Prompting

Instructor: Lara J. Martin (she/they)

TA: Omkar Kulkarni (he)

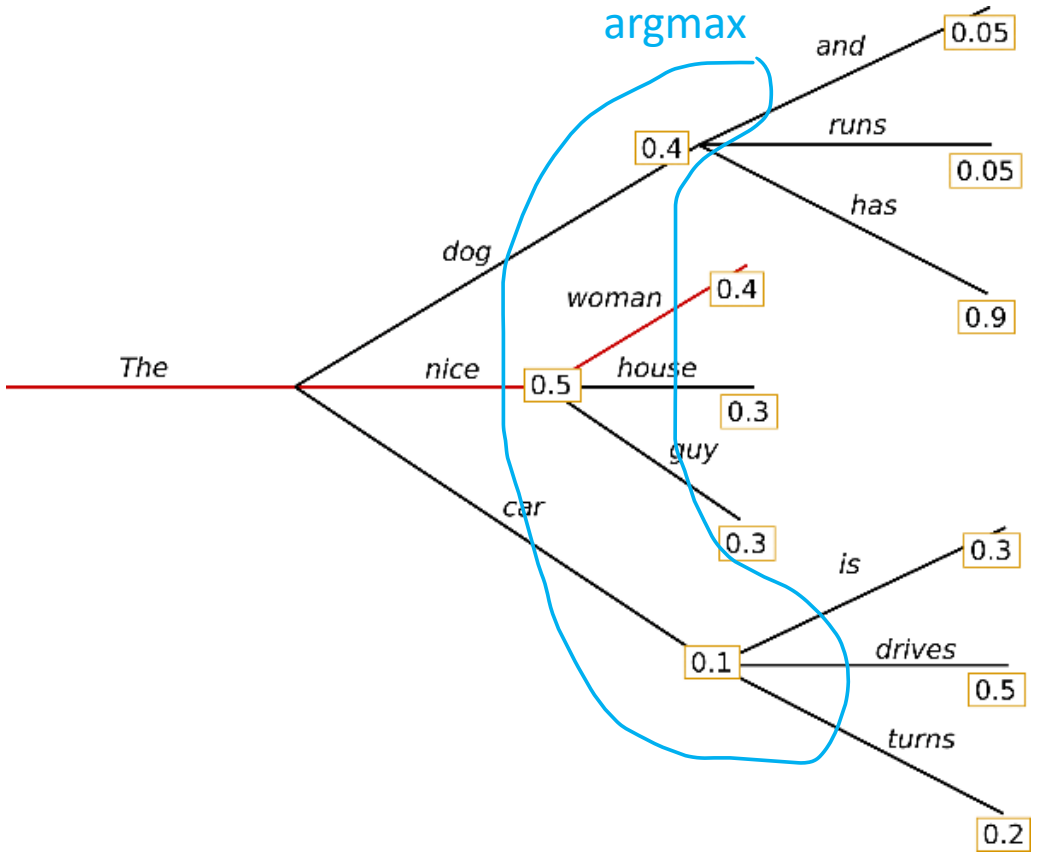
https://laramartin.net/NLP-class/

Learning Objectives

Distinguish between prompting techniques

Try common prompting techniques like chain-of-thought

