Decoding Algorithm
Language Model
X
What is the mass of Jupiter?
GPT3
Generate a question containing all of the given words.
Words: Jupiter, Mercury, Venus, mass
Table to Text
X
Y
type hotel
count 182
dogs allowed don't care
There are 182 hotels if you do not care whether dogs are allowed .
Image Captioning
X
Y
giraffe: 99%
zebra: 99%
A giraffe standing in a field with a zebra.
Machine Translation
X
Silent night: Tips to fight sleep disorders.
Y
Erholsame Nacht: Tipps gegen Schlafstörungen.

NeuroLogic Decoding in a Nut Shell



NeuroLogic Decoding in a Nut Shell

— it's a logic-guided search algorithm



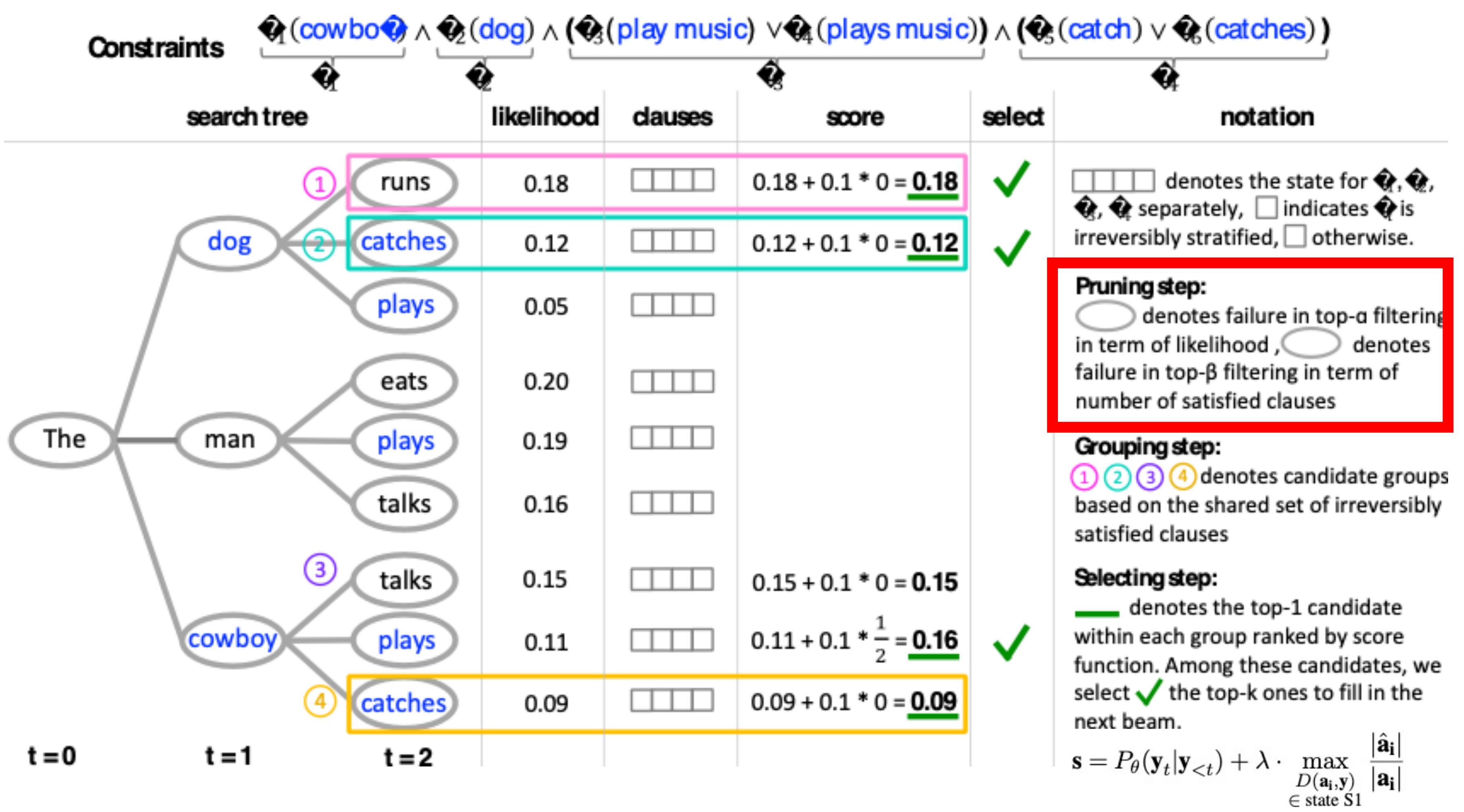
four states of clause satisfaction:

— reversible satisfaction

— irreversible satisfaction

— reversible unsatisfaction

— irreversible unsatisfaction

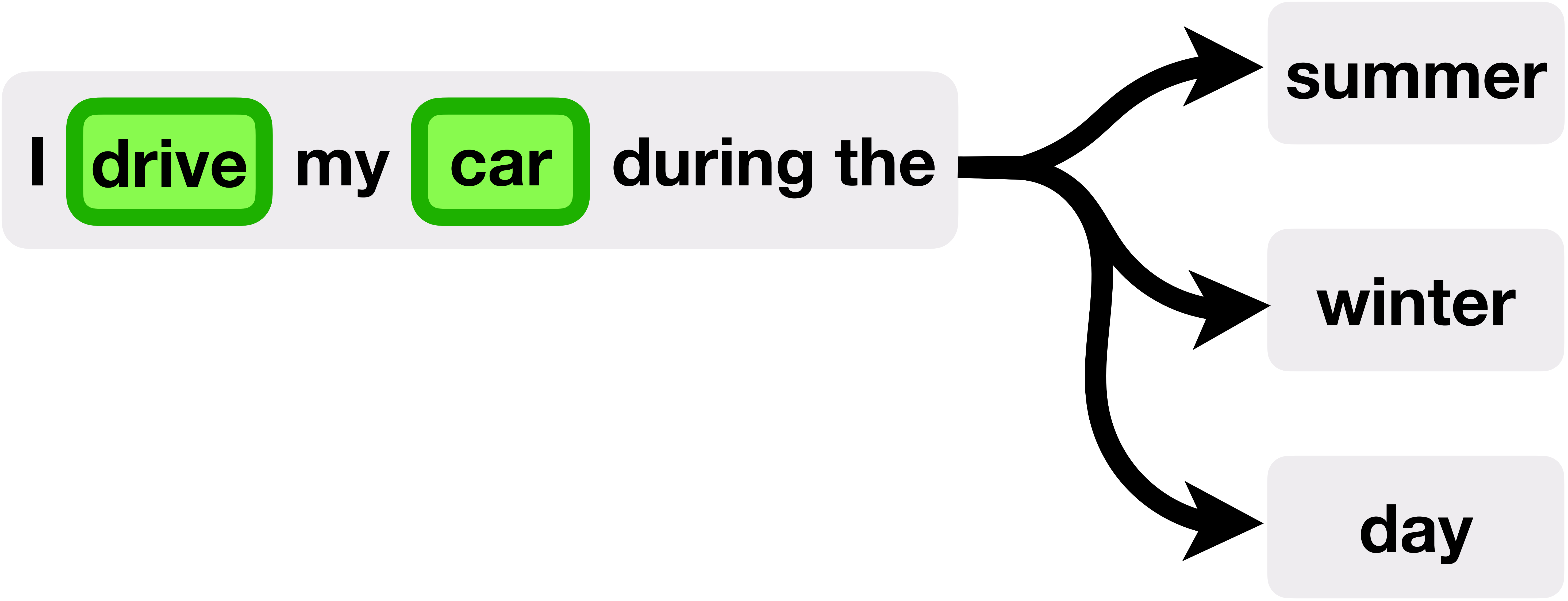


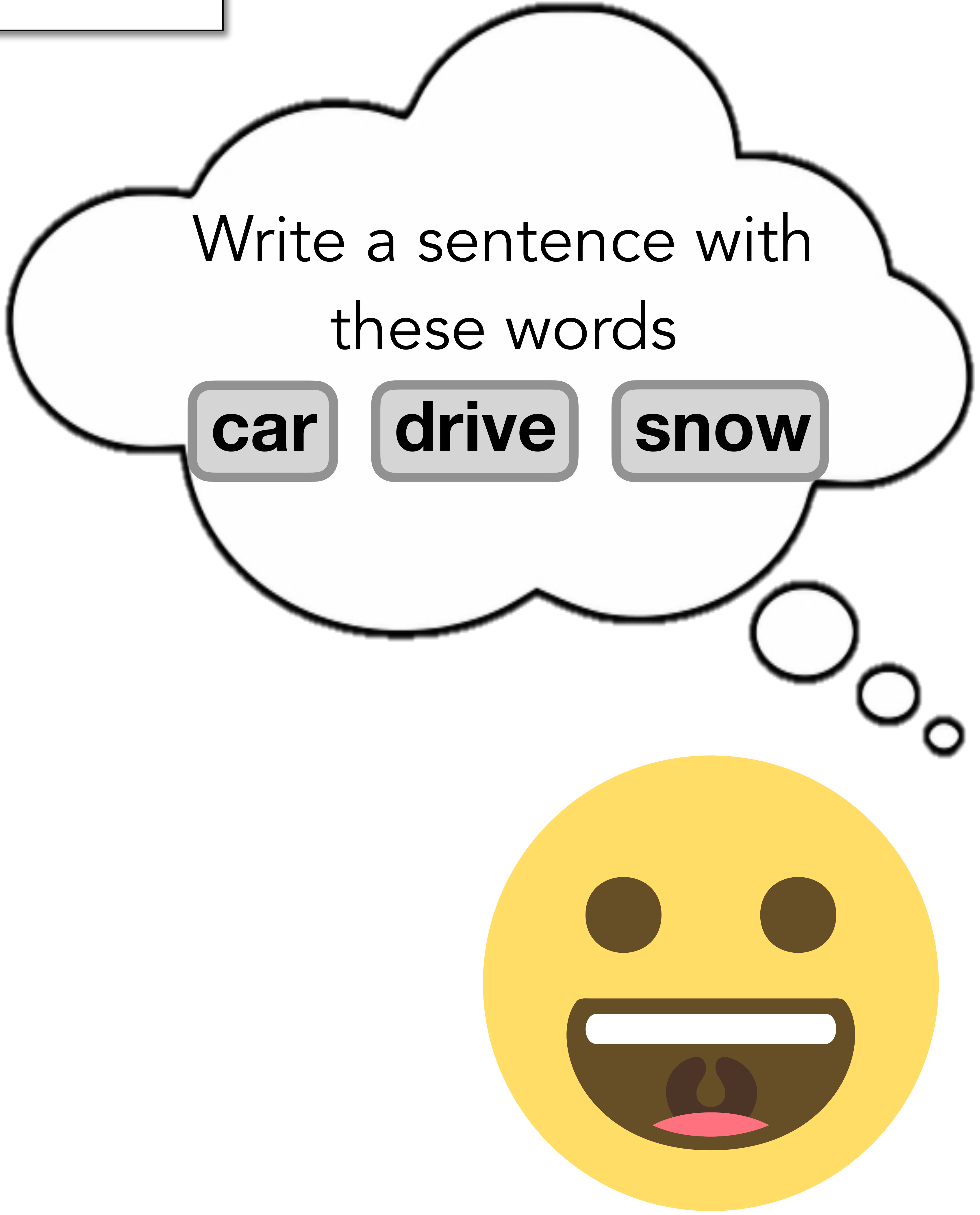
NeuroLogic Decoding

score $\quad s = \log P_\theta(\mathbf{y}_t \mid \mathbf{y}_{<t}) + \alpha' \sum_{i=1}^{m} C_i$

$D_1(\textbf{car}) \wedge D_2(\textbf{drive}) \wedge D_3(\textbf{snow})$

Off-the-Shelf GPT2





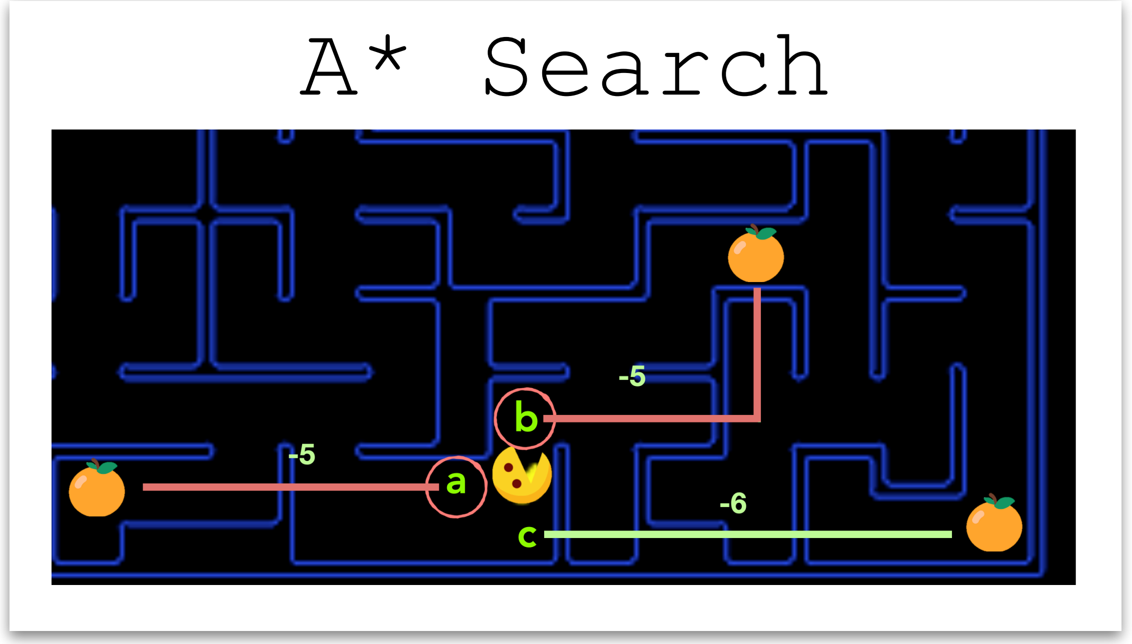
[ ]

NeuroLogic Decoding

score $\mathbf{s} = \log P_\theta(\mathbf{y}_t \mid \mathbf{y}_{<t}) + \alpha' \sum_{i=1}^{m} C_i$

$D_1(\textbf{car}) \wedge D_2(\textbf{drive}) \wedge D_3(\textbf{snow})$

[ ]



NeuroLogic A*esque

score $s = \log P_\theta(\mathbf{y}_t \mid \mathbf{y}_{<t}) + \alpha' \sum_{i=1}^{m} C_i + \lambda_1 \cdot \max_{\{D_i:\, D_i=0\}} \log P_\theta(D_i \mid \mathbf{y}_{<t+k})$

A* Heuristic

$D_1(\textbf{car}) \wedge D_2(\textbf{drive}) \wedge D_3(\textbf{snow})$

Can we use unsatisfied constraints to guide the search?

