

Reliable text generation through graph search

CSE 373

Motivation

Modeling and generating language

- Language encapsulates ideas.
- Factual knowledge
 - *Molly Seidel won the medal in the 2020 Olympic marathon.*
- State of the art **language model:** Today's lecture

- **I am a highly intelligent question answering bot.**

Q: Who was president of the United States in 1955?

A: Dwight D. Eisenhower was president of the United States in 1955.

Q: Molly Seidel won which medal in the 2020 Olympic marathon?

A: Molly Seidel won a **bronze medal in the 2020 Olympic marathon.**

Motivation

Modeling and generating language

- Language encapsulates ideas.
- Common sense

- *I tipped the bottle. As a result,*

- State of the art GPT-3 language model:

- **I will continue your sentence based on my common-sense understanding of the world:**

- I tipped the bottle. As a result, the drink spilled out.**

Motivation

Modeling and generating language

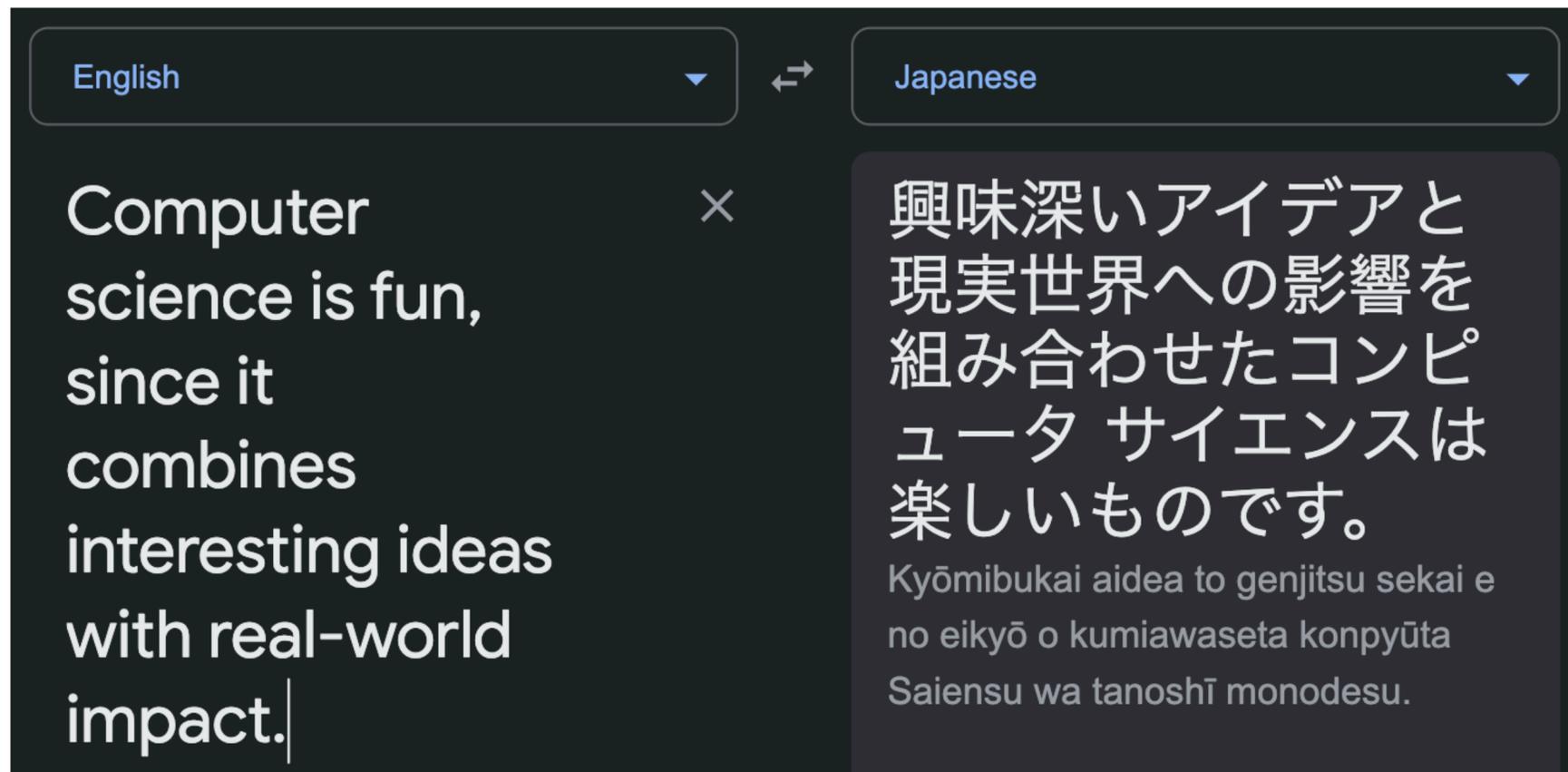
- **Generating language is useful.**
- Dialogue
 - You: What have you been up to?
Friend: Watching old movies.
You: Did you watch anything interesting?
Friend: Yes, I watched The Omen and Troy.

Motivation

Modeling and generating language

- **Generating language is useful.**
- Machine translation

-



The screenshot shows a machine translation interface with two panels. The left panel is labeled 'English' and contains the text: 'Computer science is fun, since it combines interesting ideas with real-world impact.' The right panel is labeled 'Japanese' and contains the translated text: '興味深いアイデアと現実世界への影響を組み合わせたコンピュータサイエンスは楽しいものです。' Below the Japanese text is a phonetic transcription: 'Kyōmibukai aidea to genjitsu sekai e no eikyō o kumiawaseta konpyūta Saiensu wa tanoshī monodesu.'

Motivation

Modeling and generating language

- **Generating language is useful.**
- Programming assistants

```
sentiments.ts  write_sql.go  parse_expenses.py  addresses.rb

1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch(`http://text-processing.com/api/sentiment/`, {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    },
14  });
15  const json = await response.json();
16  return json.label === "pos";
17 }
```

Copilot

Your AI pair programmer

GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in real-time, right from your editor.

