

# Towards Consistent Language Models Using Controlled Prompting and Decoding

**Jasmin Mousavi**, Arash Termehchy

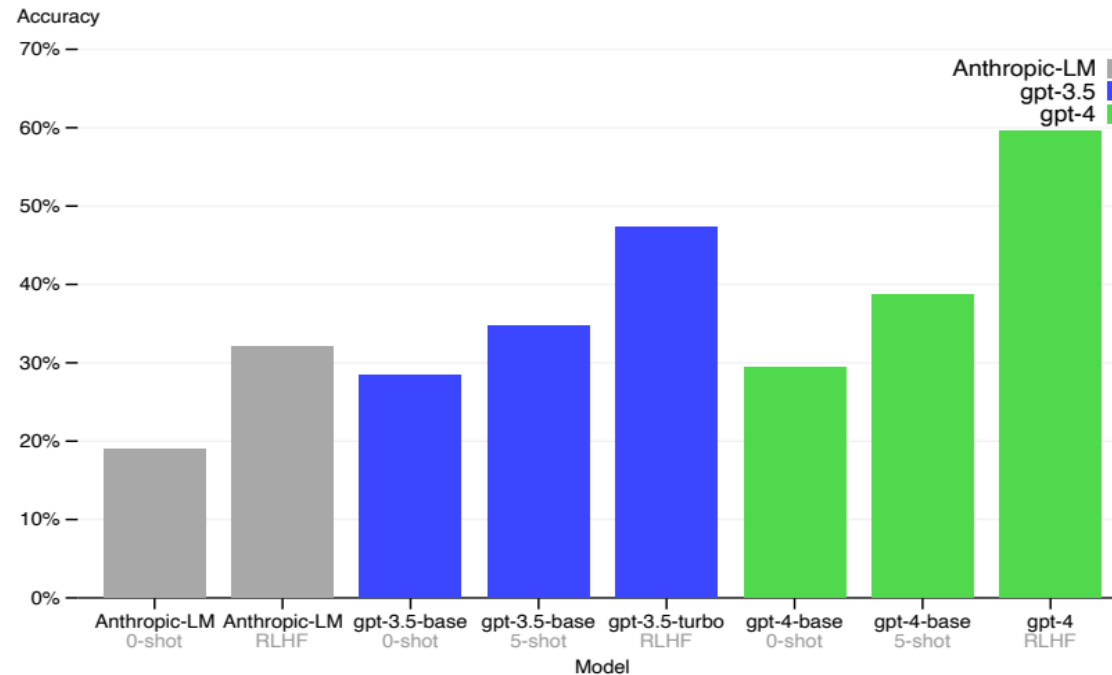


# LLMs hallucinate

- Produce inconsistent information and inaccurate results

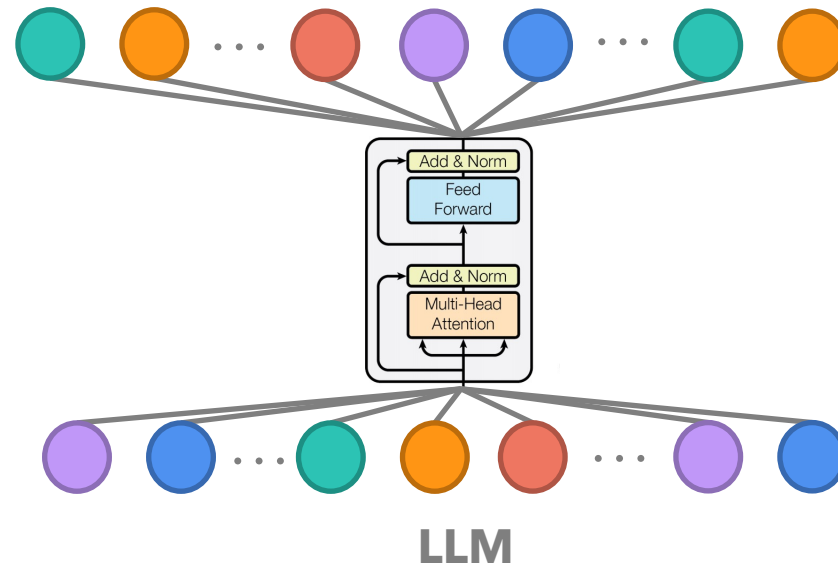
## GPT-4: 40% hallucination rate

Accuracy on adversarial questions (TruthfulQA mc1)