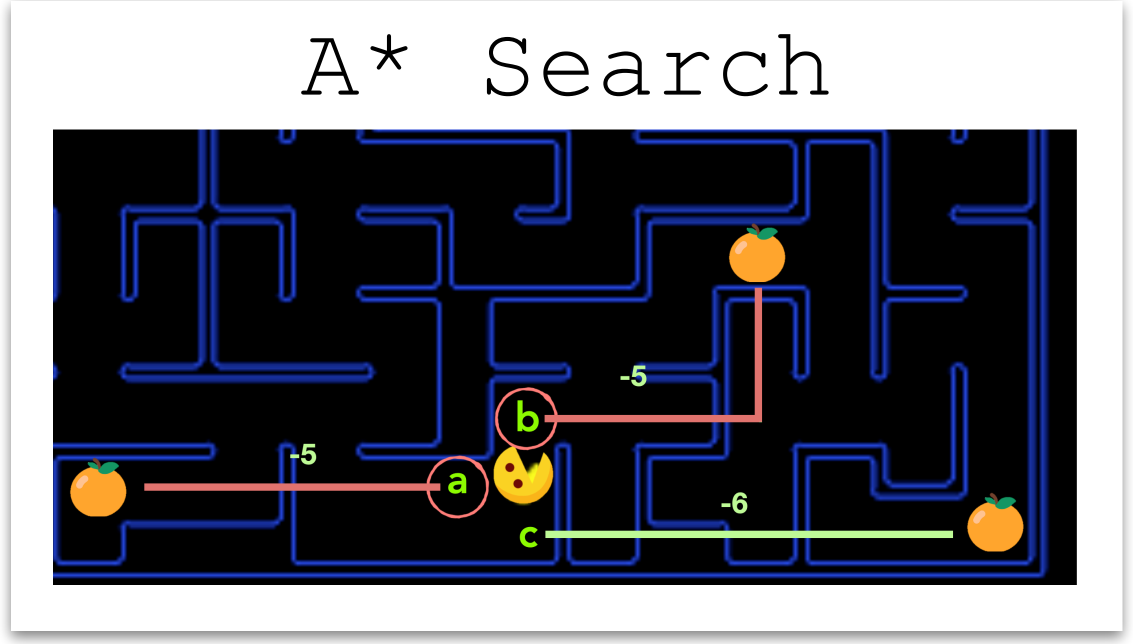
Off-the-Shelf GPT2

I drive my car during the

summer $p(w \mid past) = 0.4$ ✗

winter $p(w \mid past) = 0.2$ — **snow** ✓

day

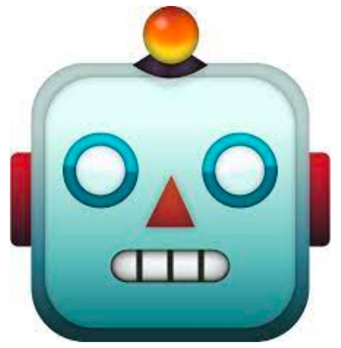


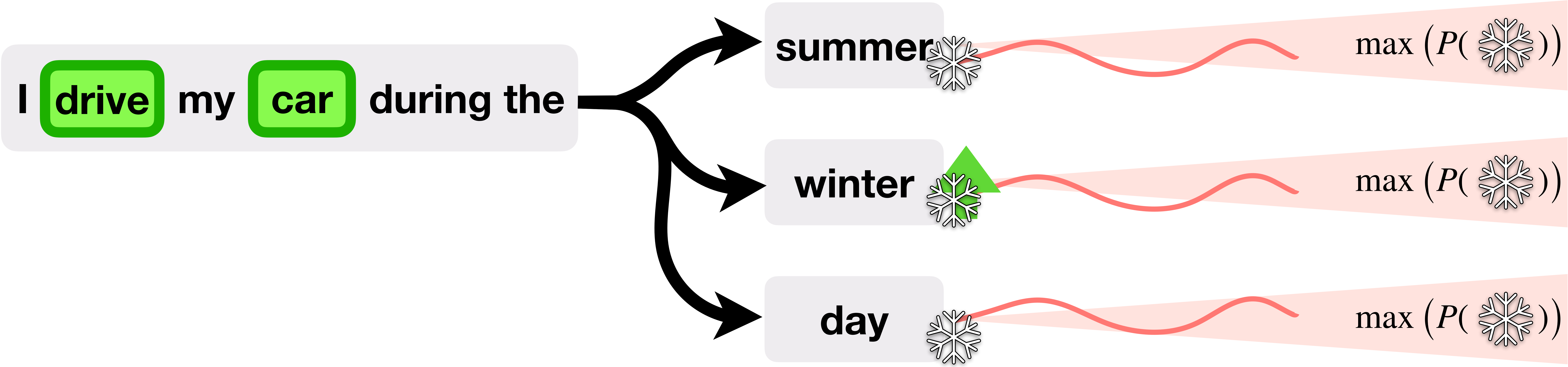
NeuroLogic A*esque

score $\mathbf{s} = \log P_\theta(\mathbf{y}_t | \mathbf{y}_{<t}) + \alpha' \sum_{i=1}^{m} C_i + \lambda_1 \cdot \max_{\{D_i:\, D_i=0\}} \log P_\theta(D_i | \mathbf{y}_{<t+k})$ ← A* Heuristic

$D_1(\textbf{car}) \wedge D_2(\textbf{drive}) \wedge D_3(\textbf{snow})$

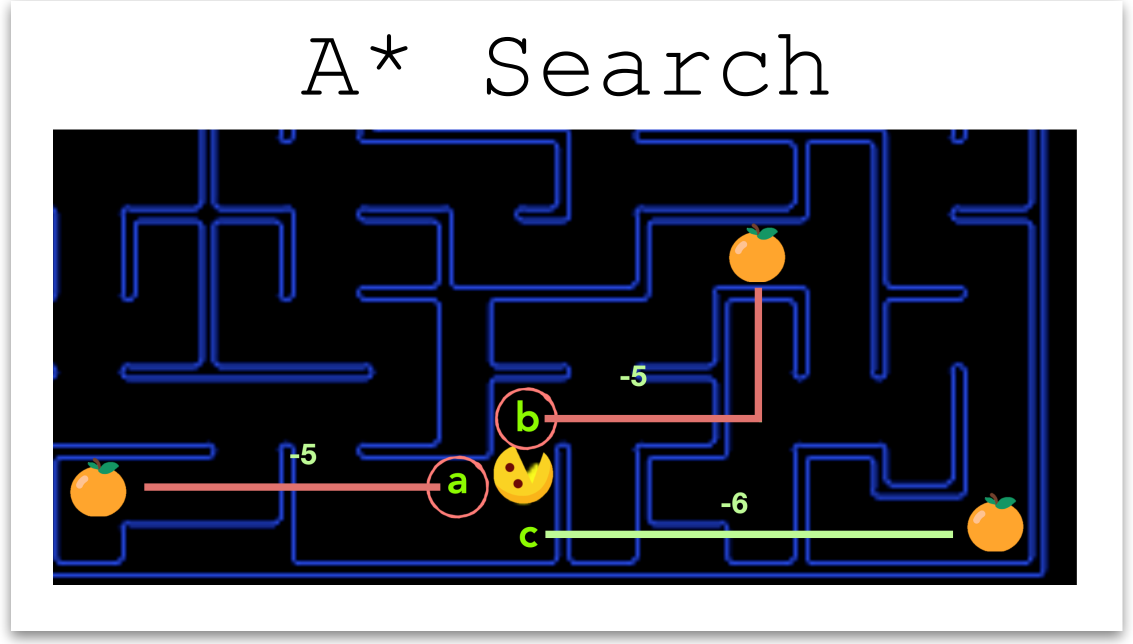
Off-the-Shelf GPT2






A* Search
-5
-5
-6
a
b
c
NeuroLogic A*esque
score $\mathbf{s} = \log P_\theta(\mathbf{y}_t \mid \mathbf{y}_{<t}) + \alpha' \sum_{i=1}^{m} C_i + \lambda_1 \cdot \max_{\{D_i:\, D_i=0\}} \log P_\theta(D_i \mid \mathbf{y}_{<t+k})$
A* Heuristic
$D_1(\text{car}) \wedge D_2(\text{drive}) \wedge D_3(\text{snow})$
greedy look-ahead $y_{t' \in [1,k]} = \arg\max_{y \in \mathcal{V}} P_\theta(y \mid \mathbf{y}_{<t'})$
A* heuristics $P_\theta(D_i(\mathbf{a}) \mid \mathbf{y}_{\leq t+k}) = \max_{i \in [1,k]} P_\theta(\mathbf{y}_{t+i:t+i+|\mathbf{a}|} = \mathbf{a} \mid \mathbf{y}_{<t+i})$
Off-the-Shelf GPT2
I drive my car during the
summer
winter
day
max (P( ) P( ) P( ))
max (P( ) P( ) P( ))
max (P( ) P( ) P( ))

NeuroLogic A*esque



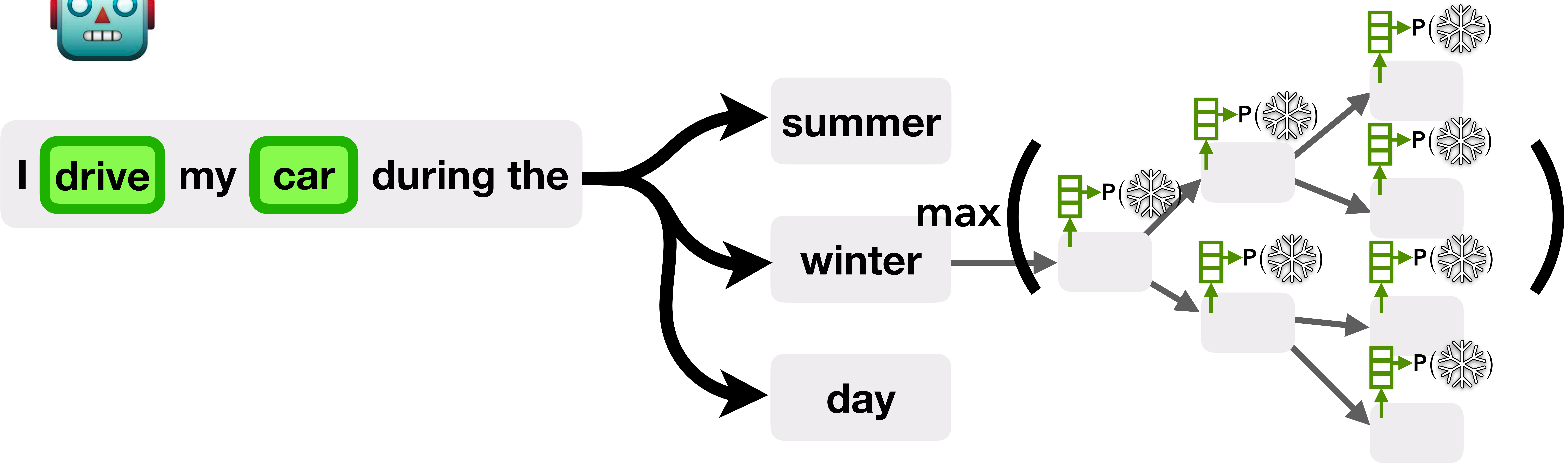
score $s = \log P_\theta(\mathbf{y}_t|\mathbf{y}_{<t}) + \alpha' \sum_{i=1}^{m} C_i + \lambda_1 \cdot \max_{\{D_i:\, D_i=0\}} \log P_\theta(D_i|\mathbf{y}_{<t+k})$ ← A* Heuristic

beam look-ahead $Y_{t'\in[1,k]} = \arg \mathbf{topk}_{y\in\mathcal{V}} P_\theta(y|\mathbf{y}_{<t'})$

A* heuristics $P_\theta(D_i(\mathbf{a})\,|\,Y_{\leq t+k}) = \max_{\mathbf{y}\in Y} \max_{i\in[1,k]} P_\theta(\mathbf{y}_{t+i:t+i+|\mathbf{a}|} = \mathbf{a}\,|\,\mathbf{y}_{<t+i})$

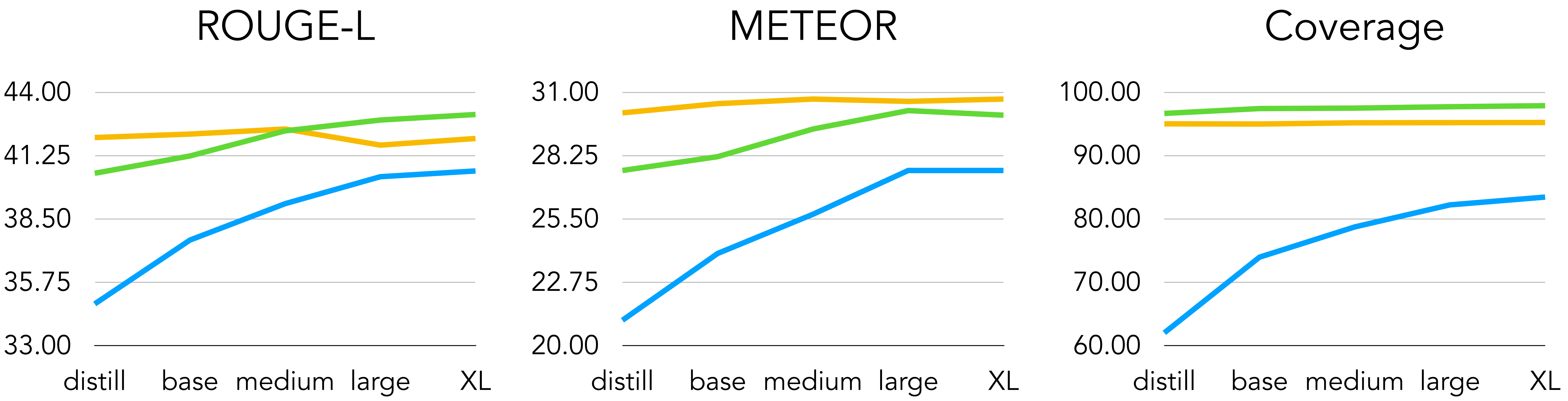
$D_1(\text{car}) \wedge D_2(\text{drive}) \wedge D_3(\text{snow})$