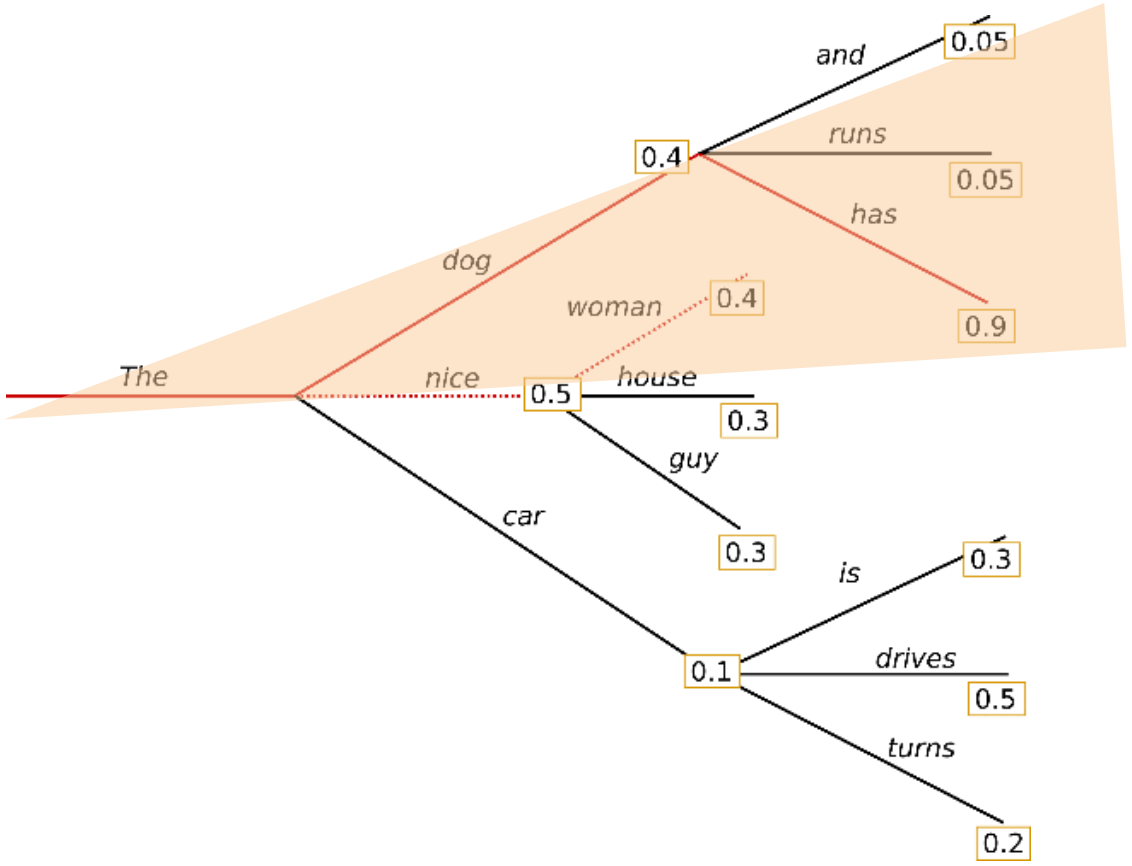
Review: Greedy Search



https://huggingface.co/blog/how-to-generate

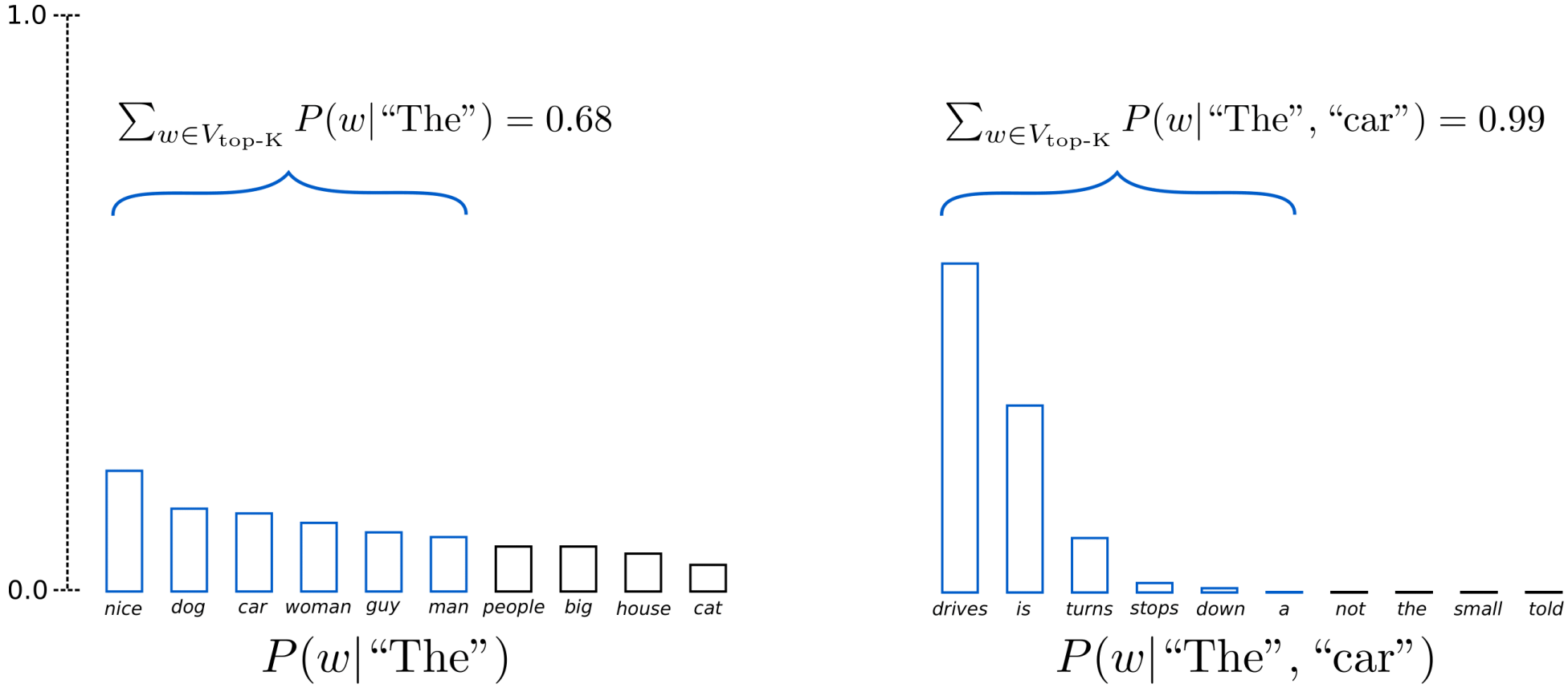
Review: Beam Search

Number of beams = 2



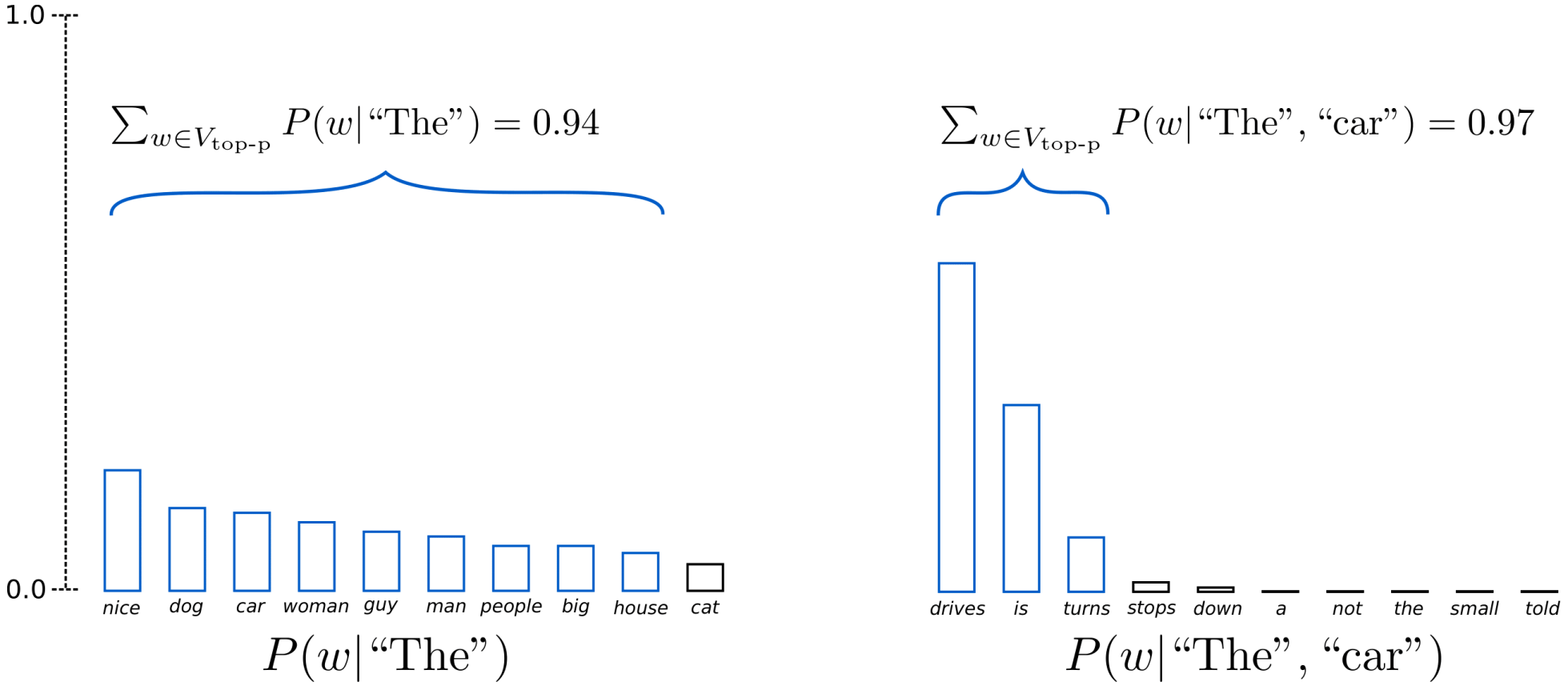
https://huggingface.co/blog/how-to-generate