Today's lecture

A common *language modeling* recipe underlies all of these applications.

Open-Ended Generation

Build next-gen apps with OpenAI's powerful models.

OpenAI's API provides access to GPT-3, which performs a wide variety of natural language tasks, and Codex, which translates natural language to code.

Long-form QA

How has technological growth increased so exponentially in the last 50 years?

< 175B best-of-64 >

There are many explanations for the exponential growth in technology in the last century. One explanation is that the pace of technological progress speeds up exponentially over time because of a common force driving it forward^[3]. Another explanation is that each new generation of technology stands on the shoulders of its predecessors, allowing for improvements that lead to the next generation of even better

Machine Translation

Google Translate

Text Documents Websites

DETECT LANGUAGE ENGLISH SPANISH ENGLISH SPANISH ARABIC

Translation

0 / 5,000

Program Synthesis

Your AI pair programmer

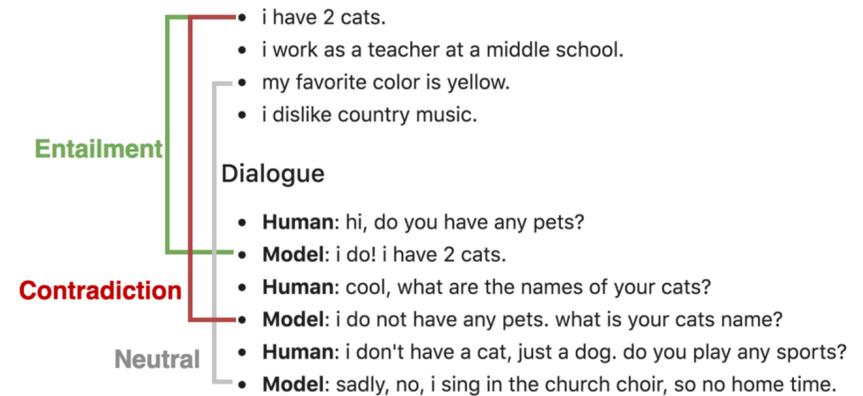
With GitHub Copilot, get suggestions for whole lines or entire functions right inside your editor.

Sign up >

```
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch("http://text-processing.com/api/sentiment/", {
9     method: "POST",
10    body: "text=${text}",
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    }
14  });
15  const json = await response.json();
16  return json.label === "pos";
17 }
```