# Over-reliance on generalization

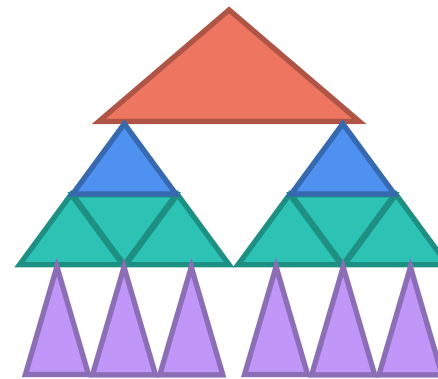
- LLMs *over-generalize patterns and relationships* from pretraining data
  - Causes inconsistent and inaccurate results



# How do we resolve inconsistencies?

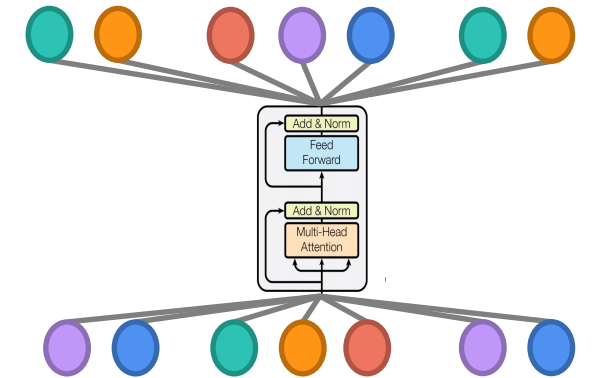
## Leverage Constraints!

- Abstracts data into logical rules
  - Reduces what is injected into LLMs
- Incorporated softly
- Provides a structured way of controlling the output of LLMs