Review: Top-K Sampling



https://huggingface.co/blog/how-to-generate

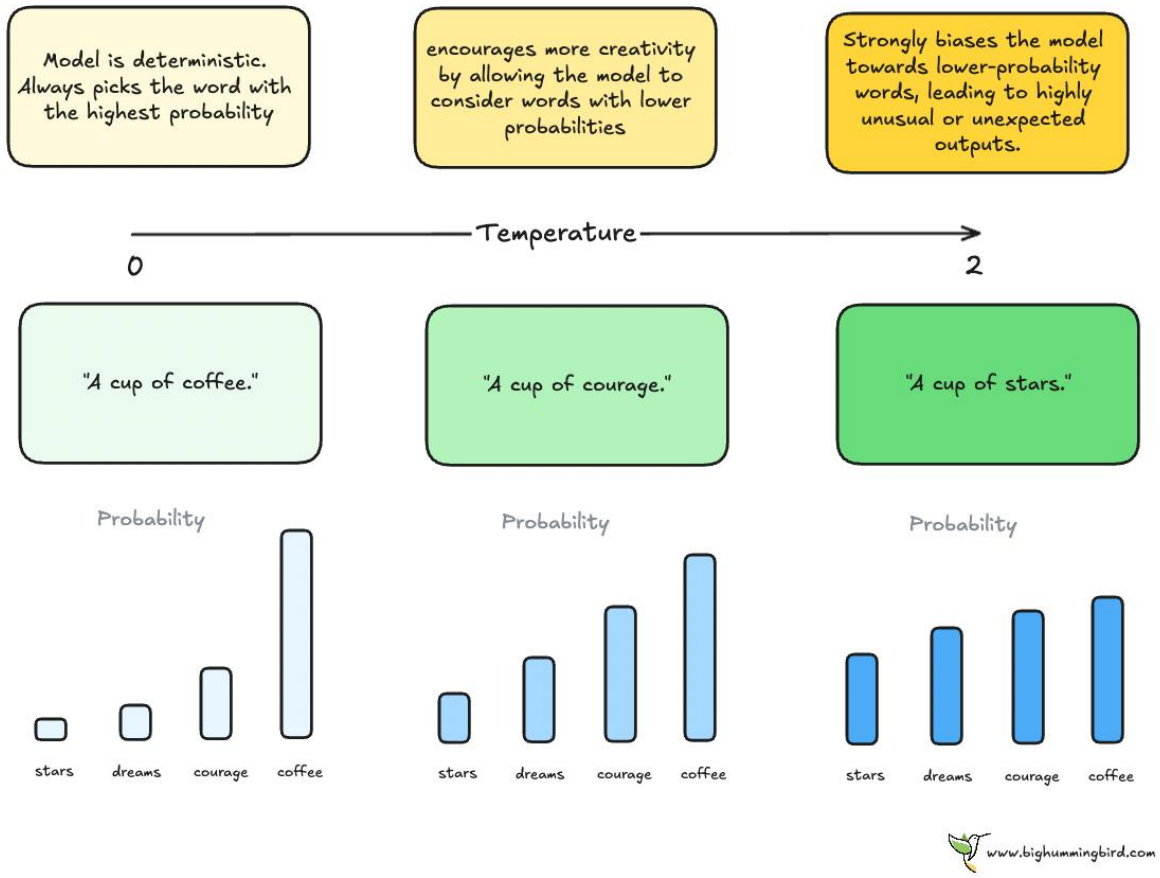
Review: Top-P/Nucleus Sampling



A. Holtzman, J. Buys, M. Forbes, and Y. Choi, “The Curious Case of Neural Text Degeneration,” in *International Conference on Learning Representations (ICLR)*, 2020, p. 16. https://openreview.net/forum?id=rygGQyrFvH

https://huggingface.co/blog/how-to-generate

Review: “Temperature”