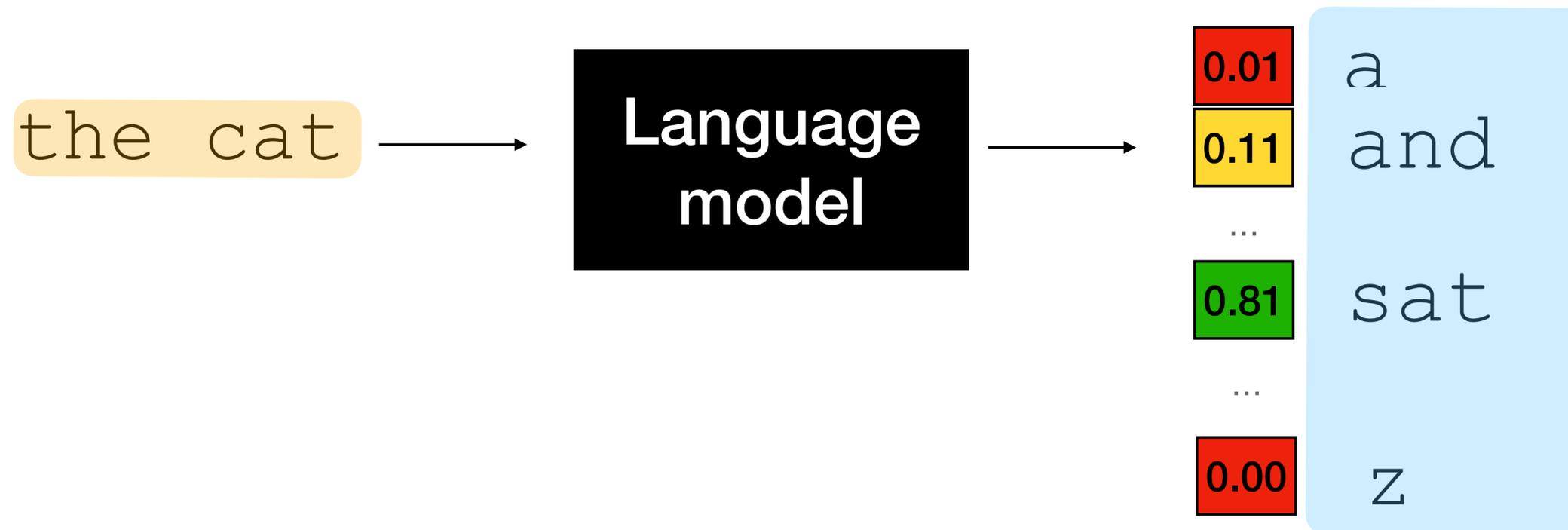
GitHub Copilot

Dialogue



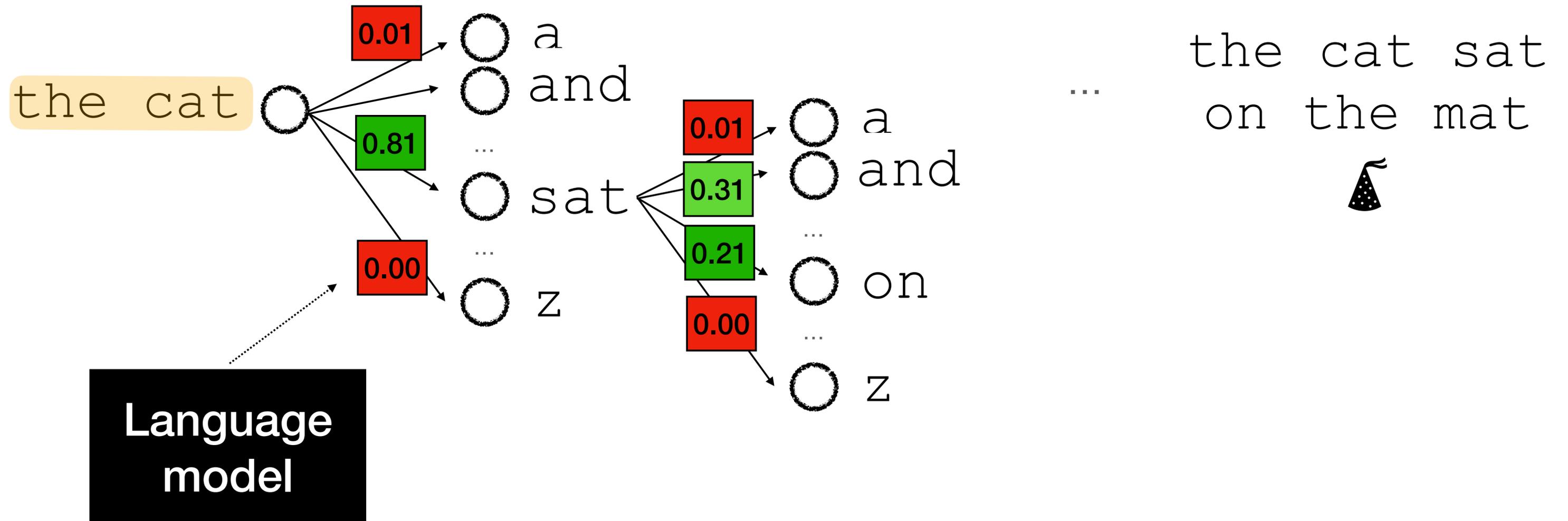
What is a language model?

- Given **previous words**, scores **next-words**



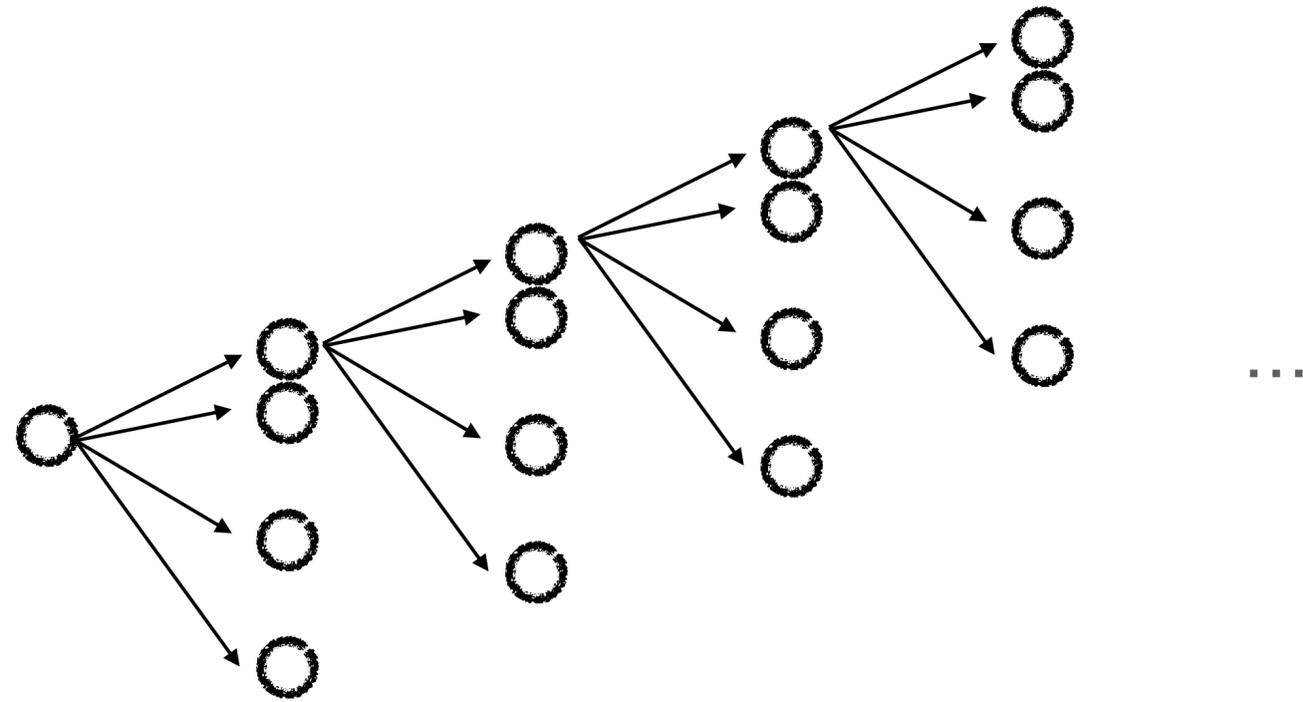
How do we generate text?

- Graph search!



What if we want the highest-scoring text?