Off-the-Shelf GPT2




A* Search
-5
-5
-6
b
a
c
NeuroLogic A*ESQUE
score $s = \log P_\theta(\mathbf{y}_t | \mathbf{y}_{<t}) + \alpha' \sum_{i=1}^{m} C_i + \lambda_1 \cdot \max_{\{D_i:\, D_i=0\}} \log P_\theta(D_i | \mathbf{y}_{<t+k})$
A* Heuristic
sampling look-ahead $y_{t' \in [1,k]} \sim P_\theta(y | \mathbf{y}_{<t'})$
A* heuristics $P_\theta(D_i(\mathbf{a}) | Y_{\le t+k}) = \max_{\mathbf{y} \in Y} \max_{i \in [1,k]} P_\theta(\mathbf{y}_{t+i:t+i+|\mathbf{a}|} = \mathbf{a} | \mathbf{y}_{<t+i})$
D_1(car) $\wedge$ D_2(drive) $\wedge$ D_3(snow)
Off-the-Shelf GPT2
I drive my car during the
summer
winter
day
sample
max
P( )

CommonGen (Zero-shot)

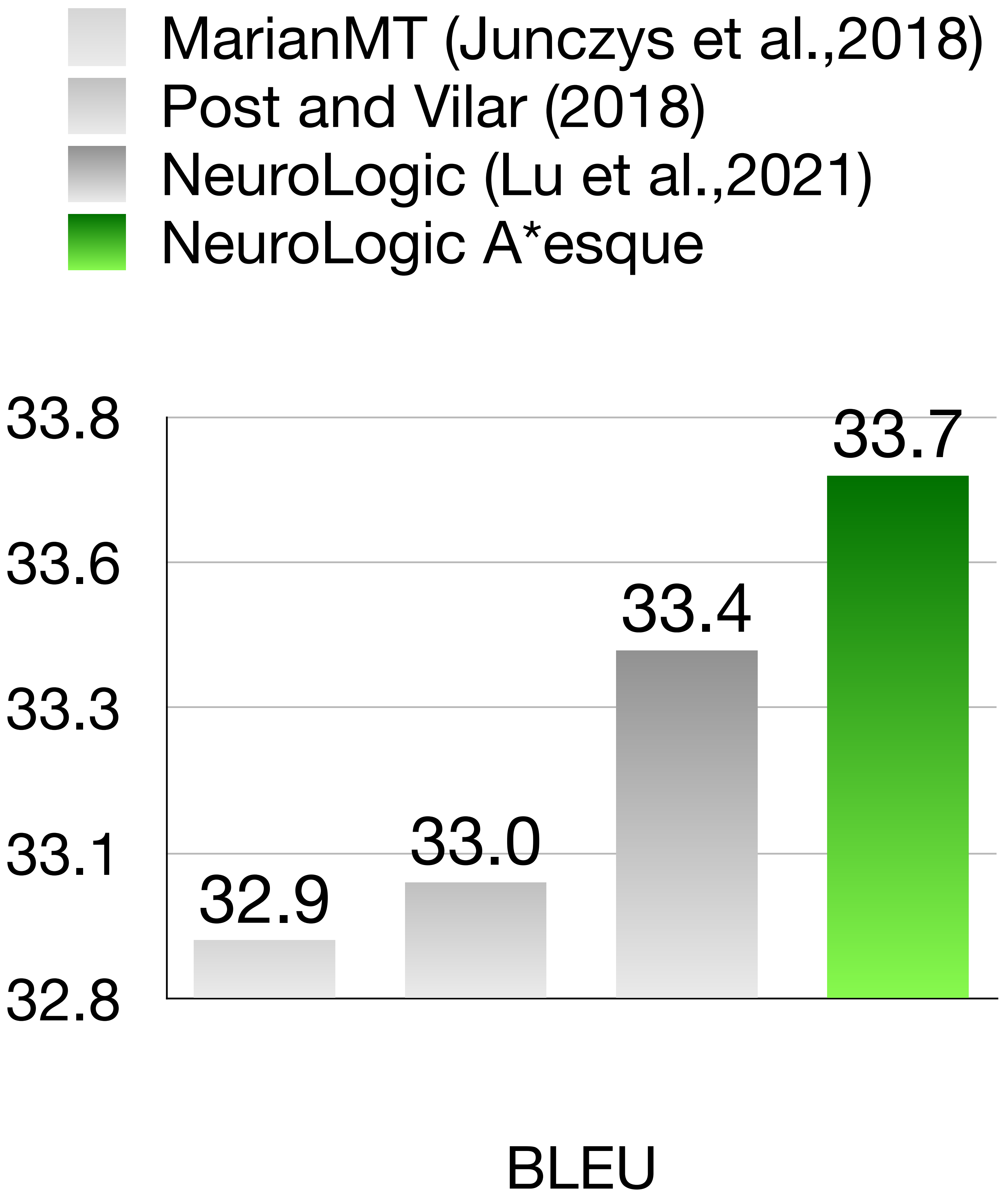


Unsupervised NeuroLogic outperforms **supervised approaches**

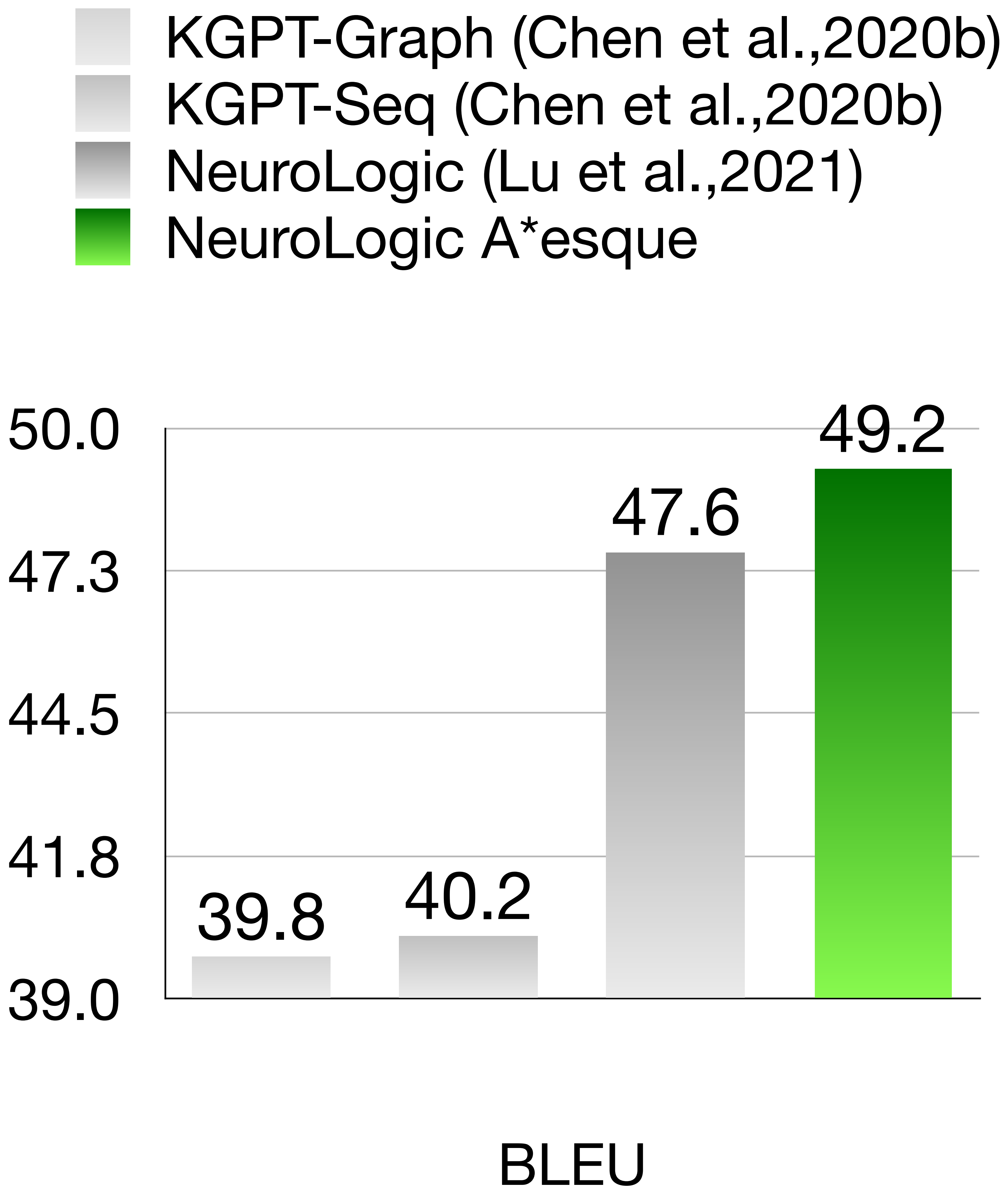
Unsupervised NeuroLogic on smaller networks outperforms **supervised approaches on larger networks!**

Neurologic A* esque **generalize** to many downstream tasks

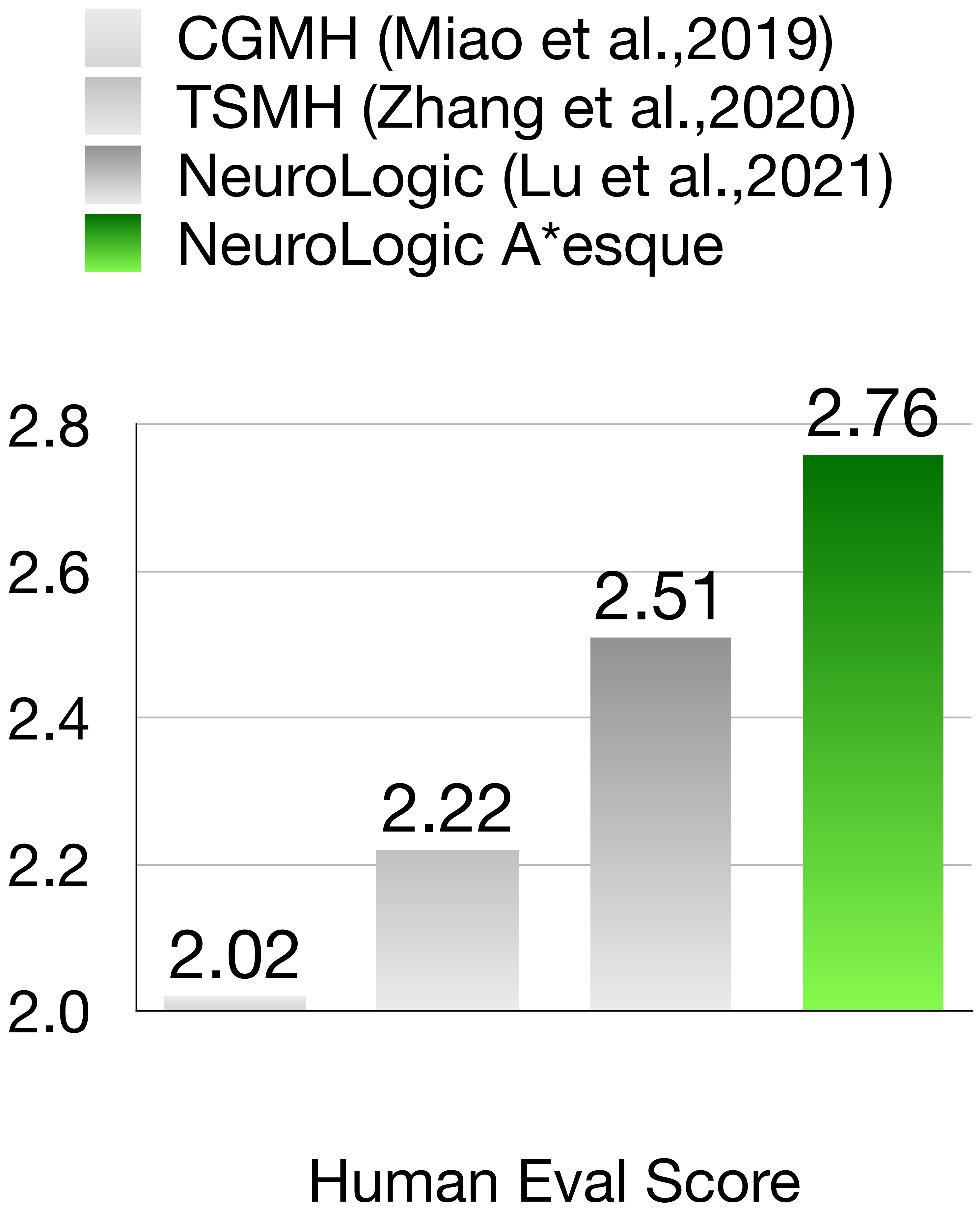
Constrained MT
(Dinu et al., 2019)



Few-Shot E2ENLG
(Chen et al., 2020)



Question Generation
(Zhang et al., 2020)