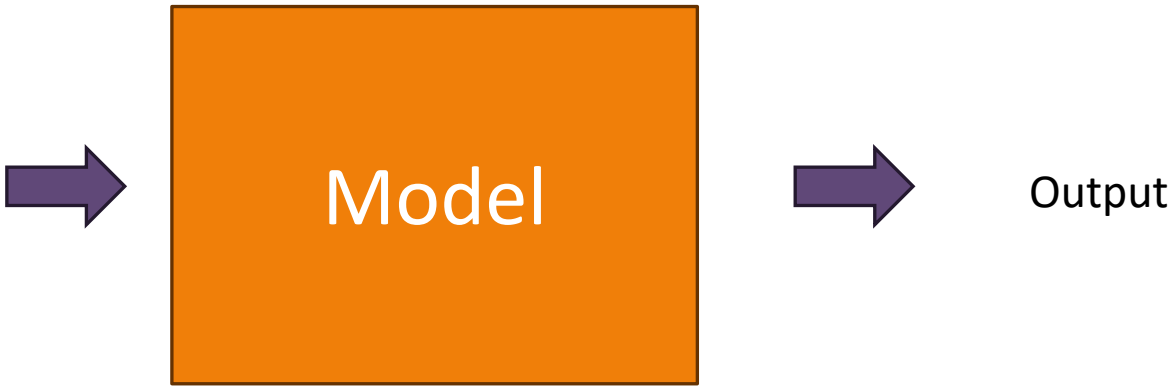
https://www.bighummingbird.com/blogs/llm-hyperparameter

Review: Zero-shot Prompting

You are a helpful assistant.
You will be tagging the parts of speech in sentences.

Instructions

Task



Output

Sentence:
The dog ate the giant fish.

Review: Few-shot Prompting

Instructions

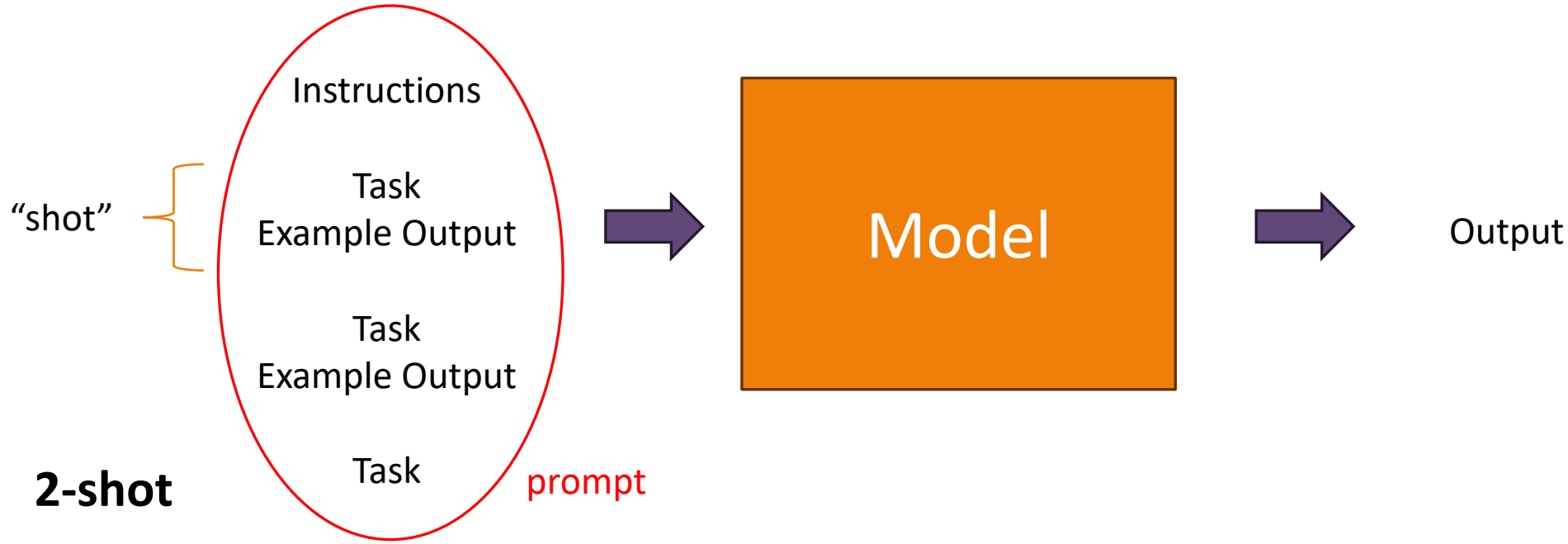
You are a helpful assistant. You will be tagging the parts of speech in sentences.

Task

Sentence:
The dog ate the giant fish.