興味深いアイデアと
現実世界への影響を
組み合わせたコンピ
ュータサイエンスは
楽しいものです。



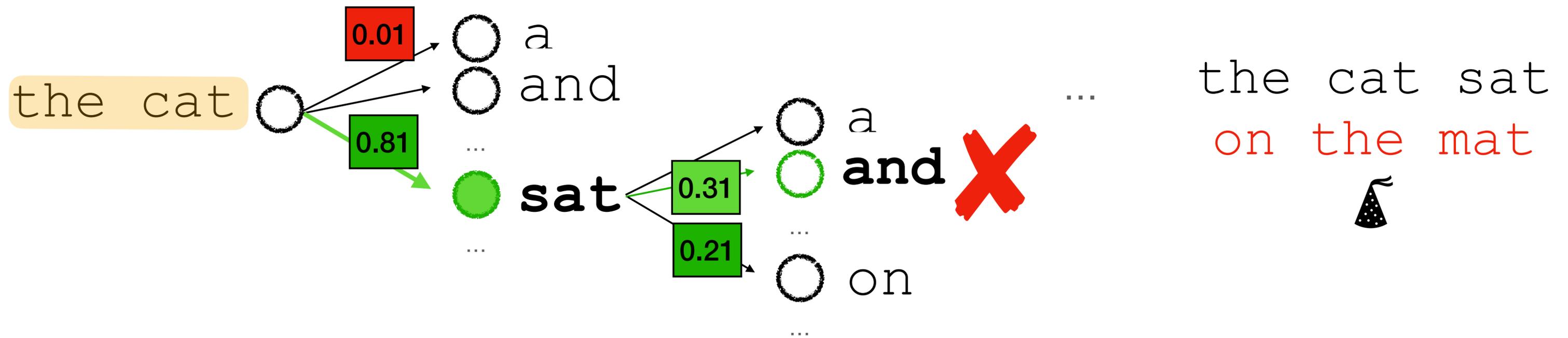
Computer
science is fun,
since it
combines
interesting ideas
with real-world
impact.