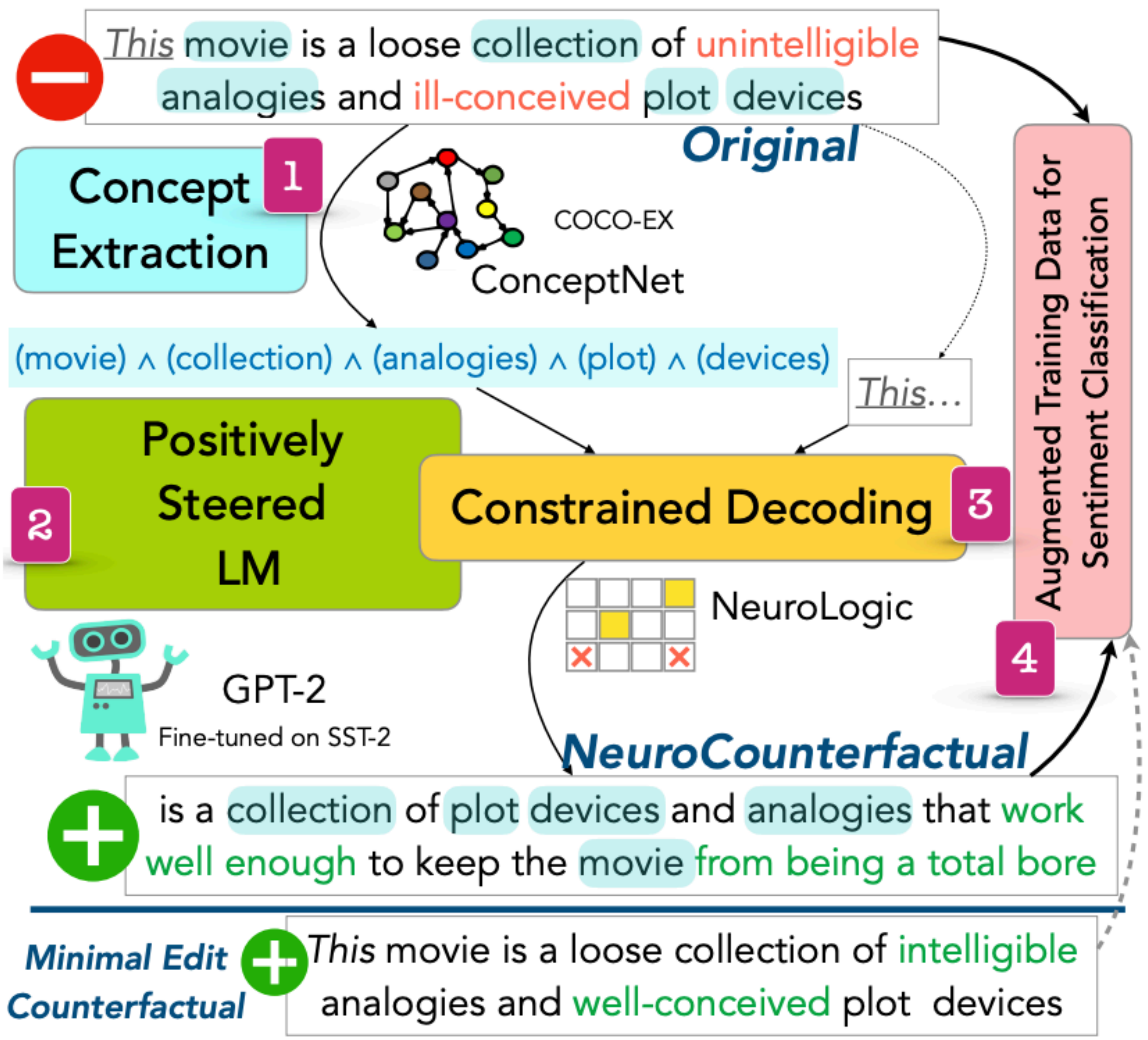


NeuroCounterfactuals: Beyond Minimal-Edit Counterfactuals for Richer Data Augmentation

Phillip Howard[◇] **Gadi Singer**[◇] **Vasudev Lal**[◇] **Yejin Choi**[♡♣] **Swabha Swayamdipta**[♣♠]
[◇]Intel Labs [♣]Allen Institute for AI [♠]University of Southern California
[♡]Paul G. Allen School of Computer Science & Engineering, University of Washington
phillip.r.howard@intel.com

Abstract

While counterfactual data augmentation offers a promising step towards robust generalization in natural language processing, producing a set of counterfactuals that offer valuable inductive bias for models remains a challenge. Most existing approaches for producing counterfactuals, manual or automated, rely on small perturbations via minimal edits, resulting in simplistic changes. We introduce NeuroCounterfactuals, designed as loose counterfactuals, allowing for larger edits which result in naturalistic generations containing linguistic diversity, while still bearing similarity to the origi-



2082: An AI Odyssey

Prolog: what AI in 2082 be like

Chapter 1: The ambiguity

Chapter 2: The continuum

➡ Chapter 3: The dark matter

Speculations

Epilog: a confession of an alien

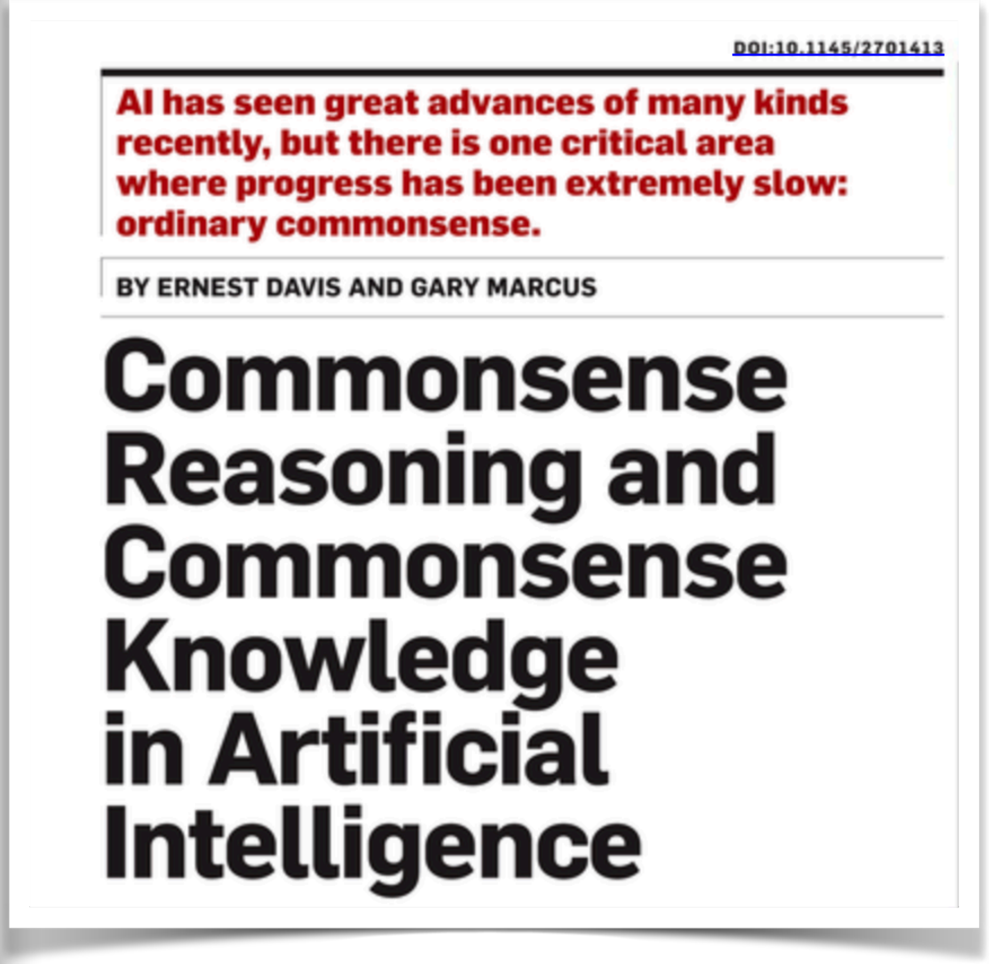
Dark matter is what matters in modern physics

- Only 5% of universe is normal matter. The remaining 95% is dark matter and dark energy.
- Dark matter is completely invisible, yet affects what are visible: the orbits of stars and the trajectory of light

Dark matter of language?

Normal matter: visible text (words, sentences)

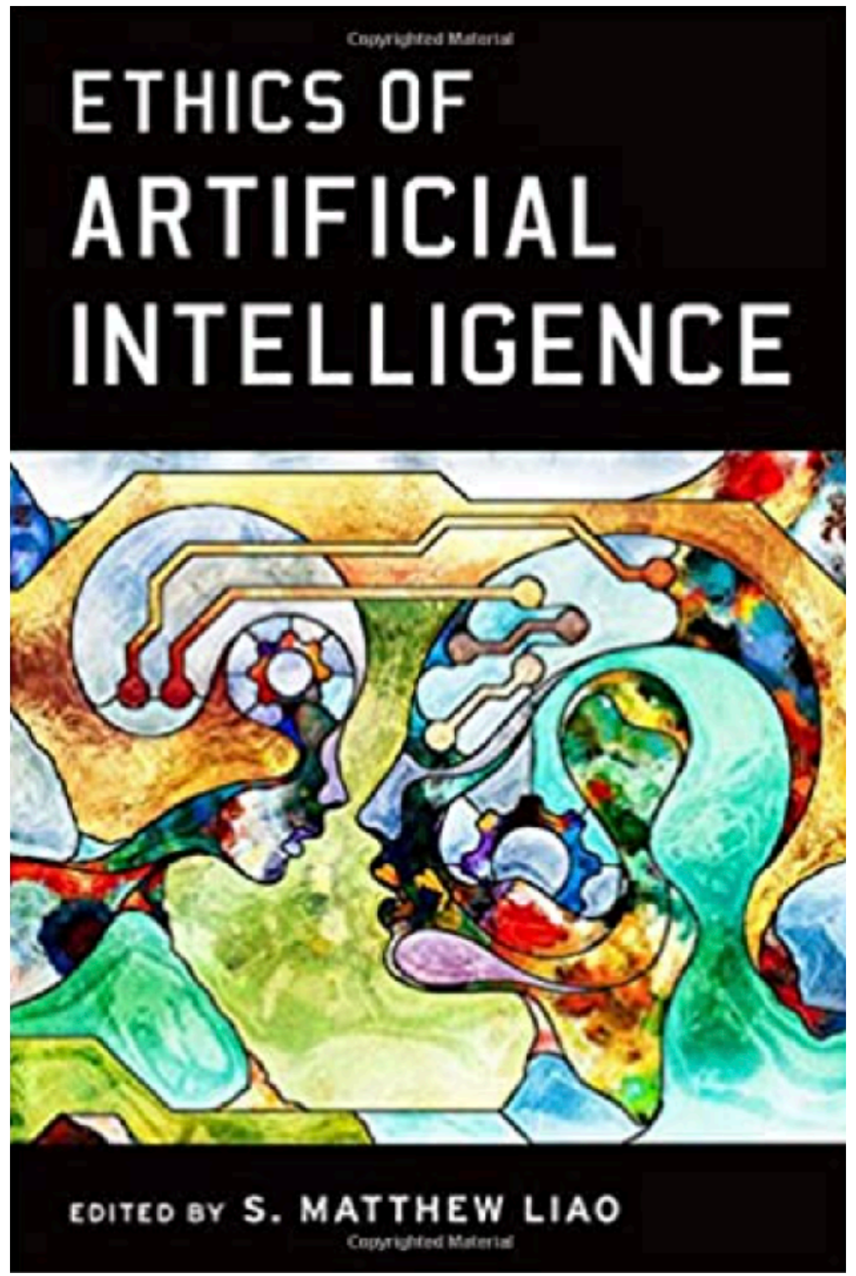
Dark matter: the unspoken rules of how the world works, which influence the way people use and interpret language


DOI:10.1145/2701413

AI has seen great advances of many kinds recently, but there is one critical area where progress has been extremely slow: ordinary commonsense.