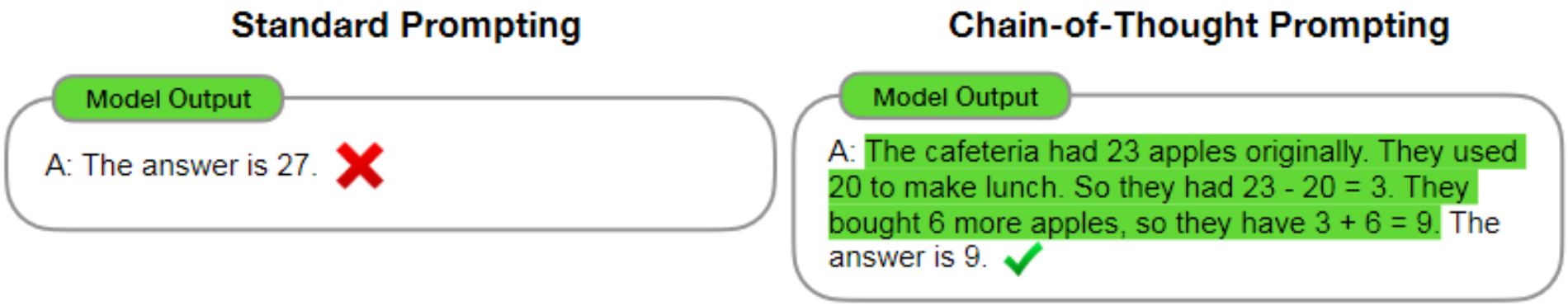
Example Output

The dog ate the giant fish.
D N V D Adj N



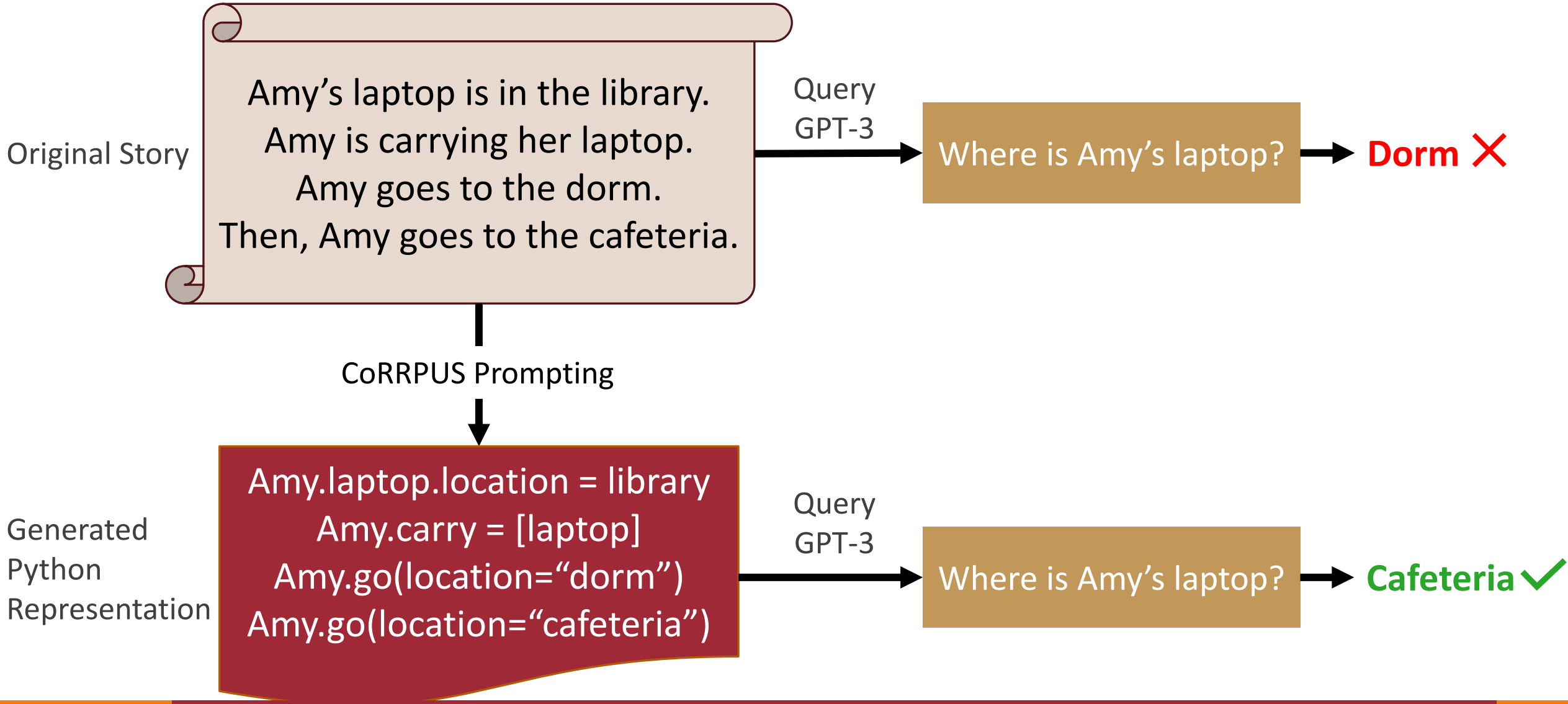
Review: Chain-of-Thought Prompting

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?



Part of Figure 1 from J. Wei *et al.*, "Chain of Thought Prompting Elicits Reasoning in Large Language Models," in *International Conference on Neural Information Processing Systems (NeurIPS)*, New Orleans, LA & Online, Jun. 2022. doi: 10.48550/arXiv.2201.11903.

CoRRPUS (Code Representations to Reason & Prompt over for Understanding in Stories)



Dong, Y. R., **Martin, L. J.,** & Callison-Burch, C.
"CoRRPUS: Code-based Structured Prompting for Neurosymbolic Story Understanding." *Findings of ACL 2023.*

CoRRPUS Chain-of-Thought Prompting

Three versions that are initialized the same:

Comment

```
def story(self):
    ## Mary moved to the bathroom.
    self.Mary.location = “bathroom”
    ## Mary got the football there.
    self.Mary.inventory.append(“football”)
    …
```

Specific Functions

```
self.Mary_moved_to_the_bathroom()
self.Mary_got_the_football_there()
self.John_went_to_the_kitchen()
self.Mary_went_back_to_the_garden()

def Mary_moved_to_the_bathroom()
    self.Mary.location=“bathroom”
def Mary_got_the_football_there():
…
```

Abstract Functions

```
def go(self, character, location):
    character.location = location
    for item in character.inventory:
        item.location = location
def pick_up(): …

def story(self):
    ## Mary moved to the bathroom.
    self.go(character=self.Mary,
    location = “bathroom”)
    …
```


Dong, Y. R., **Martin, L. J.,** & Callison-Burch, C.
“CoRRPUS: Code-based Structured Prompting for Neurosymbolic Story Understanding.” *Findings of ACL 2023.*

Tested On 2 Tasks

bAbI (Weston et al. 2015)

- Task 2: Stories tracking objects that characters carry

Re^3 (Yang et al. 2022)

- Identifying inconsistencies in stories (e.g., descriptions of characters' appearances, relationships)
- Stories were generated from a list of facts (the premise). They also generated premises with a contradiction.