$O(\text{num words}^{\text{sequence length}})$

- Scoring all edges with the neural network & selecting maximum is infeasible

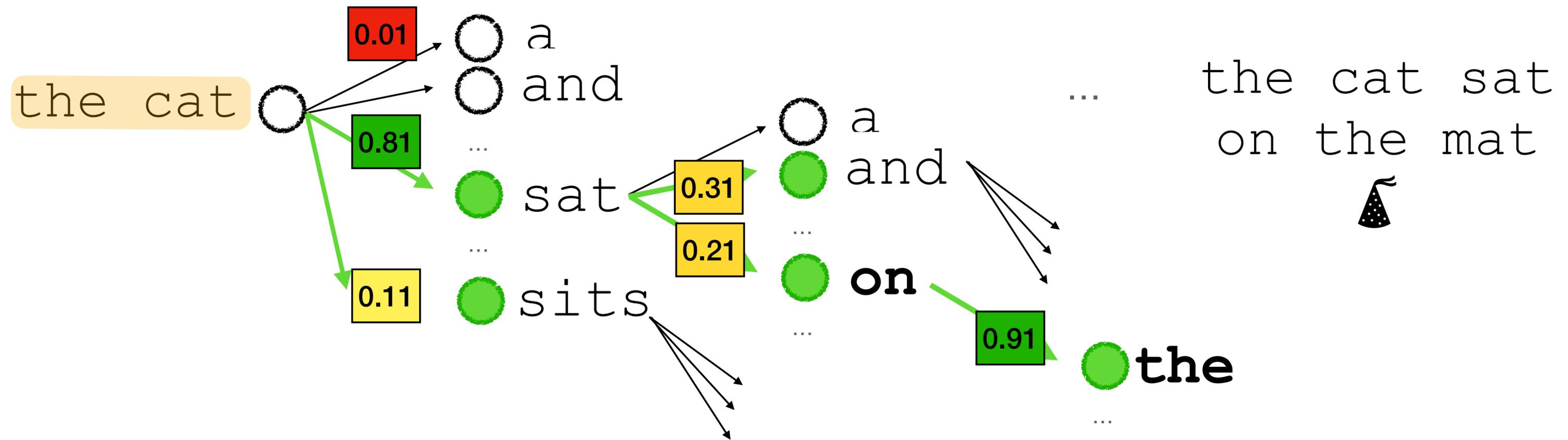
Greedy search

- Choose the highest scoring token at each step

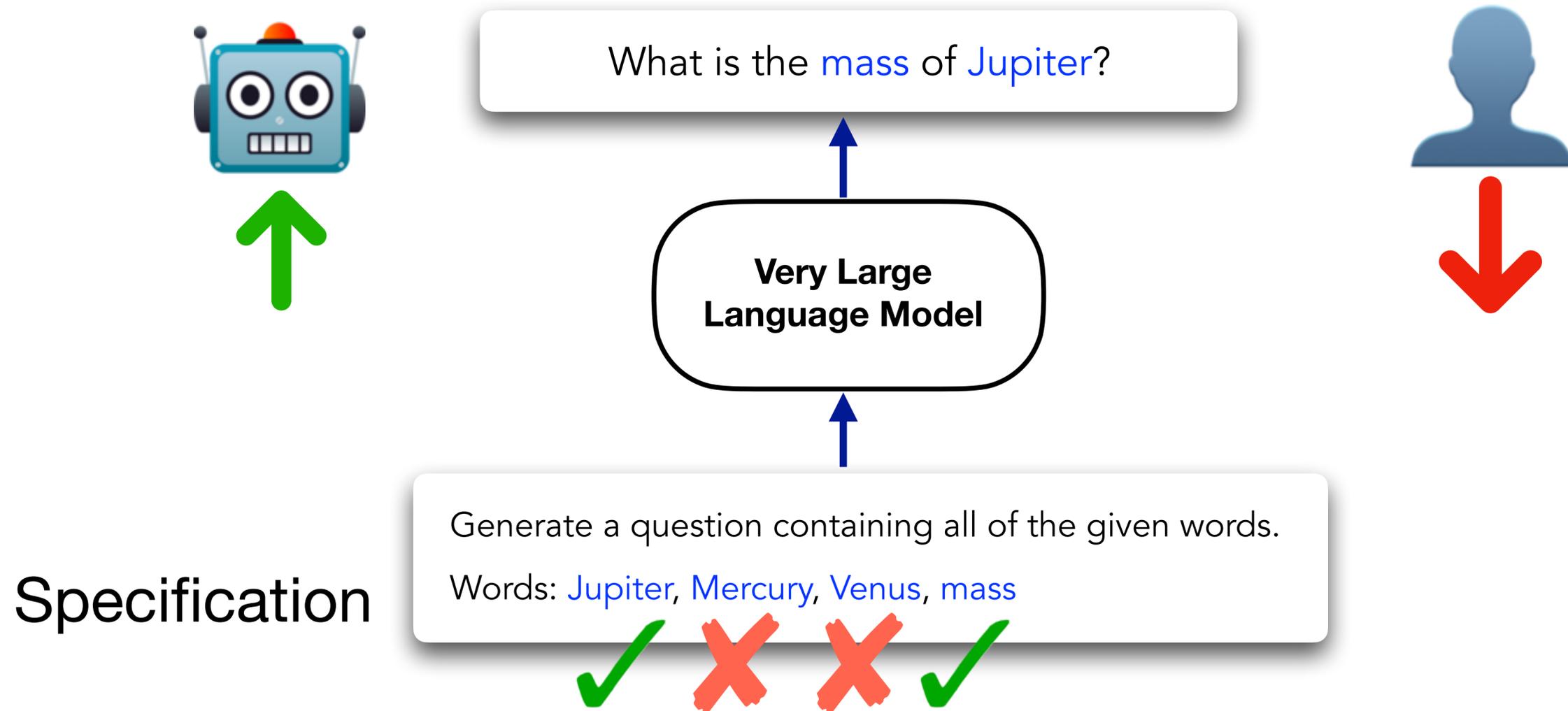


Beam search

- Expand a “beam” of paths at each step

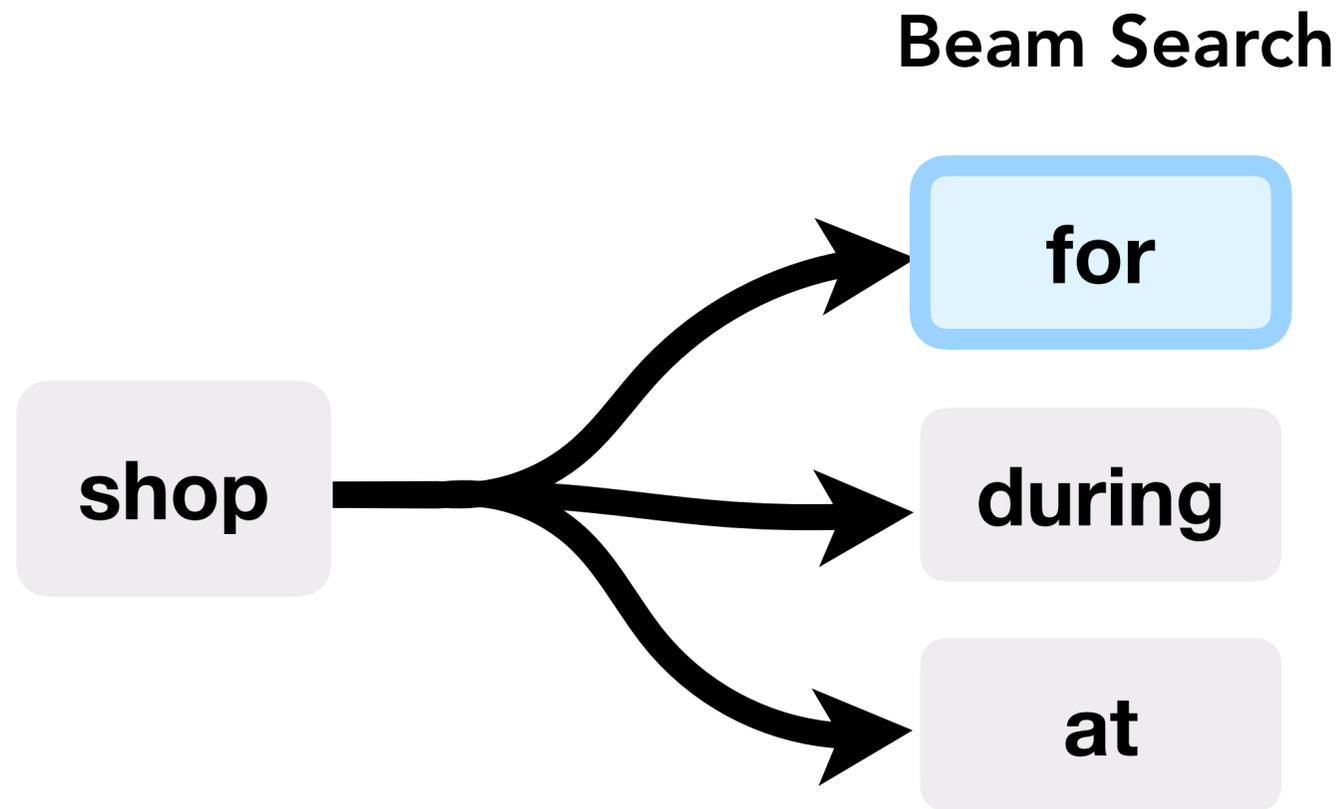


Constrained search



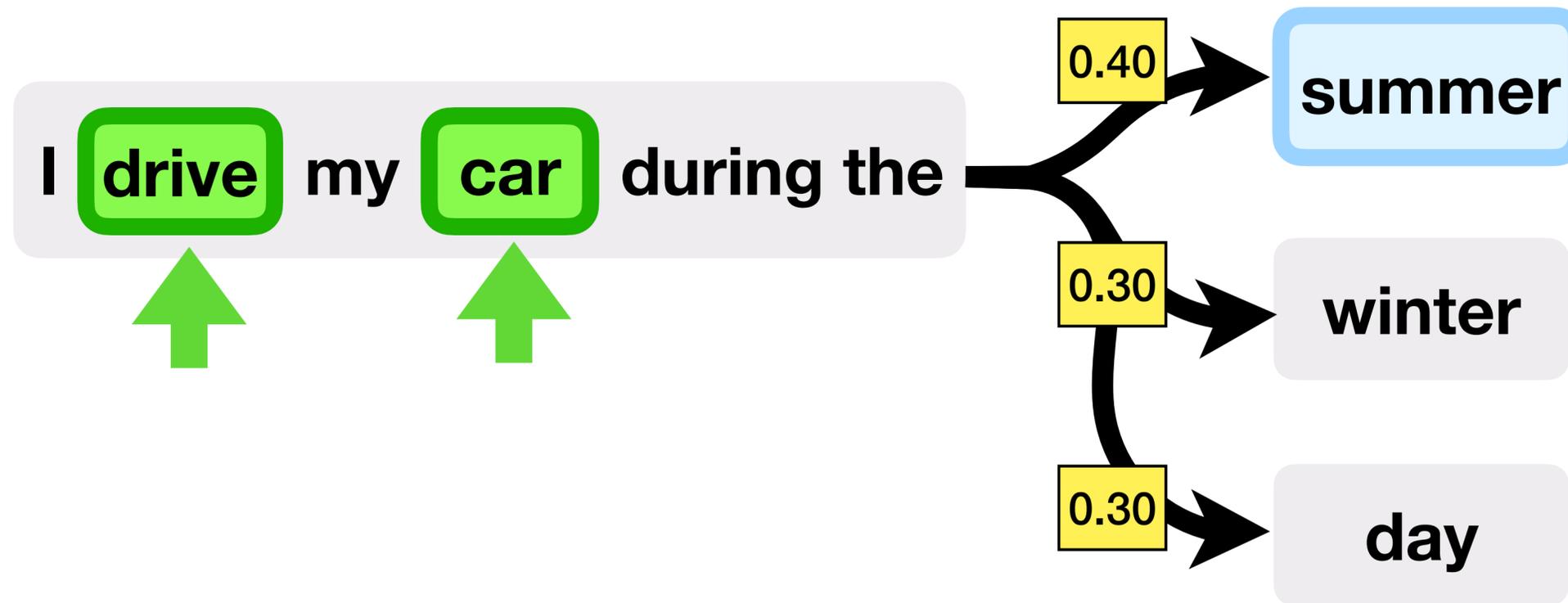
Beam search : no constraints!

Write a sentence with: car \wedge drive \wedge snow



Left-to-right search : myopic!

Write a sentence with: **car** ^ **drive** ^ snow



since it's hot outside.

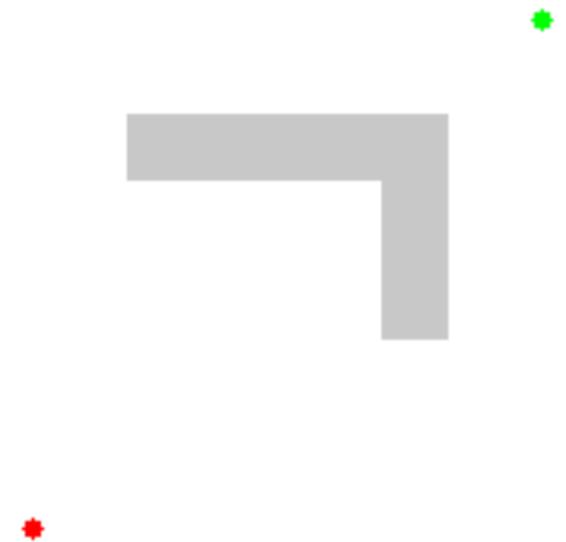
to avoid the snow.

A* search