BY ERNEST DAVIS AND GARY MARCUS

Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence



Symbolic Knowledge Distillation

From General Language Models to Commonsense Models

— NAACL 2022 —


Peter
West

Chandra
Bhagavatula


Jack
Hessel

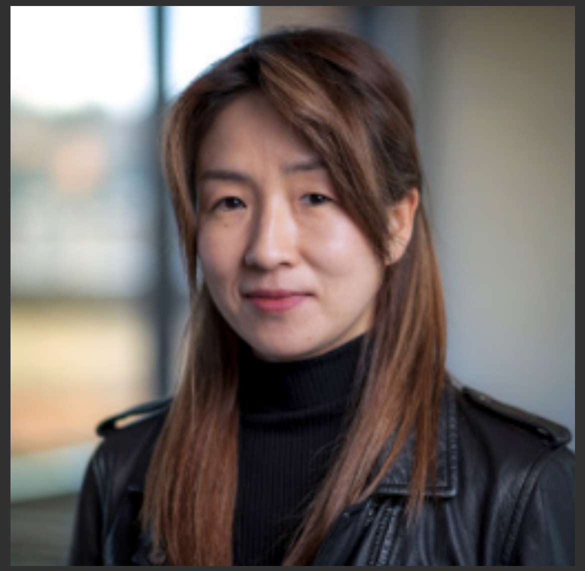

Jena
Hwang


Liwei
Jiang

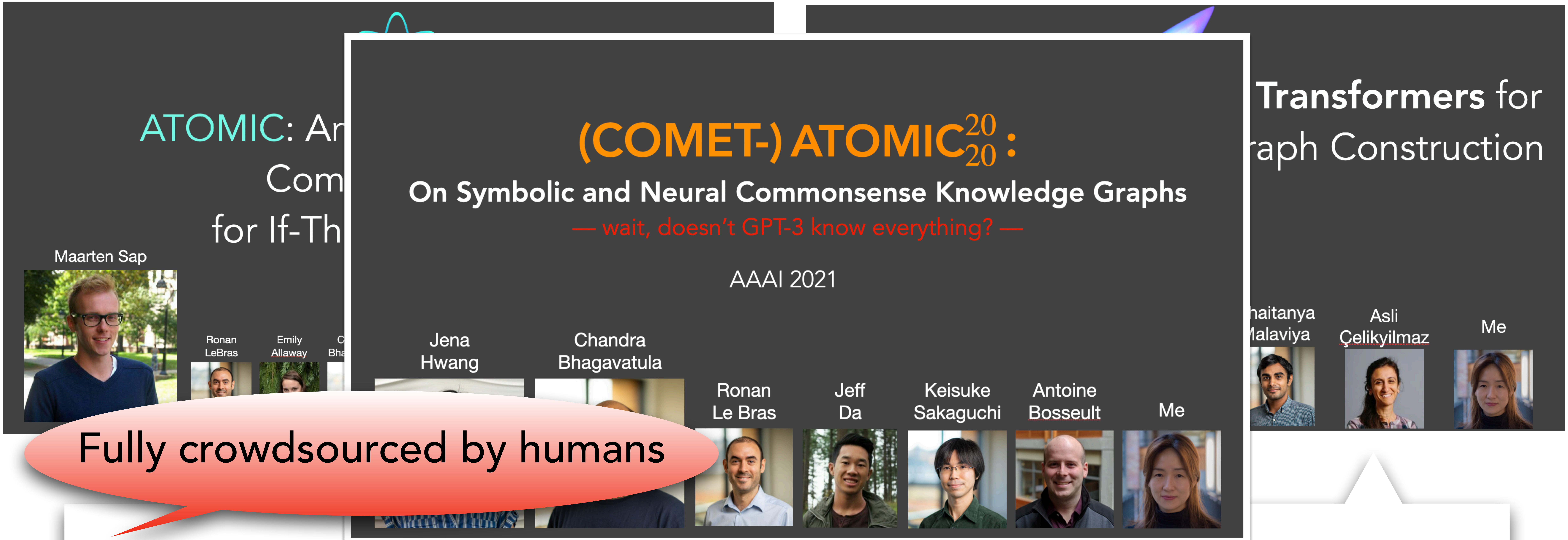

Ronan
Le Bras


Ximing
Lu


Sean
Welleck


Yejin
Choi


Language models != knowledge models



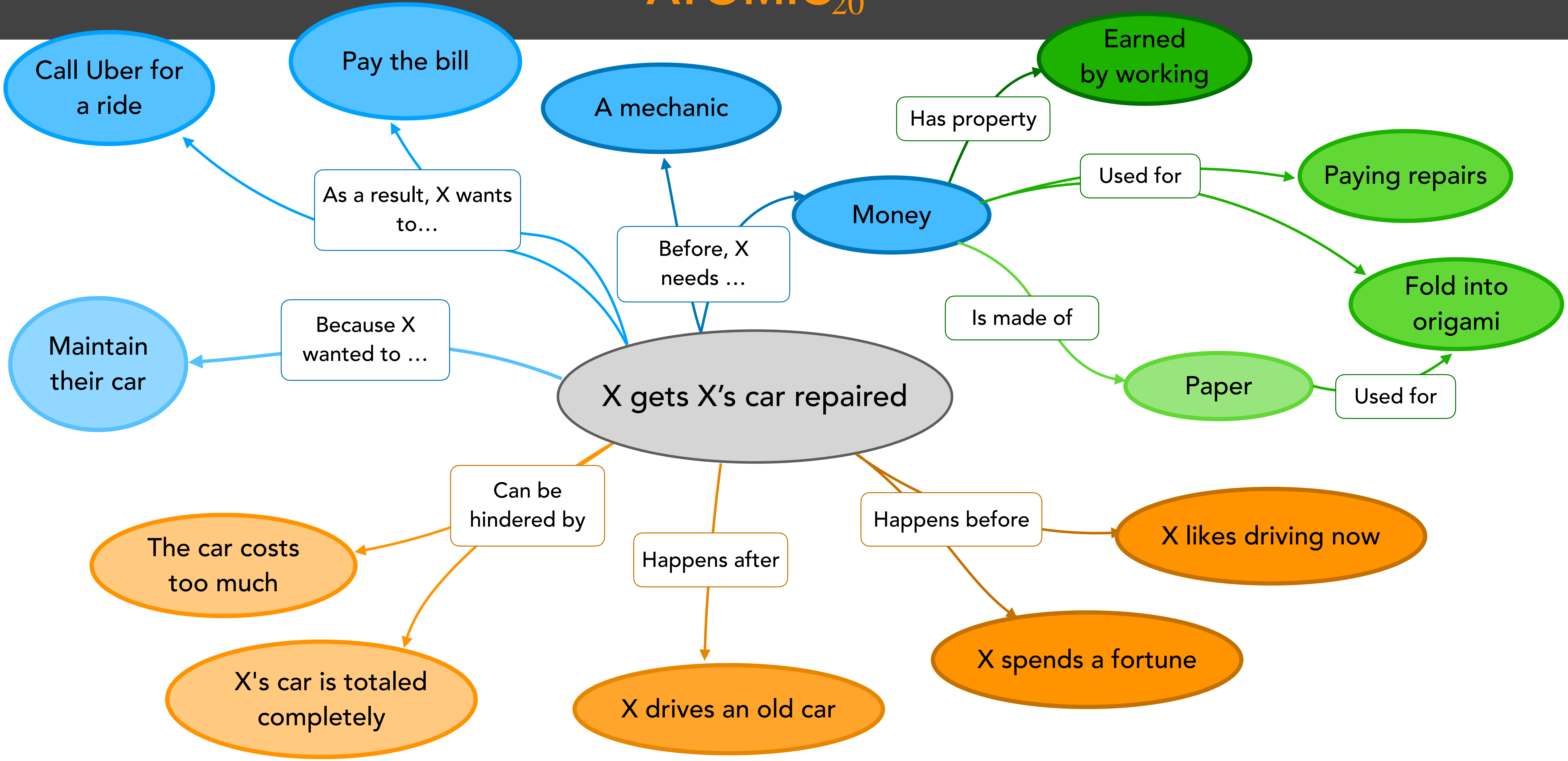
Fully crowdsourced by humans

Symbolic commonsense knowledge graph

Neural commonsense model

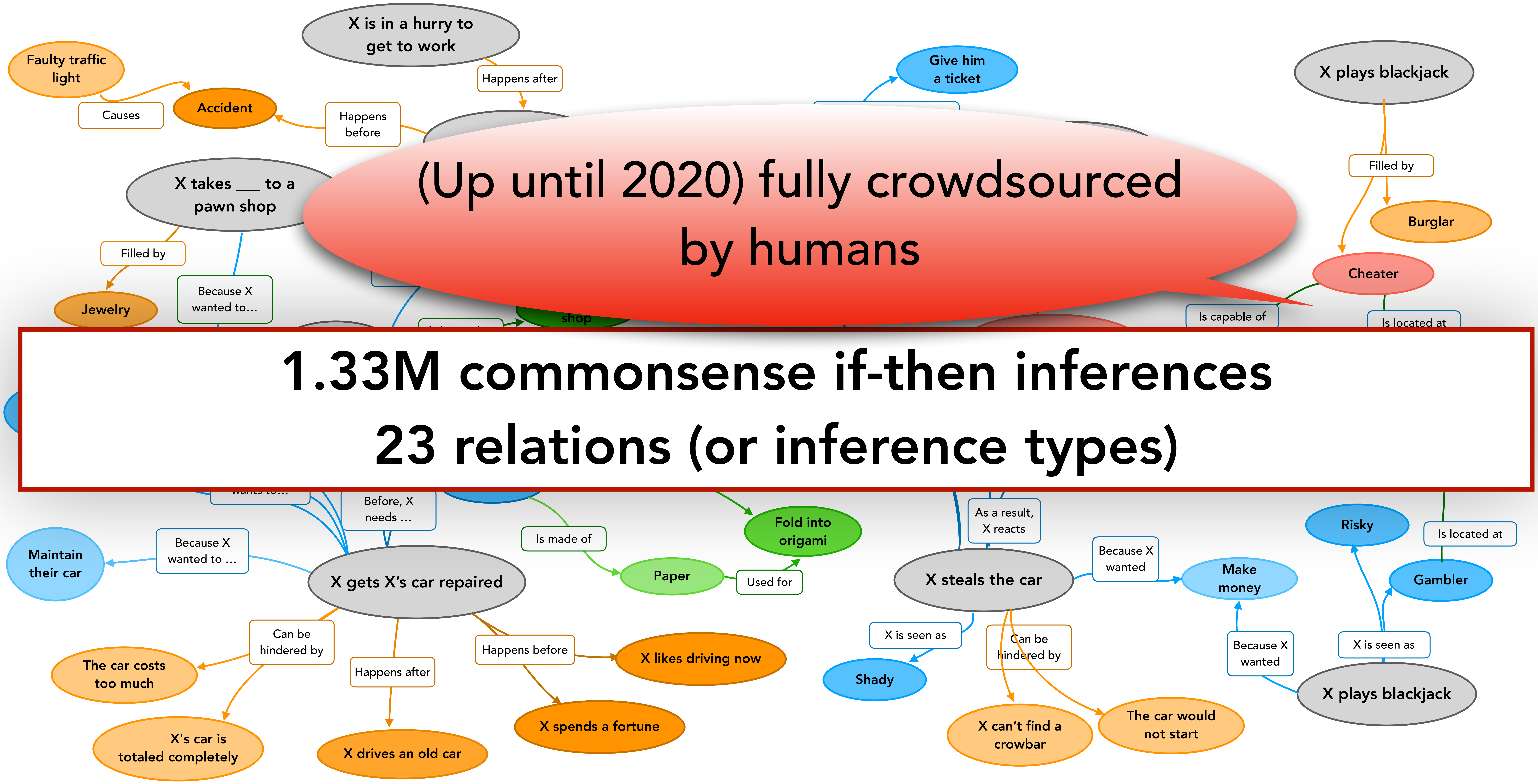

ATOMIC$^{20}_{20}$
Physical-Entity Commonsense
Call Uber for a ride
Pay the bill
A mechanic
Earned by working
Has property
Used for
Paying repairs
ATOMIC
As a result, X wants to…
Money
Before, X needs …
Fold into origami
Is made of
Maintain their car
Because X wanted to …
Paper
Used for
X gets X's car repaired
Social-Interaction Commonsense
Can be hindered by
Happens before
X likes driving now
The car costs too much
Happens after
X's car is totaled completely
X spends a fortune
X drives an old car
Event-Centered Commonsense

ATOMIC$^{20}_{20}$



Call Uber for a ride
Pay the bill
A mechanic
Earned by working
Has property
Used for
Paying repairs
Money
As a result, X wants to…
Before, X needs …
Is made of
Fold into origami
Because X wanted to …
Maintain their car
X gets X's car repaired
Paper
Used for
Can be hindered by
Happens before
X likes driving now
The car costs too much
Happens after
X spends a fortune
X's car is totaled completely
X drives an old car

ATOMIC$^{20}_{20}$

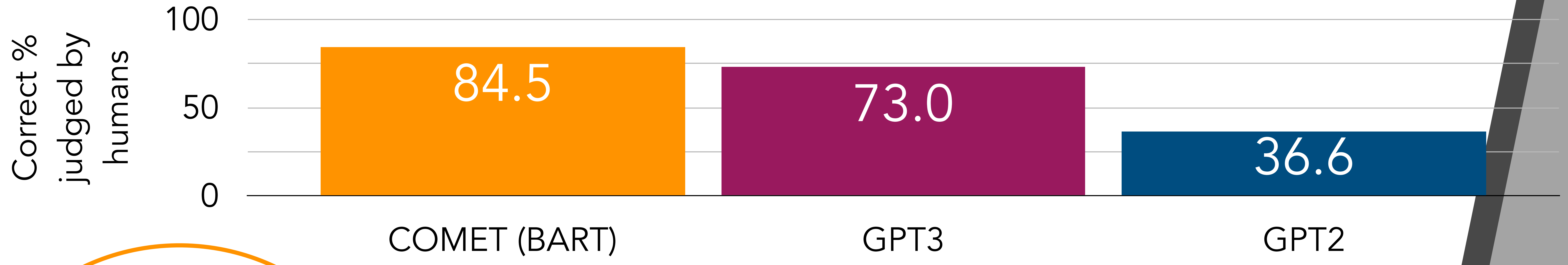


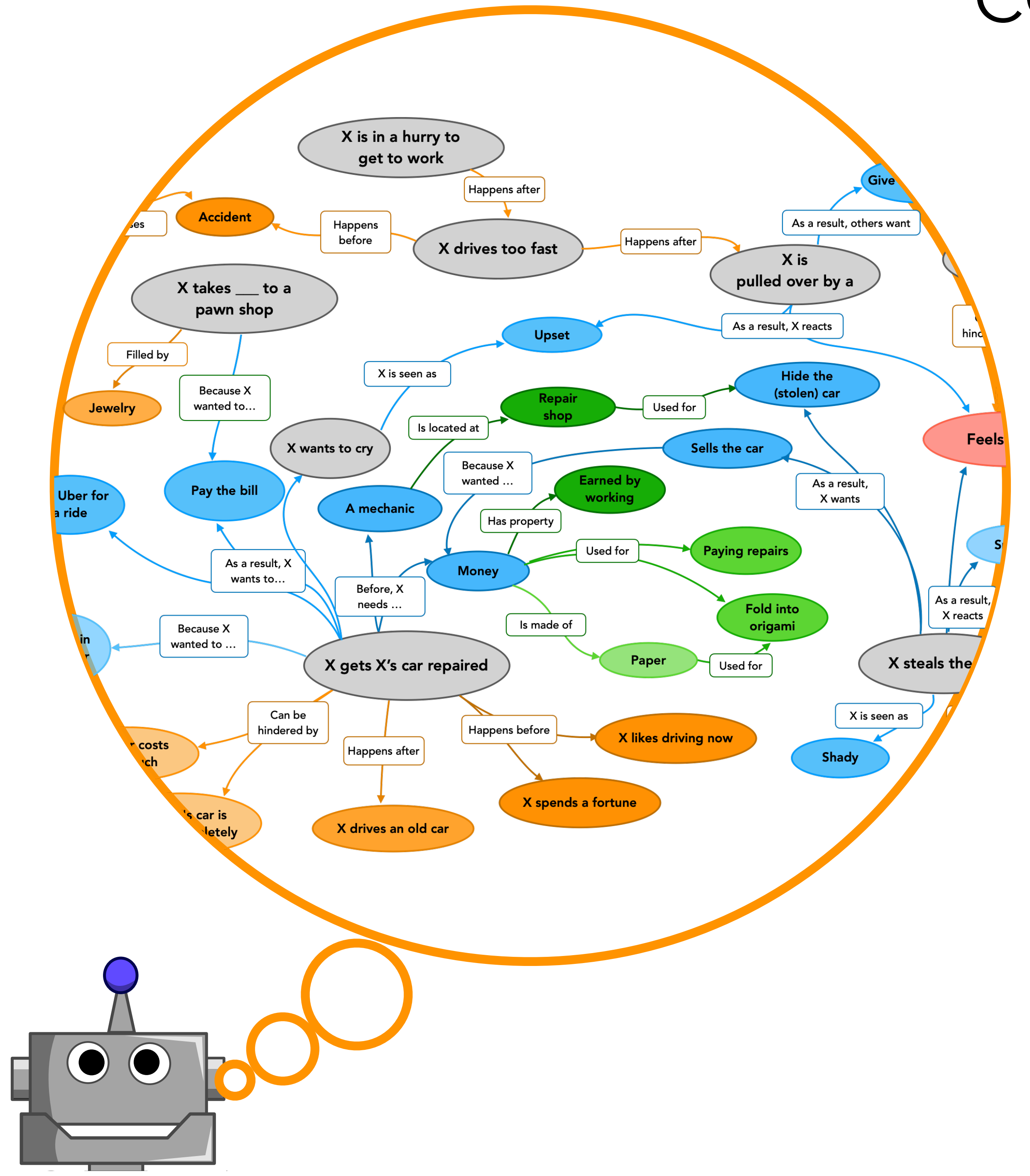
(Up until 2020) fully crowdsourced by humans

1.33M commonsense if-then inferences

23 relations (or inference types)

Knowledge Models | Off-the-shelf Language Models





COMeT (BART): x435 smaller model (~400M parameters), informed by ATOMIC$^{20}_{20}$

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

COMET-ATOMIC 2020

Persona-aware Conversations

Like Hiking? Person-grounded Dialog
(Majumder et al, 2020)
EMNLP '20

Health Counseling Dialogue
(Kearns et al, 2020)
CHI EA '20

COSMIC: Emotion Identification in Conversations
(Ghosal et al, 2020)
EMNLP '20

Figurative Language Understanding

Metaphor Generation with Conceptual Mapping
(Stowe et al, 2021)
ACL '21

MERMAID: Metaphor Generation
(Chakrabarty et al, 2021)
NAACL '21

Interactive Learning Enhancement

Conversation Multi-hop Reasoning through Neural Commonsense
(Forough et al, 2021)
EMNLP '21