Dong, Y. R., **Martin, L. J.,** & Callison-Burch, C.
"CoRRPUS: Code-based Structured Prompting for Neurosymbolic Story Understanding." *Findings of ACL 2023.*

bAbI (Weston et al. 2015)

Method	# Shot	Accuracy ↑
Random	-	25%
GPT-3	1	56.5%
Chain of Thought (Creswell et al. 2022)	1	46.4%
Selection-Inference (Creswell et al. 2022)	1	29.3%
Dual-System (Nye et al. 2021)	10	100%
CoRRPUS (comment)	**1**	**67.0%**
CoRRPUS (specific)	**1**	**78.7%**
CoRRPUS (abstract)	**1**	**99.1%**

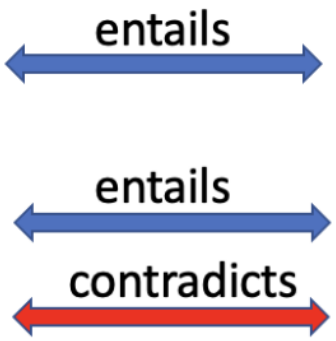
Dong, Y. R., **Martin, L. J.,** & Callison-Burch, C.
"CoRRPUS: Code-based Structured Prompting for Neurosymbolic Story Understanding." *Findings of ACL 2023.*

Re^3

The task is to see what stories match what premises based on the facts extracted from both.

Joan Westfall premise

Attribute	Value
Gender	Female
Occupation	Teacher
Brother	Brent Westfall
Appearance	Blue eyes



Joan Westfall in story

Attribute	Value
Gender	Female
Father	Jason Westfall
Brother	Brent Westfall
Appearance	Brown eyes

Takeaway: structured representations help!

Re^3 (Yang et al. 2022)

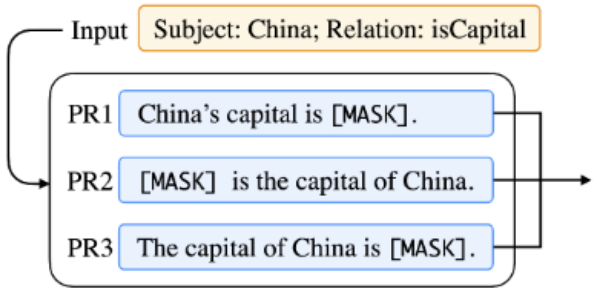
Method	ROC-AUC ↑
Random	0.5
GPT-3	0.52
Entailment (Yang et al. 2022)	0.528
Entailment with Dense Passage Retrieval (Yang et al. 2022)	0.610
Attribute Dictionary → Sentence (Yang et al. 2022)	0.684
CoRRPUS (comment)	**0.751**
CoRRPUS (specific)	**0.794**
CoRRPUS (abstract)	**0.704**

→ Probably because functions like `set_age(self, character, age)` complicate more than they help.

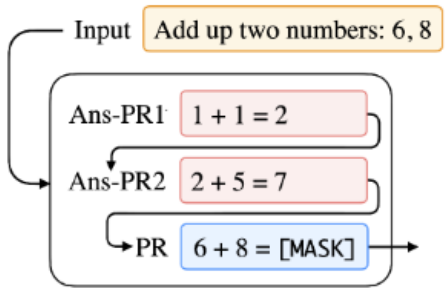
Dong, Y. R., **Martin, L. J.,** & Callison-Burch, C.
"CoRRPUS: Code-based Structured Prompting for Neurosymbolic Story Understanding." *Findings of ACL 2023.*

Multi-prompt “Learning”

Multiple unanswered prompts



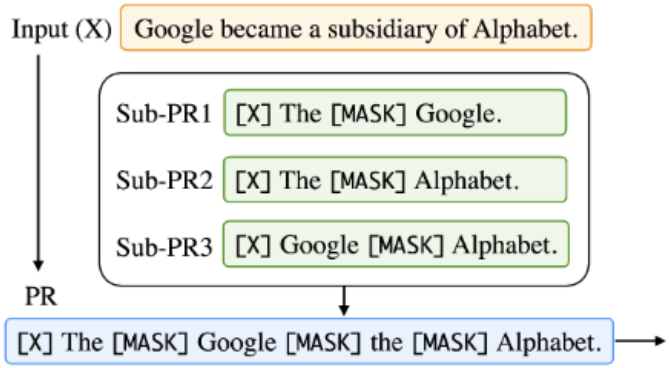
(a) Prompt Ensembling.



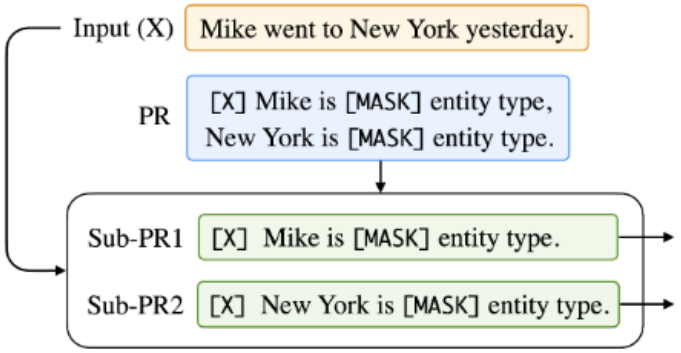
(b) Prompt Augmentation.

Demonstration learning

Merging subprompts



(c) Prompt Composition.



(d) Prompt Decomposition.

For multi-label predictions

“ ” for input text, “ ” for prompt, “ ” for answered prompt. “ ” for sub-prompt.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. ACM Comput. Surv. 55, 9, Article 195 (September 2023), 35 pages. https://doi.org/10.1145/3560815