- A* Search: best-first search with future heuristics

- $f(n) = \underbrace{s(n)} + \underbrace{h(n)}$

- $s(n)$: score to reach node n
- $h(n)$: estimated cost from n to goal.



A* search

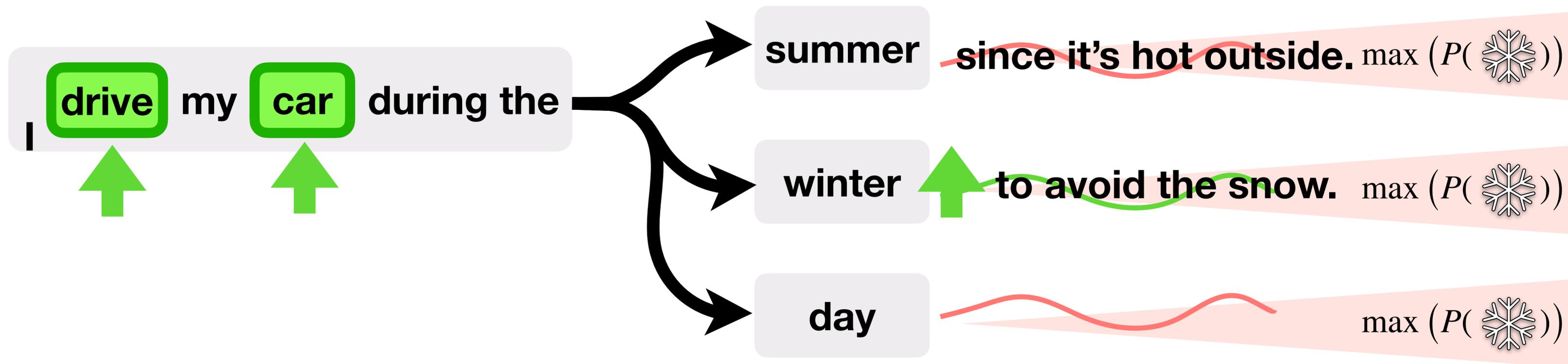
- Initialize priority queue of nodes Q_f
- Until goal reached:
 - Expand best node from Q_f
 - Score and add the node's neighbors to Q_f

Q_f : $O(\text{num words}^{\text{sequence length}})$ storage

Use a beam instead : "A*esque" search

NeuroLogic A*-Esque Decoding [Lu et al 2022]

Write a sentence with: **car** \wedge **drive** \wedge snow



Constraints: {sponge, pour, pool, side, clean}

Example output: Pour water on a sponge and use it to clean the side of the pool.