Storytelling and Fantasy Gaming

How to Motivate Your Dragon
(Ammanabrolu et al, 2021)
AAAI '21

Commonsense Story Generation
(Guan et al, 2020)
TACL '20

Symbolic Knowledge Distillation

From Neural Language Models to Causal Commonsense Models

New:

ATOMIC-10x

COMET-distill

Peter West

Chandra Bhagavatula

Jack Hessel

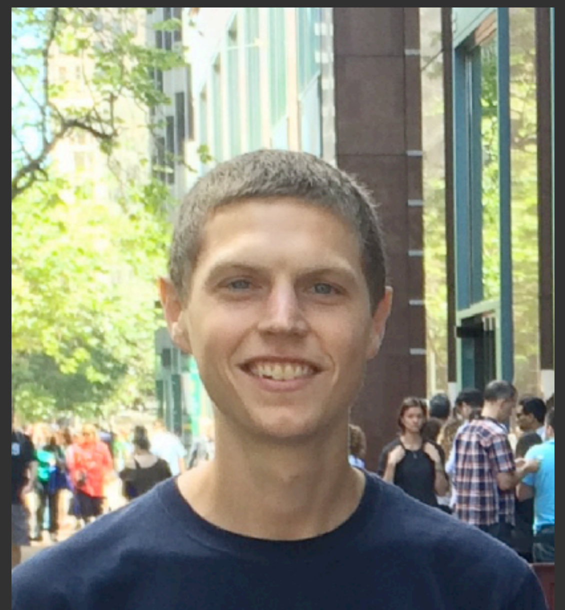
Jena Hwang

Liwei Jiang

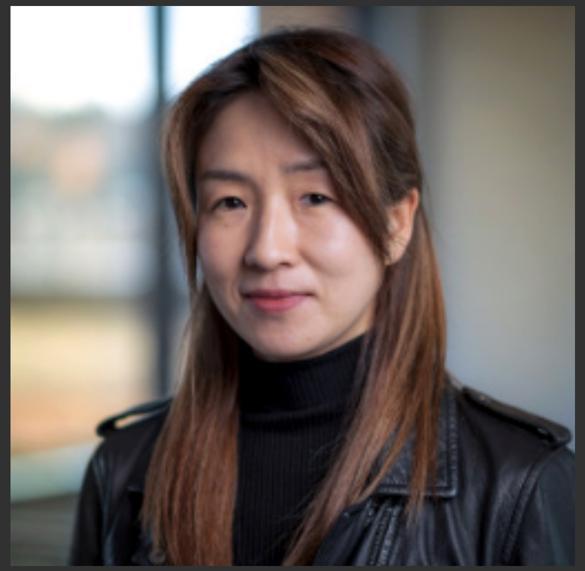
Ronan Le Bras

Ximing Lu

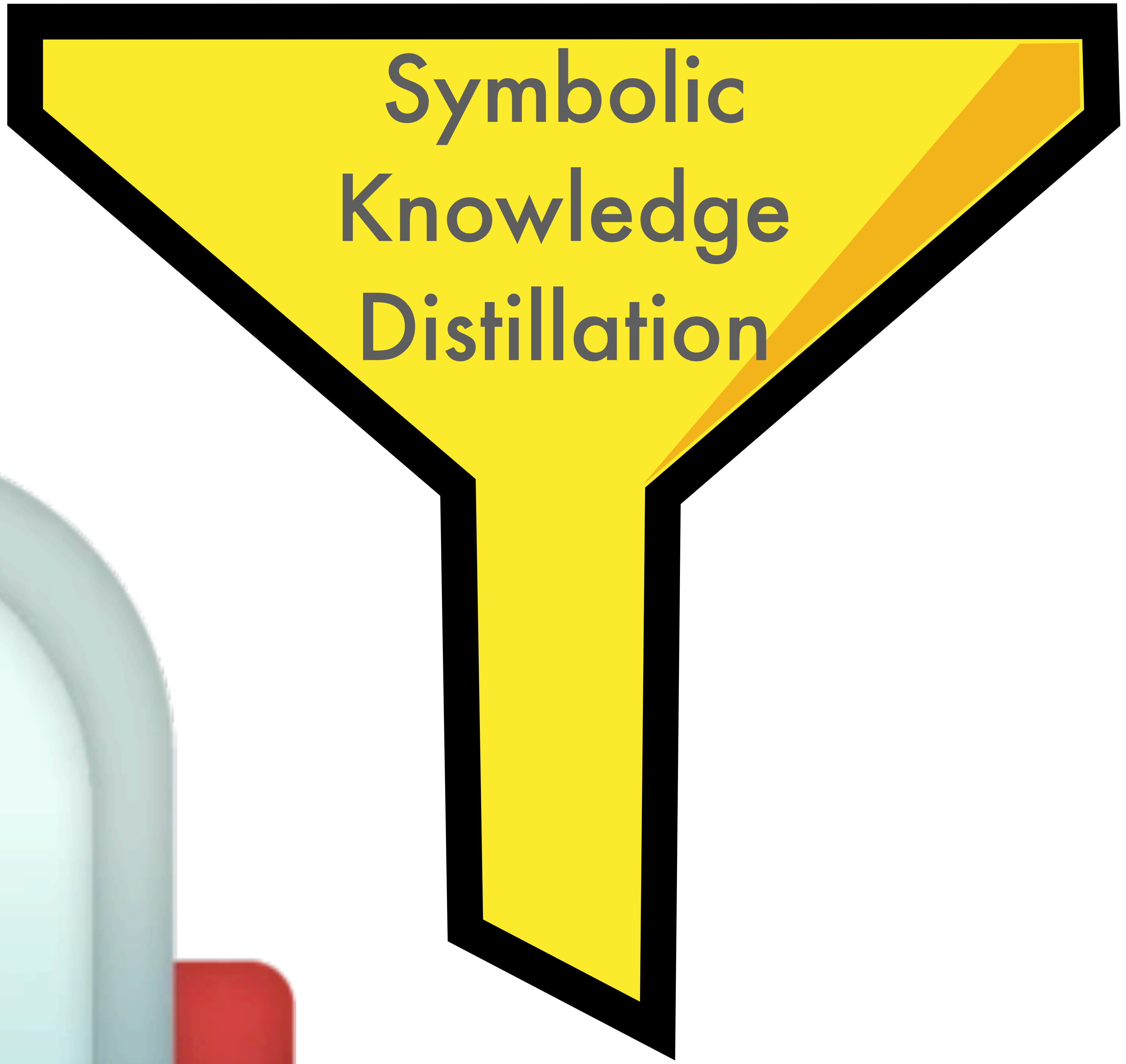
Sean Welleck

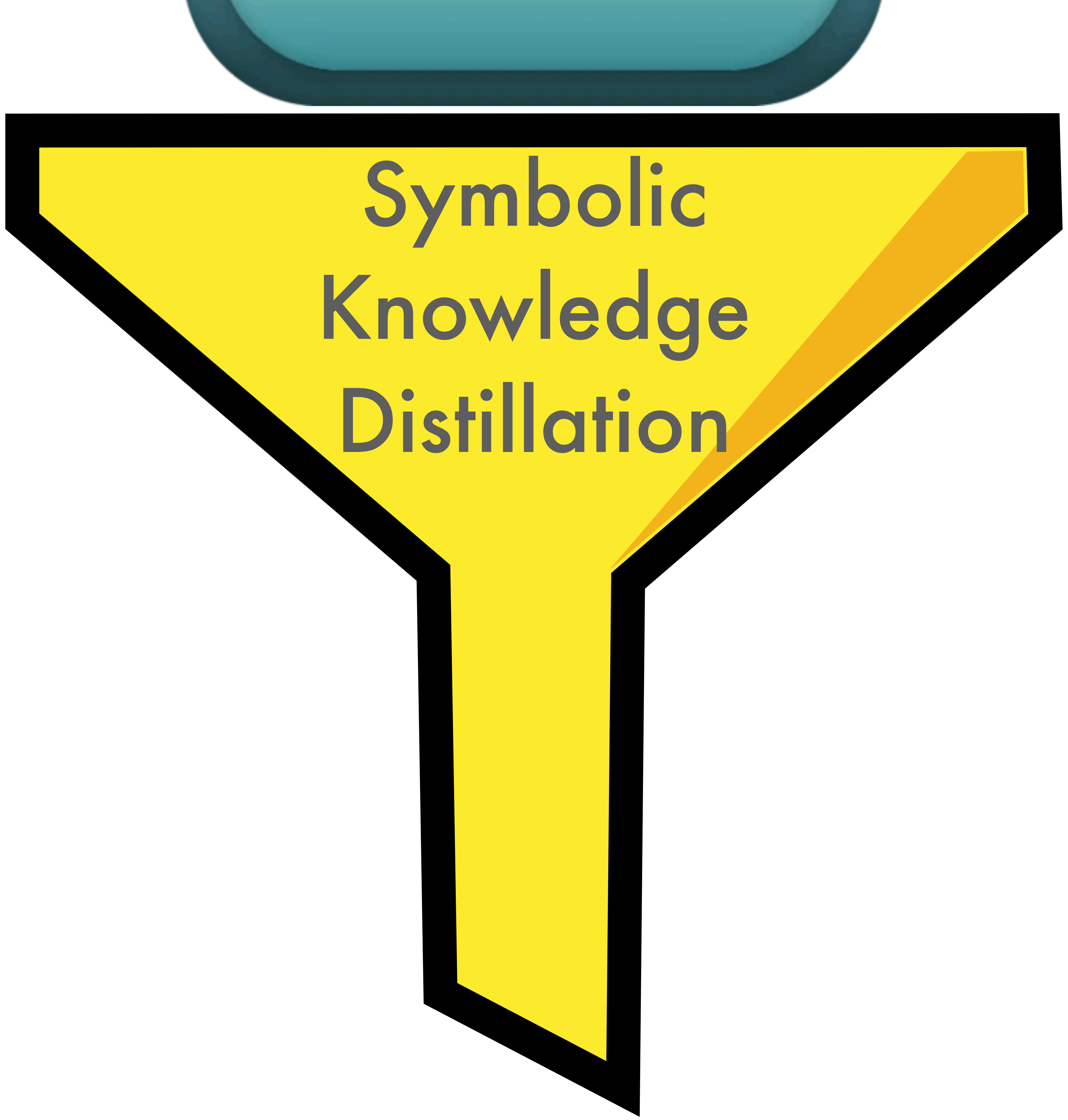


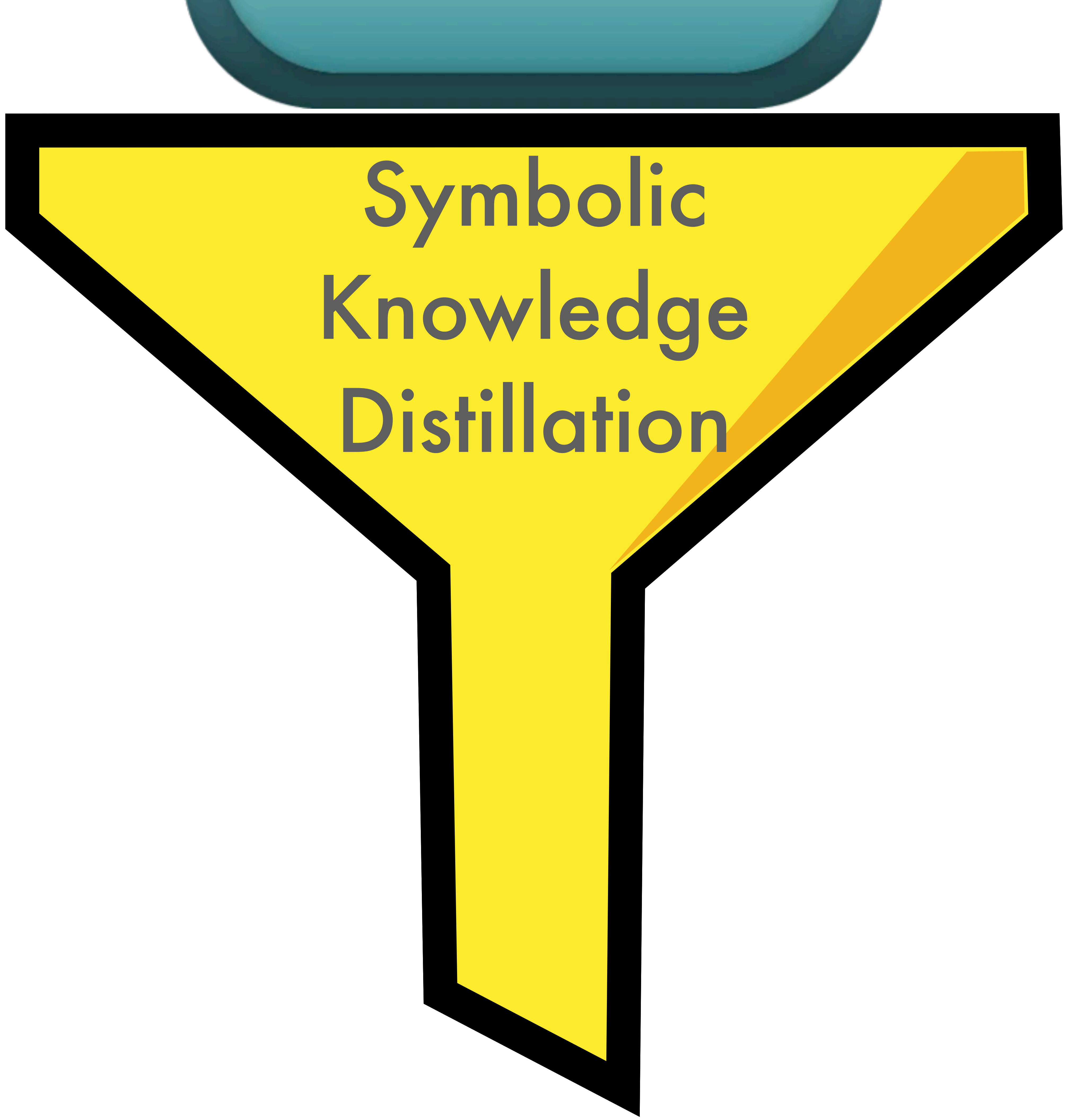
Yejin Choi



GPT-3




Symbolic
Knowledge
Distillation


Symbolic