Constraints

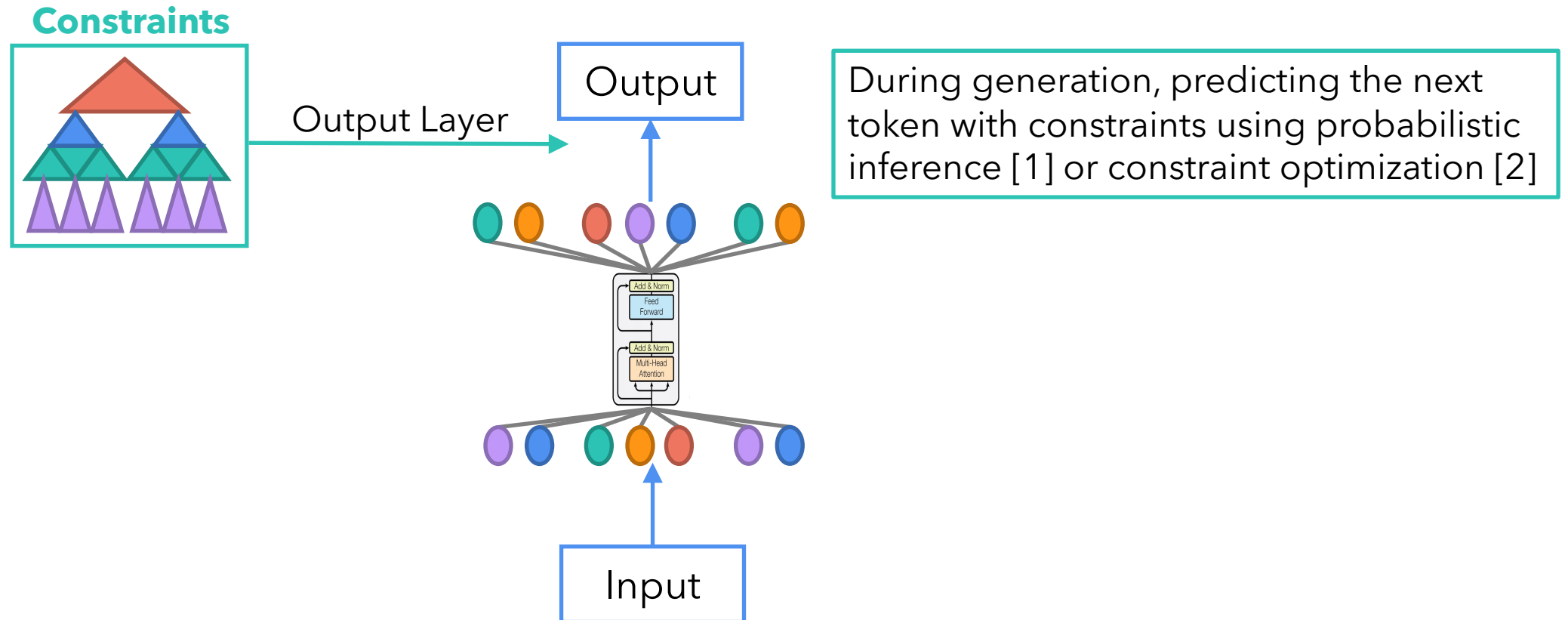
+



LLM

# Decoding under constraints: current efforts towards reducing inconsistencies in LLMs

- Ensuring the output of LLMs is consistent with domain constraints

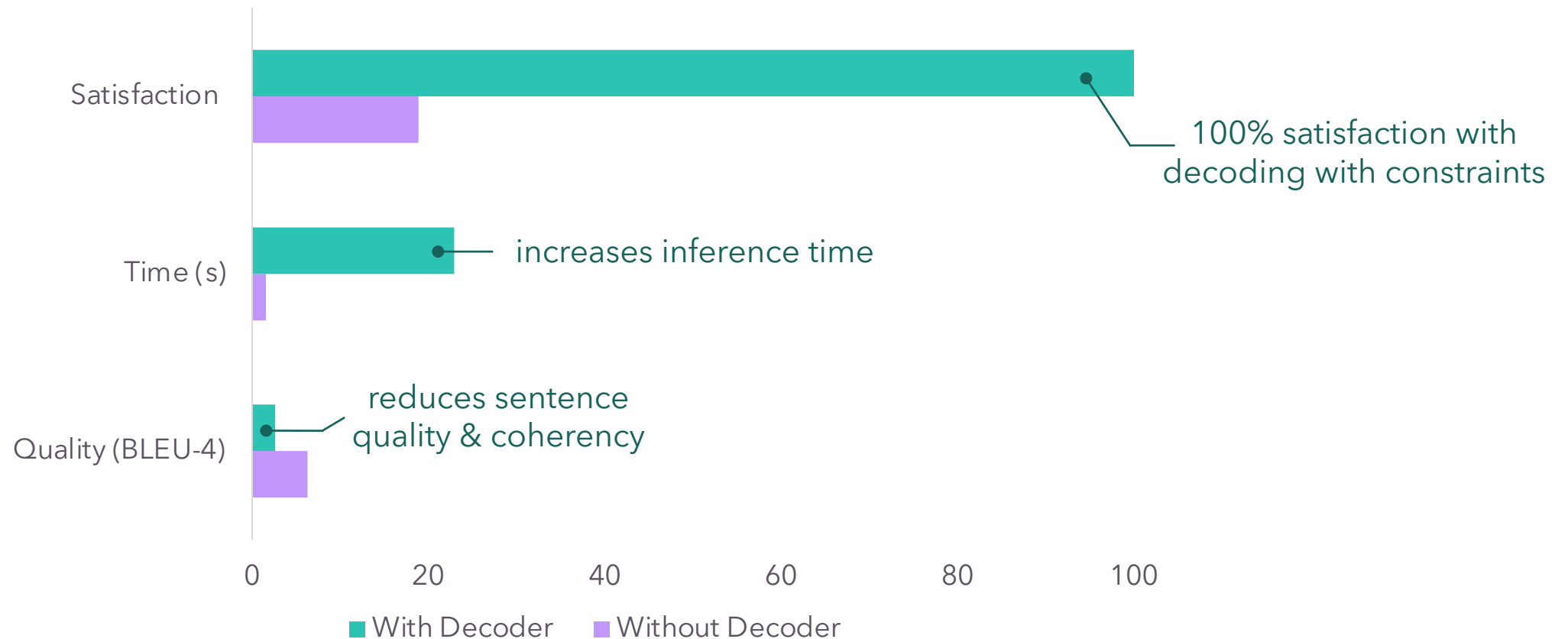


[1] Sequential Monte Carlo Steering of Large Language Models using Probabilistic Programs , 2023