(sponge ∨ sponges) ∧ (pour ∨ pours ∨ pouring ∨ poured) ∧ (pool ∨ pools) ∧ (side ∨ sides) ∧ (clean ∨ clean ∨ cleans ∨ cleaning)

C

beam search

The woman, whose name has not been released, was taken to a local hospital, where she was listed in stable condition, according to the sheriff's office.

completely irrelevant

NeuroLogic

The man **cleans** a **sponge** in a **pouring pool** at the **side** of the road.

slightly awkward

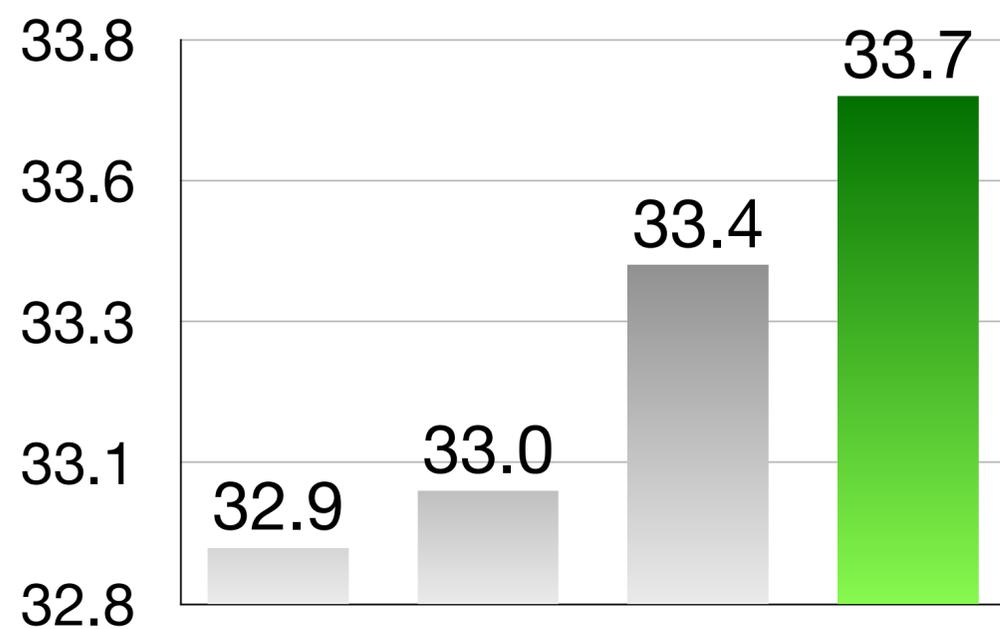
A* NeuroLogic

The boy **cleaned** the **side** of the **pool** with a **sponge**, and **poured** water over it .

Constrained MT

(Dinu et al., 2019)

- MarianMT (Junczys et al., 2018)
- Post and Vilar (2018)
- NeuroLogic (Lu et al., 2021)
- NeuroLogic A*esque

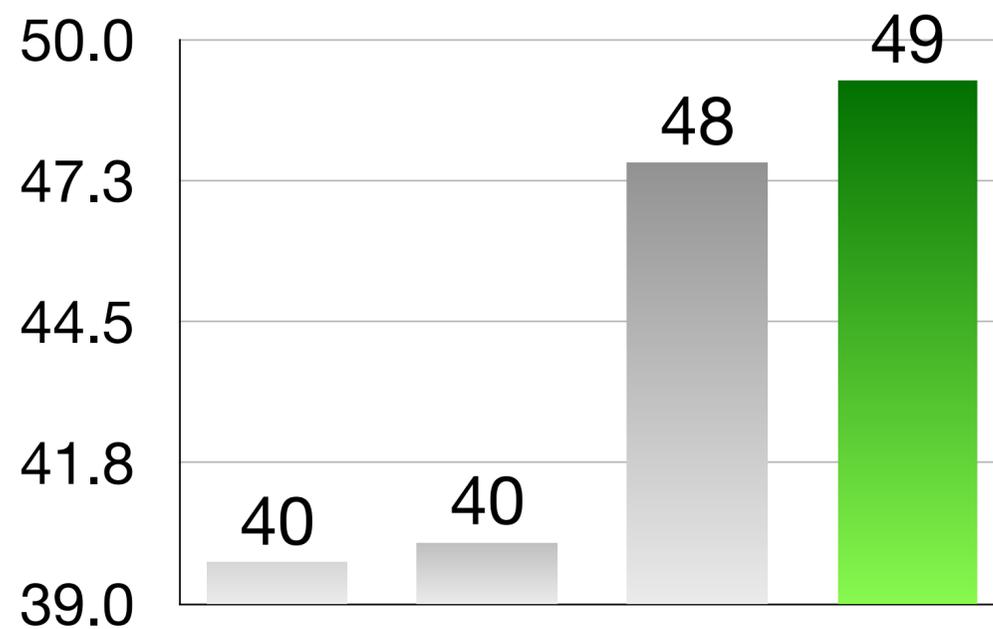


BLEU

Few-Shot E2ENLG

(Chen et al., 2020)

- KGPT-Graph (Chen et al., 2020b)
- KGPT-Seq (Chen et al., 2020b)
- NeuroLogic (Lu et al., 2021)
- NeuroLogic A*esque