Knowledge
Distillation

SMALLER

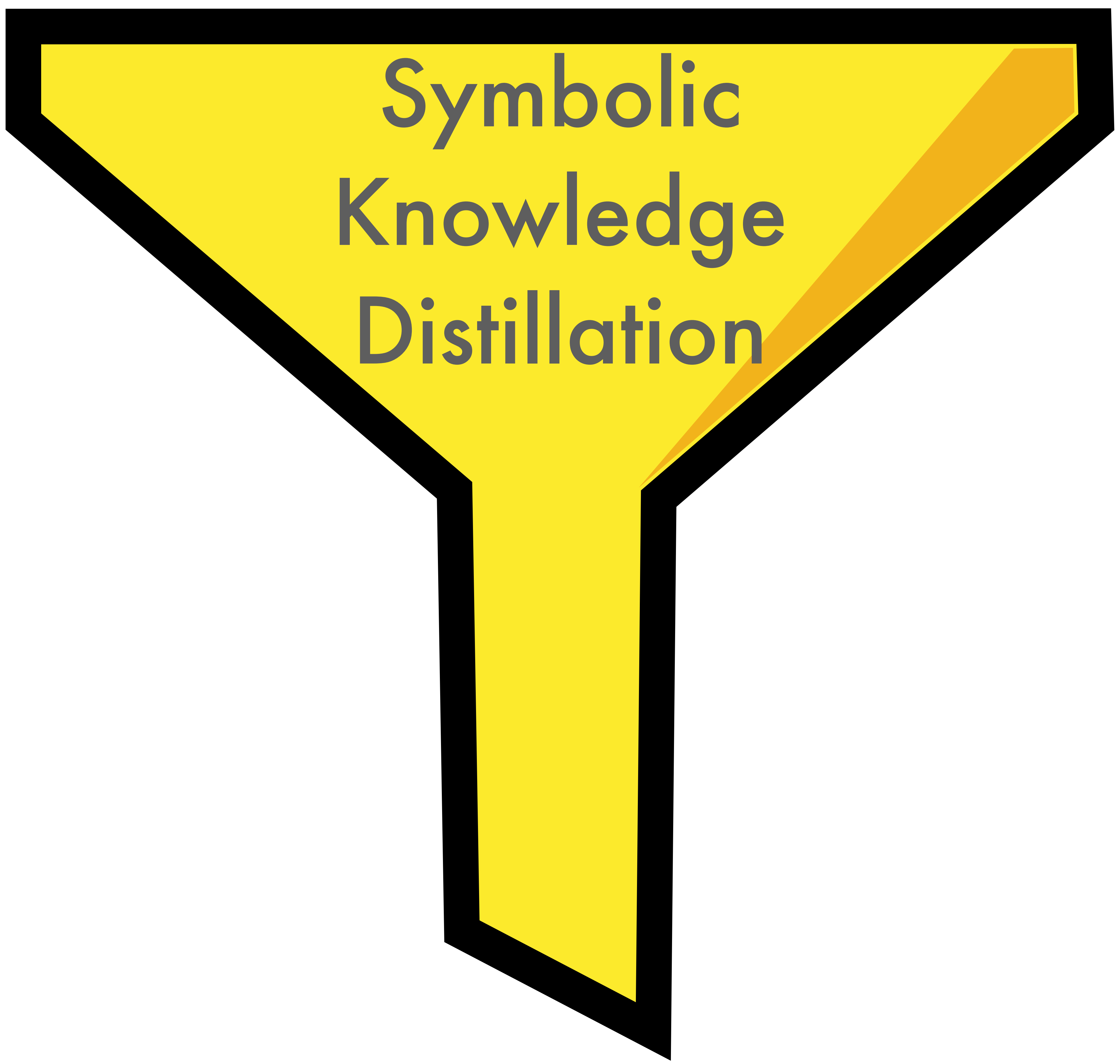
AND

BETTER

🤖

EVEN POSSIBLE

???



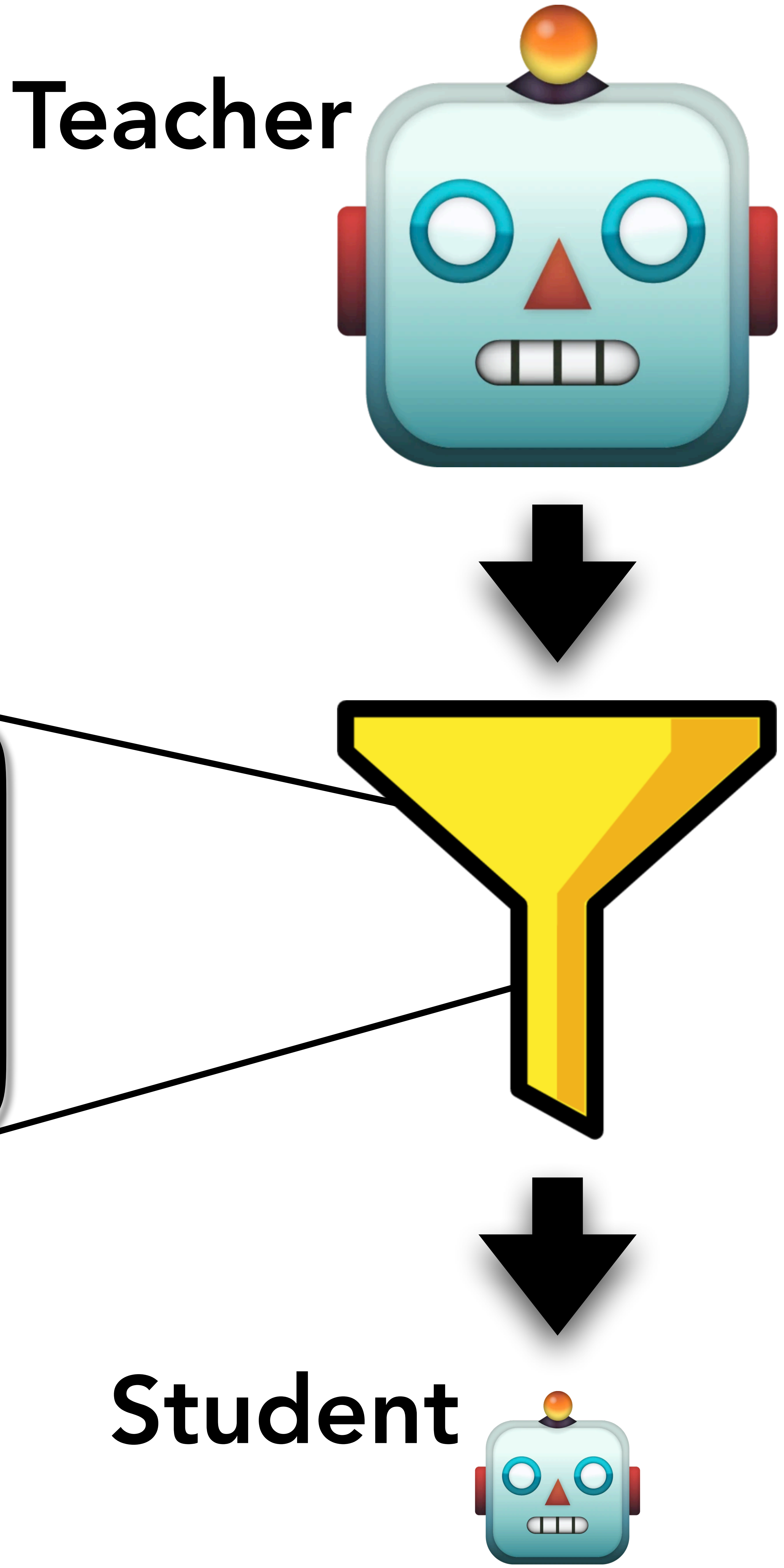
Student Model
smaller & better
Symbolic
Knowledge
Distillation
Knowledge Graph
6.5M high quality examples
Critic
sort good and
bad knowledge

Knowledge Distillation

(Hinton et al. 2015)

$$H(P_t, P_s) = -\sum_{y \in Y} P_t(y) \log P_s(y)$$

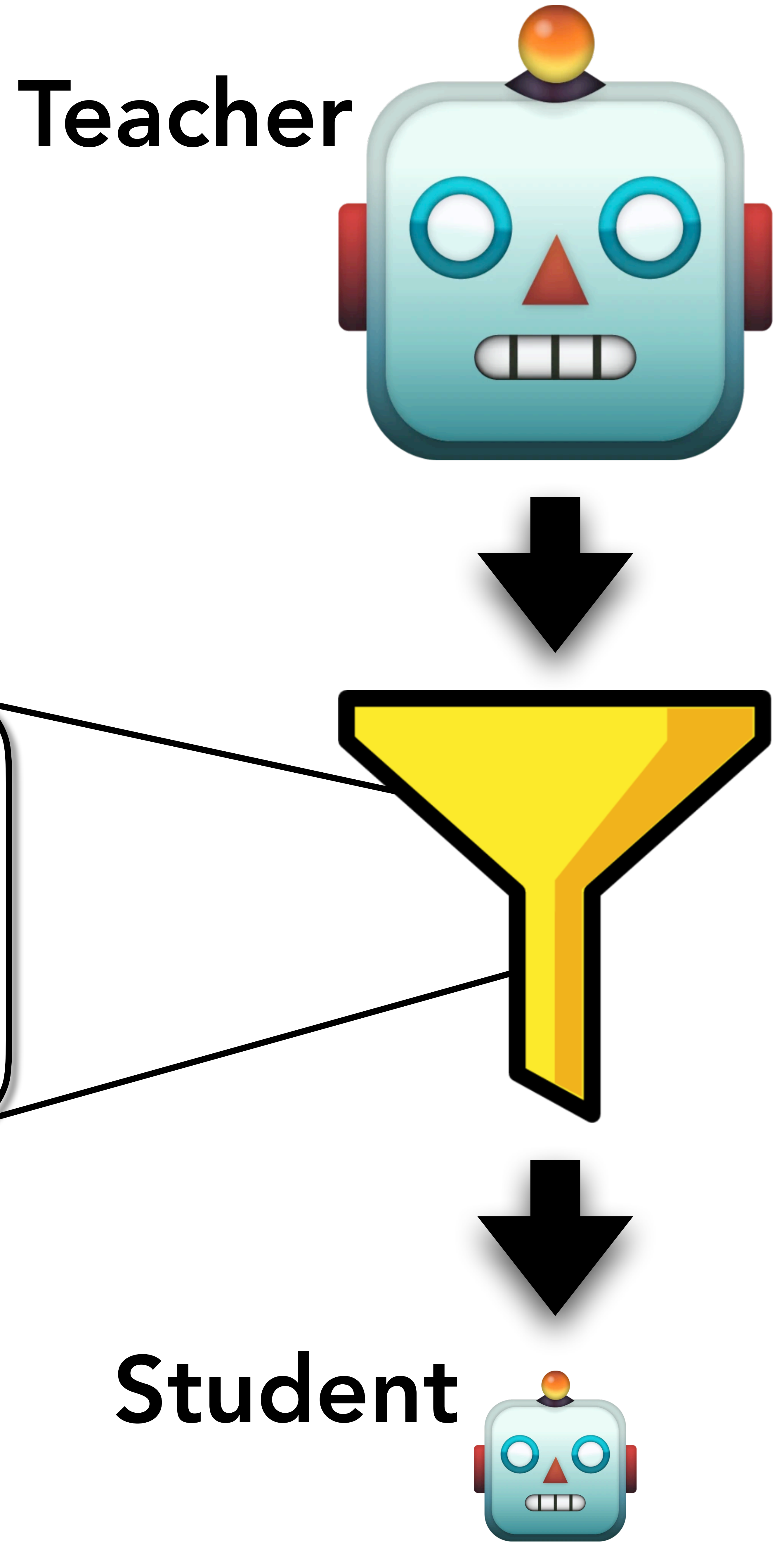
Train student to match teacher probabilities



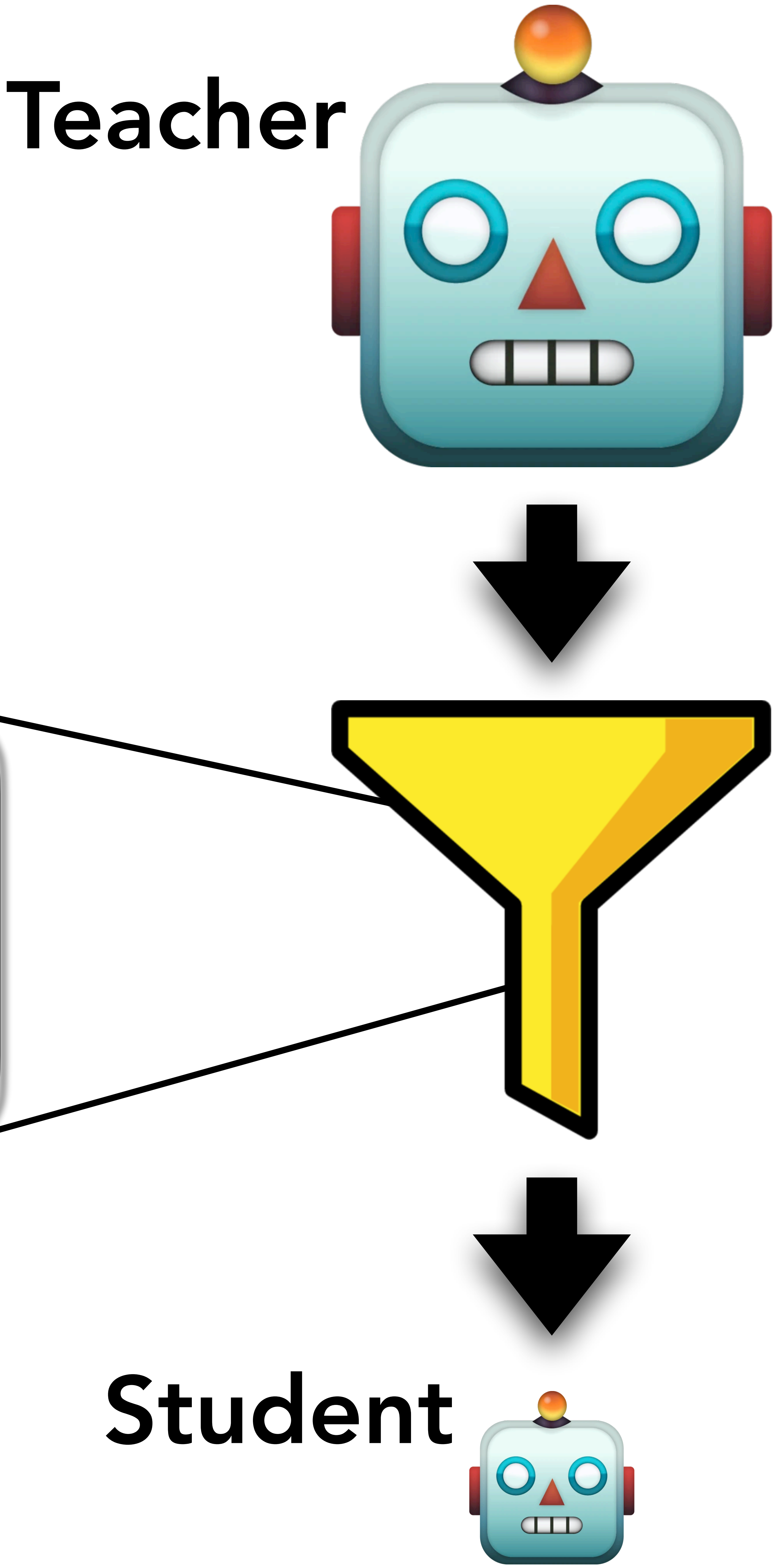
Symbolic
Knowledge Distillation

$$\cancel{H(P_t, P_s) = \sum_{y \in Y} P_t(y) \log P_s(y)}$$

In generation, Y is all strings – intractable!