Prompt Composition

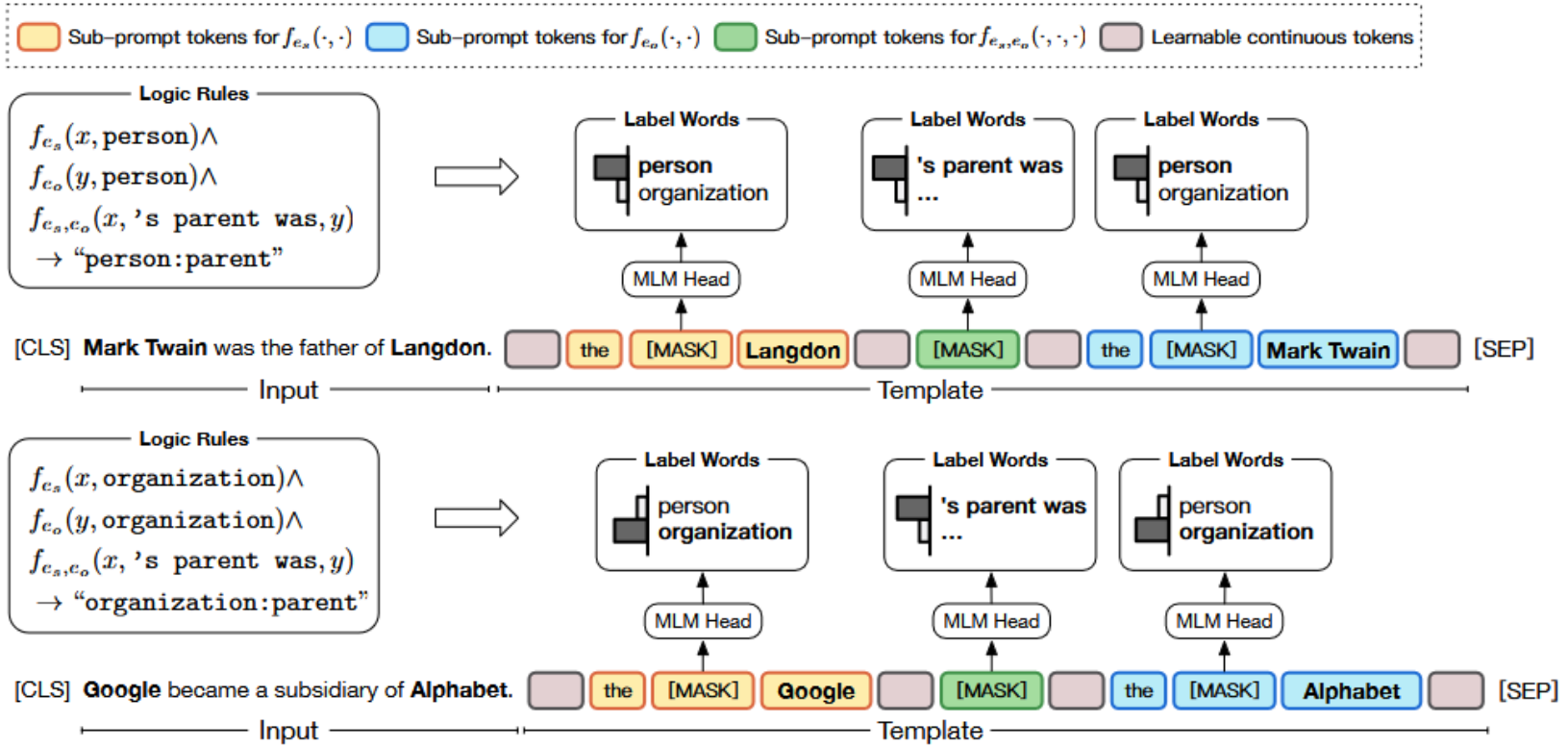


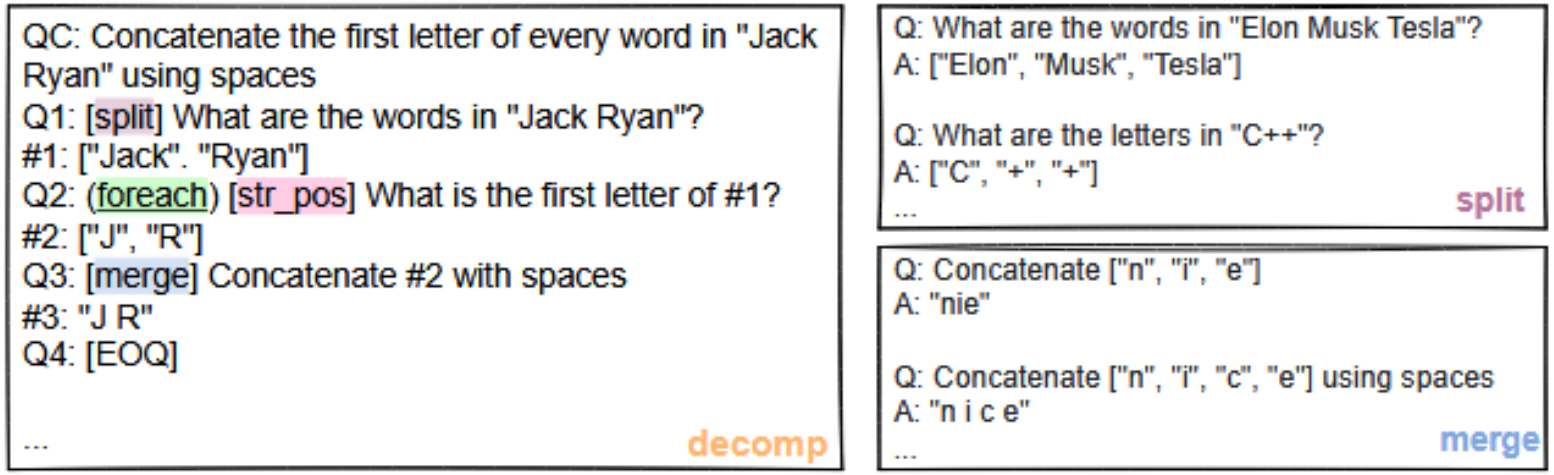
Figure 2: Illustration of PTR. By composing the sub-prompts of the conditional functions $f_{e_s}(\cdot,\cdot)$, $f_{e_o}(\cdot,\cdot)$ and $f_{e_s,e_o}(\cdot,\cdot,\cdot)$, we could easily get an effective prompt to distinguish "person:parent" and "organization:parent".

Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, Maosong Sun. 2021. PTR: Prompt Tuning with Rules for Text Classification. arXiv: 2105.11259

Prompt Decomposition

Breaking down complicated problems into components that feed into each other

Each instance of a model does a different step or function



Tushar Khot, Harsh Trivedi, Matthew Finlayson, Yao Fu, Kyle Richardson, Peter Clark, Ashish Sabharwal. 2023. Decomposed Prompting: A Modular Approach for Solving Complex Tasks. In *Proceedings of the International Conference on Learning Representations.*

Self-Criticism

LLM “ruminates” on its output to try to come up with better output

(Along with chain-of-thought) Precursor to reasoning models that are finetuned to do this automatically

```
Question: Who was the third president of the United States?
Here are some brainstormed ideas: James Monroe
Thomas Jefferson
John Adams
Thomas Jefferson
George Washington
Possible Answer: James Monroe
Is the possible answer:
 (A) True
 (B) False
The possible answer is:
```


Saurav Kadavath, Tom Conerly, Amanda Askell, Tom Henighan, Dawn Drain, Ethan Perez, *et al.* 2022. Language Models (Mostly) Know What They Know. arXiv: 2207.05221

Meta-Prompting

Prompting to generate prompts

Meta-prompts…

- “encode what the LLM should “do” given different descriptions of the same task” [1]
- “will return the relevant outputs for an arbitrary task, provided that the task description is provided as an input” [1]

[1] Adrian de Wynter, Xun Wang, Qilong Gu, Si-Qing Chen. 2023. On Meta-Prompting. arXiv: 2312.06562

Meta-Prompting Example (WHAT-IF)

```
{{
“branching_event_number”: branching_event,
“original_decision” : storyline[f“node_{branching_node}”]['decision'],
“alternate_decision” :
     storyline[f“node_{branching_node}”]['alternate_decision'],
“new_story_length”: (len(storyline) - branching_node) * 3,
“major_plot_points” : mpp,
“prompt”: f“TODO: <a prompt for ChatGPT for every branching point above with
 following requirements:
     1. Ask to use the original storyline as a reference to write an alternate
        storyline that branches out at event {branching_event} if {char_name}
        {storyline[f'node_{branching_node}']['alternate_decision']} instead
        of {storyline[f'node_{branching_node}']['decision']}.
     2. Provide 5 thought-provoking concrete guiding questions as potential
        directions to explore that expand the following:\n
           a. How would alternate decision change or replace {mpp}?
           b. How would {char_name} make key decisions that overcome new
              challenges and propel the story forward?
     3. Describe what an ideal alternate storyline should look like.
     4. Ask to output the alternate storyline as a list of
        {(len(storyline) - branching_node + 1) * 3} events that has
        {storyline[f'node_{branching_node}']['alternate_decision']} as the
        first event>”
}}
```


Runsheng "Anson" Huang, Lara J. Martin, Chris Callison-Burch. 2025. WHAT-IF: Exploring Branching Narratives by Meta-Prompting Large Language Models. Wordplay: When Language Meets Games Workshop (EMNLP 2025)

Other Tricks of the Trade

Instruction-tuned models like GPT-3.5 and Mistral-7B-Instruct like to be given a "role" first (e.g., "You are a helpful writing assistant.")

The more defined the task, the better

- More details
- One thing to do at a time

LLMs are overly confident (like people on the internet)

- To "objectively" have the model evaluate something, you should create a new instance and ask it

Chain-of-thought prompting helps models come up with better answers

They will "Yes and…" your prompt

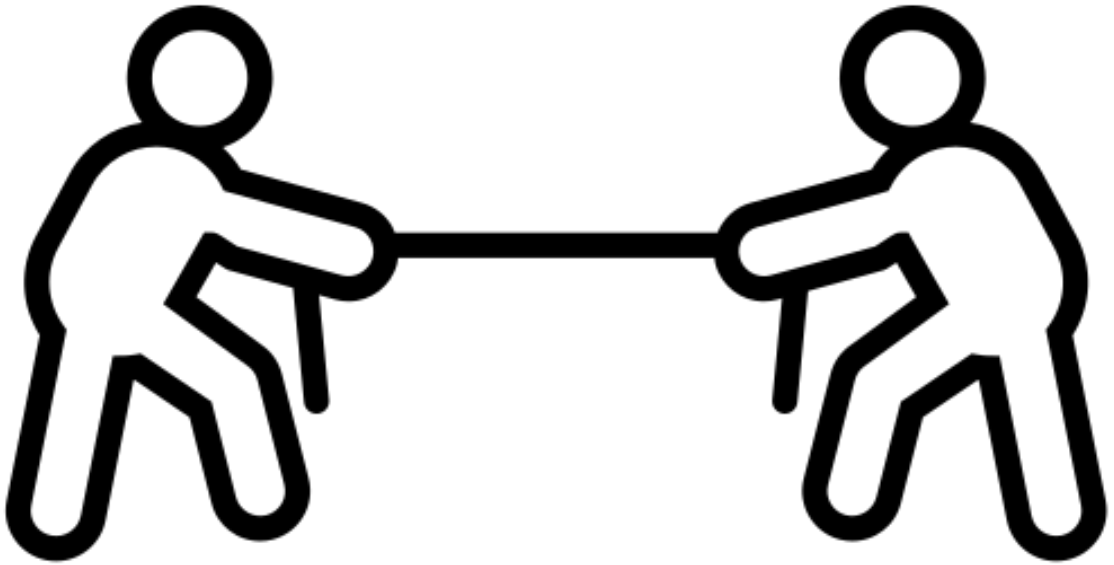
Dealing with any language models

Likelihoods → Not cause & effect

What is probable might not be possible.

Lara's Language Model Tradeoff

Coherence



Originality

https://thenounproject.com/icon/tug-of-war-1016981/

Your Turn!

This is very similar to HW 3!

Think of something you're an expert in. It can be anything!

Pick an LLM that's hosted online such as ChatGPT (https://chatgpt.com/) or Claude (https://claude.ai).

Ask your LLM to give you information about that topic. Ask in different ways about different things and use different prompting techniques.

What does the LLM do well with?

What does it not do well with?

Some Prompting Techniques:

Few-shot

Zero-shot

Chain of thought

Decomposition

Self-Criticism