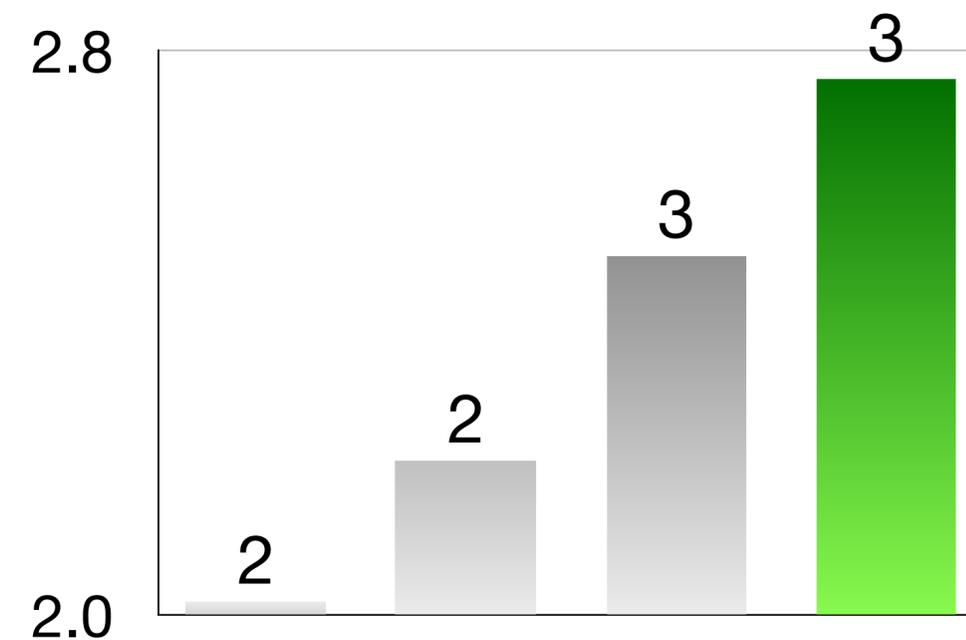


BLEU

Question Generation

(Zhang et al., 2020)

- CGMH (Miao et al., 2019)
- TSMH (Zhang et al., 2020)
- NeuroLogic (Lu et al., 2021)
- NeuroLogic A*esque

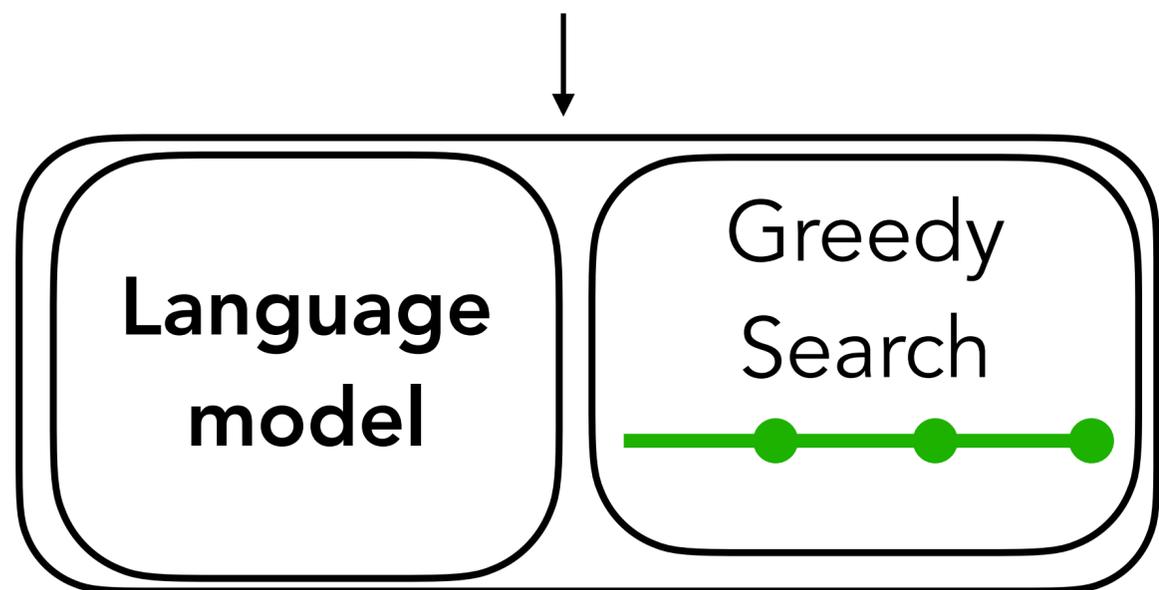


Human Eval Score

Theorem

Let x be an even integer.

Then $x + 5$ is odd.



Proof by Contradiction: Aiming for a contradiction, suppose $x + 5$ is even.

Then there exists $r \in \mathbb{Z}$ such that $x + 5 = 2r$.

But from **Odd Integer Plus Odd Integer is Odd**, $2r + 1$ is odd.

The result follows by **Proof by Contradiction**.

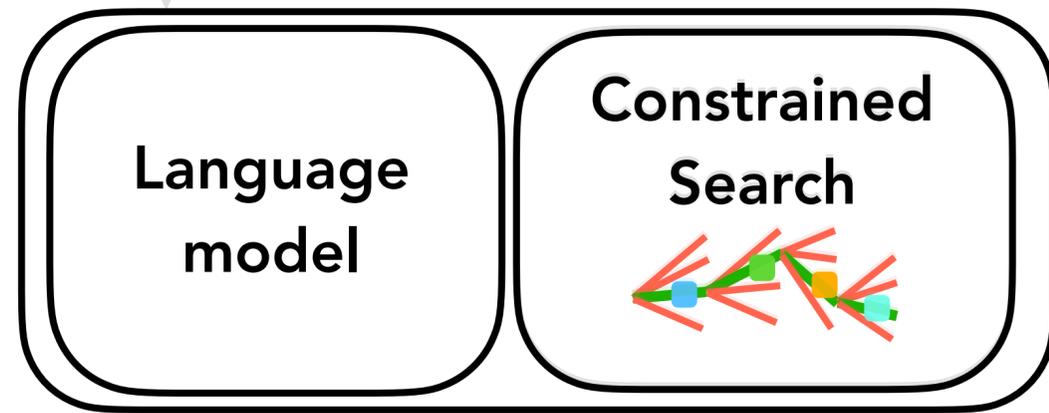
Fake theorem

Invalid reasoning

Theorem

Let x be an even integer.

Then $x + 5$ is odd.



Proof by Contradiction: Aiming for a contradiction, suppose $x + 5$ is even. Then there exists an integer k such that $x + 5 = 2k$. This contradicts the premise that x is even. Hence, by Proof by Contradiction, $x + 5$ is odd.

- **Stepwise Stochastic Beam Search**

[Welleck et al 2022]

- Beam-search over arbitrary-length segments with a constraint scoring function.

Recap

- Text generation is graph search!
 - Searching for a path in a directed graph of tokens
 - Edges: scored by a neural language model
- **New search algorithms** : higher quality, constraint satisfaction, etc.

Thanks for your attention!