Symbolic
Knowledge Distillation



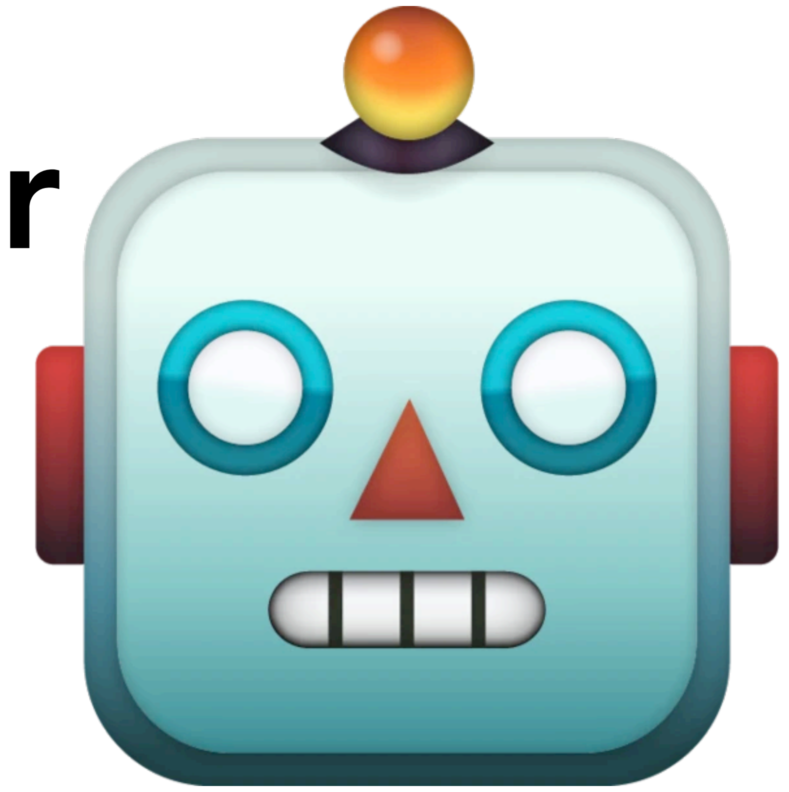
$$H(P_t, P_s) = \underset{y \sim P_t(y)}{\mathbb{E}} [-\log P_s(y)]$$

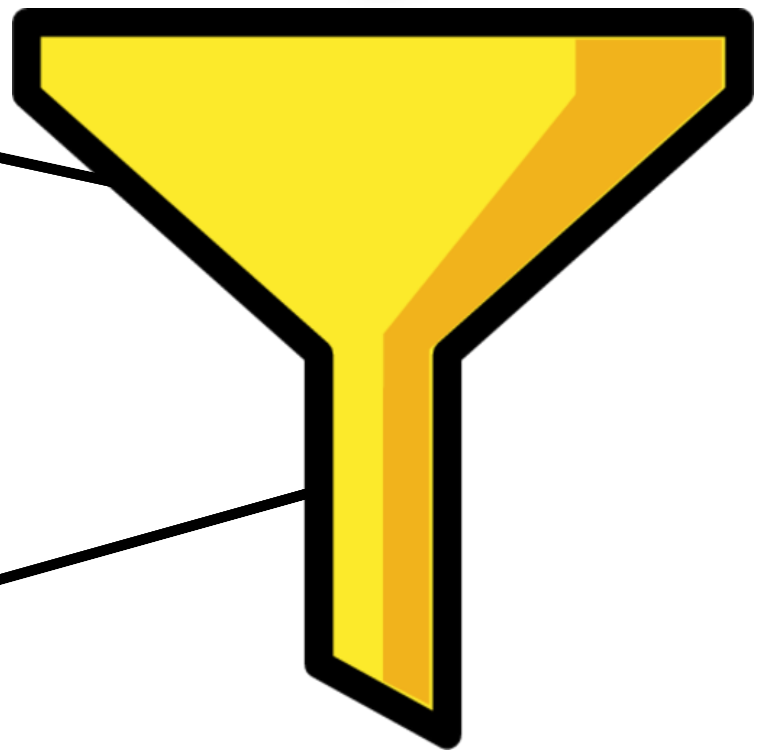
Estimate instead by generating examples!

Natural byproduct is a knowledge graph

Symbolic
Knowledge Distillation

Teacher



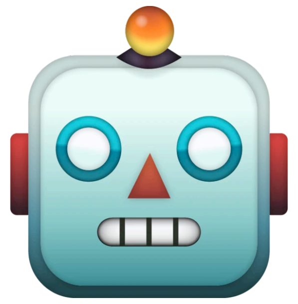


$$H(P_t, P_s) = \mathop{\mathbb{E}}_{y \sim P_t(y)} [-\log P_s(y)]$$

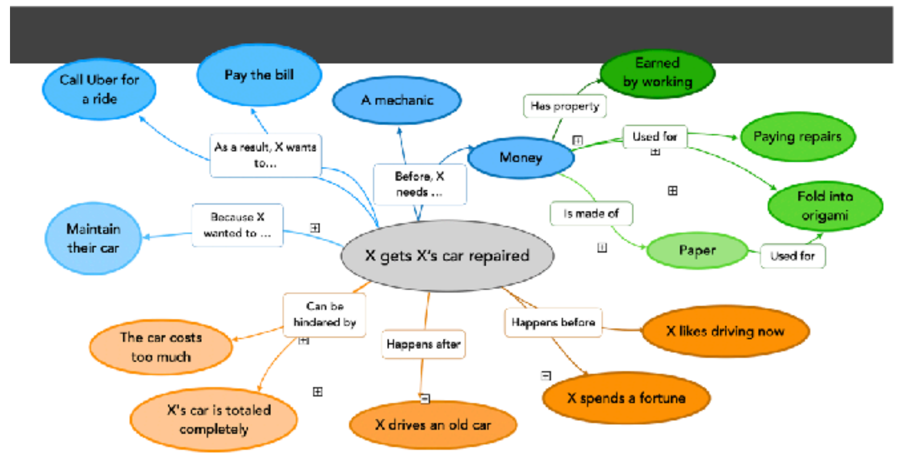
Estimate instead by generating examples!

Natural byproduct is a knowledge graph

Student



KG



Does Symbolic Knowledge Distillation Produce Good knowledge?

Student COMET^distill beats the teacher GPT-3 — smaller & better
Critical teacher results in a better student than human knowledge
Generation Accuracy
100
90
80
70
Loose Teacher
Baseline: Teacher
GPT-3
COMET^distill
GPT-2
Loose Teacher
GPT-3
COMET$^{20}_{20}$
GPT-2
Human
COMET^distill
GPT-2
Critical Teacher
GPT-3
+
Critic


ATOMIC$^{20}_{20}$
VS
ATOMIC^{10X}
Better Value
Higher Accuracy
Larger Scale
More Diverse
100
75
50
0
ATOMIC-10x: a machine-authored KB that wins, for the first time,
over a human-authored KB in all criteria: scale, accuracy, and diversity.
Accuracy
Scale
(10K examples)
Value
(ex/10¢)
Diversity
(10K unique 1-grams)

Commonsense AI: Closing Remarks & Open Research Questions

FAQ these days…

"[NLP/common sense/…] is almost solved by ChatGPT and I have an existential crisis"

Hasty generalization fallacy?

Premise of this talk: scaling laws are real

- Denial is futile
- Confession I: among my all time favorite papers are…
- Confession II: can't do without GPT-3/2, T5, CLIP, Roberta, …

Scaling Laws for Autoregressive Generative Modeling

Tom Henighan[*] Jared Kaplan[*†] Mor Katz[*]

Mark Chen Christopher Hesse Jacob Jackson Heewoo Jun

Tom B. Brown Prafulla Dhariwal Scott Gray Chris Hallacy Benjamin Mann

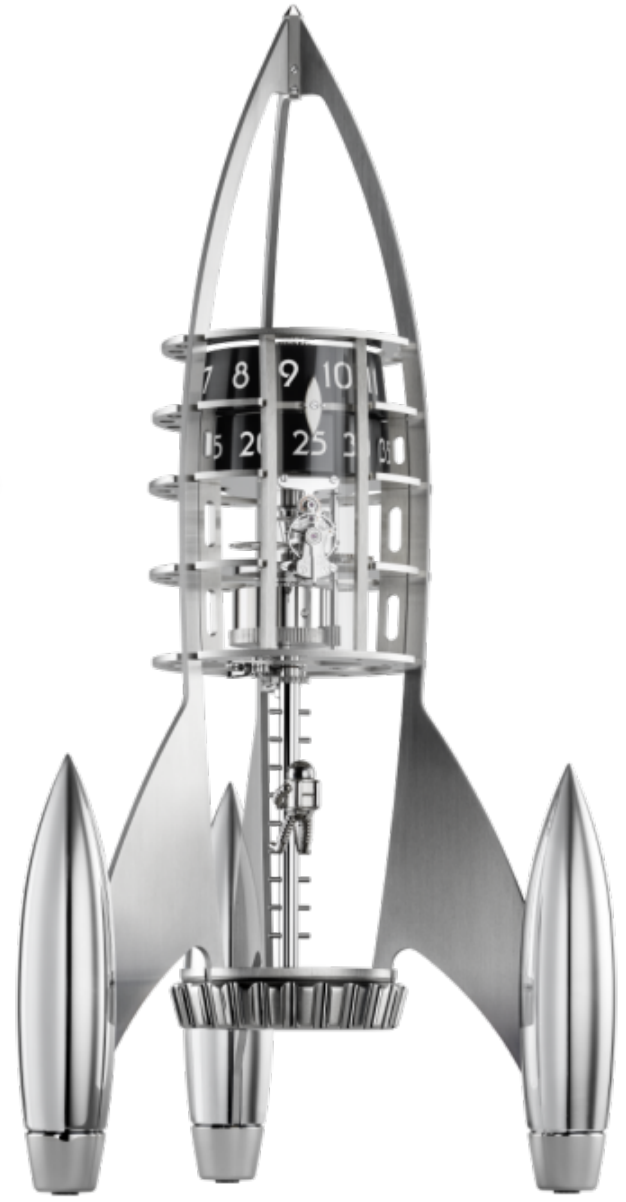
Alec Radford Aditya Ramesh Nick Ryder Daniel M. Ziegler

John Schulman Dario Amodei Sam McCandlish

OpenAI

Premise of this talk: on the other hand…

- Unlikely that we reach AGI by just scaling things up
- Scaling laws explain a **necessary condition** of AGI, not the **sufficient condition**



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

FAQs in the Era of Extreme-Scale Neural Models

"Are you sure we can't reach AGI by just scaling things up? GPT-3 is so magical!"

WSJ Wall Street Journal

Self-Driving Cars Could Be Decades Away, No Matter What Elon Musk Said

Experts aren't sure when, if ever, we'll have truly autonomous vehicles that can drive anywhere without help. First, AI will need to get a lot...

1 month ago

Hasty generalization fallacy?

Language models are “*sometimes*” amazing!

Language ↔ Reasoning

If you travel west far enough from the west coast, you will reach the east coast?

The world is round, so you will reach the east coast. Therefore, the answer is ***True***.

Hmm...

If you travel west far enough from the west coast, can you reach the east coast?

No, if you travel west from the west coast of the United States, you will eventually reach Asia, but you will not reach the east coast of the United States. The east and west coasts are located on opposite sides of the country and are separated by the majority of the landmass of the United States.

If you travel west far enough from the west coast, can you reach the east coast?

No, if you travel west from the west coast of the United States, you will eventually reach Asia, but you will not reach the east coast of the United States. The east and west coasts are located on opposite sides of the country and are separated by the majority of the landmass of the United States.

If you travel west far enough from the west coast, can you eventually reach the east coast?

No, you cannot reach the east coast by traveling west from the west coast. The earth is a sphere and there is no endpoint in any single direction of travel. If you were to travel west from the west coast, you would eventually circumnavigate the earth and end up back where you started.

Observations & Speculations

- Humans tend to assume that if AI can do seemingly **hard** tasks (e.g., chess, protein folding, shockingly impressive stylistic revision) than it must be that AI can also do seemingly **easy** tasks (e.g., commonsense)

- The empirical reality suggests otherwise

- As is, we simply do not know the depth and the breadth of human intelligence compared to what AI can and cannot achieve primarily through scale

- My speculation is that large-scale models provide “proofs of existence of solutions”, suggesting that there likely be solutions at smaller scales

- Such solutions will likely require (entirely) novel ways to integrate symbolic reasoning to neural representation


Symbolic Knowledge Distillation
From Neural Language Models to Causal Commonsense Models
New:
ATOMIC-10x
COMET-distill
Peter West
Chandra Bhagavatula
Jack Hessel
Jena Hwang
Liwei Jiang
Ronan Le Bras
Ximing Lu
Sean Welleck
Yejin Choi
Maieutic Prompting:
Logically Consistent Reasoning with Recursive Explan
— https://arxiv.org/abs/2205.11822 —
Jaehun Jung
Lianhui Qin
Sean Welleck
Faeze Brahman
Chandra Bhagavatula
Ronan Le Bras
Language
Knowledge
Reasoning
Dædalus
AI & Society
Dædalus
AI & Society
Dædalus
AI & Society
"The Curious Case of Commonsense Intelligence"
https://www.amacad.org/publication/curious-case-commonsense-intelligence
PAUL G. ALLEN SCHOOL
W
— 🏆 Best Method Paper Award at NAACL 2022 🏆 —
NeuroLogic A*esque
Constrained Text Generation with Lookahead Heuristic
Ximing Lu
Sean Welleck
Peter West
Liwei Jiang
Lianhui Qin
Youngjae Yu


Thanks!