[2] NeuroLogic Decoding: (Un)supervised Neural Text Generation with Predicate Logic Constraints , 2021

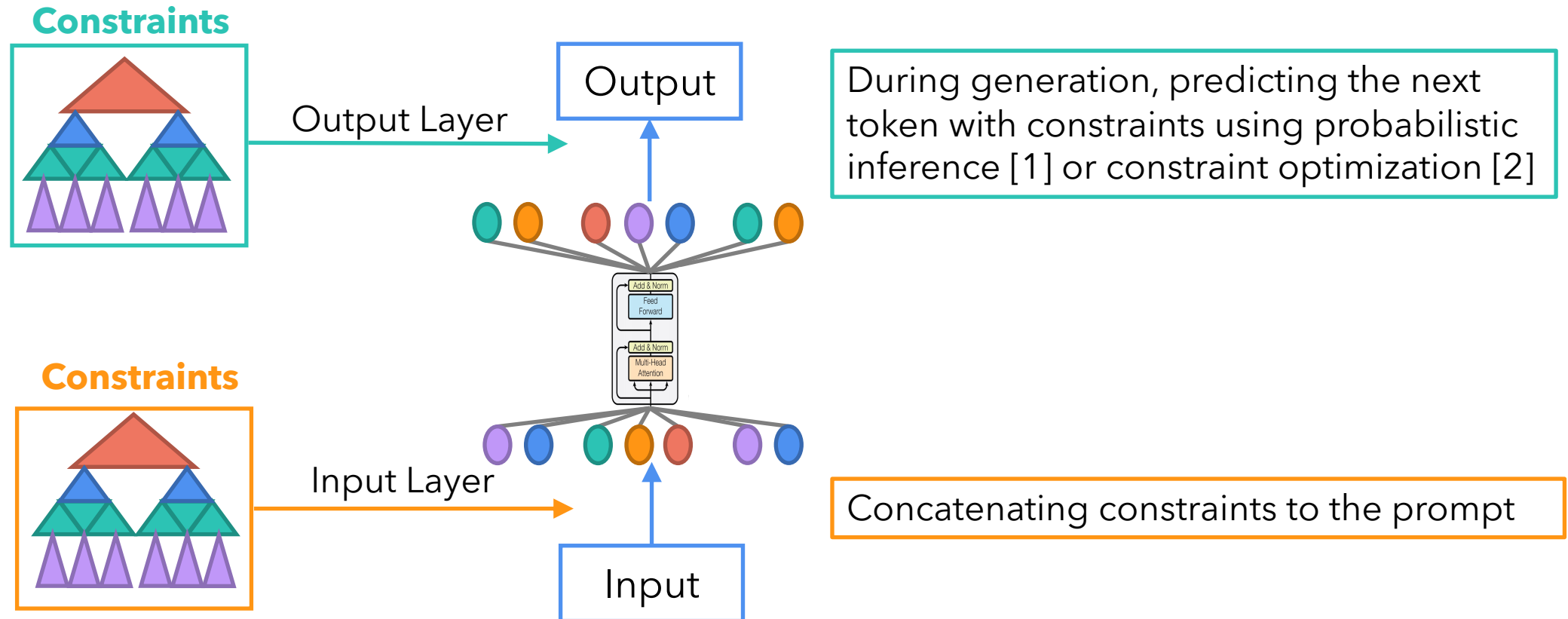
# Limitations of decoding with constraints (on CommonGen dataset)

- CommonGen constraints = contains key words or their inflections



# Can prompting mitigate the limitations of constrained decoding?

- Adding constraints to the input of LLMs

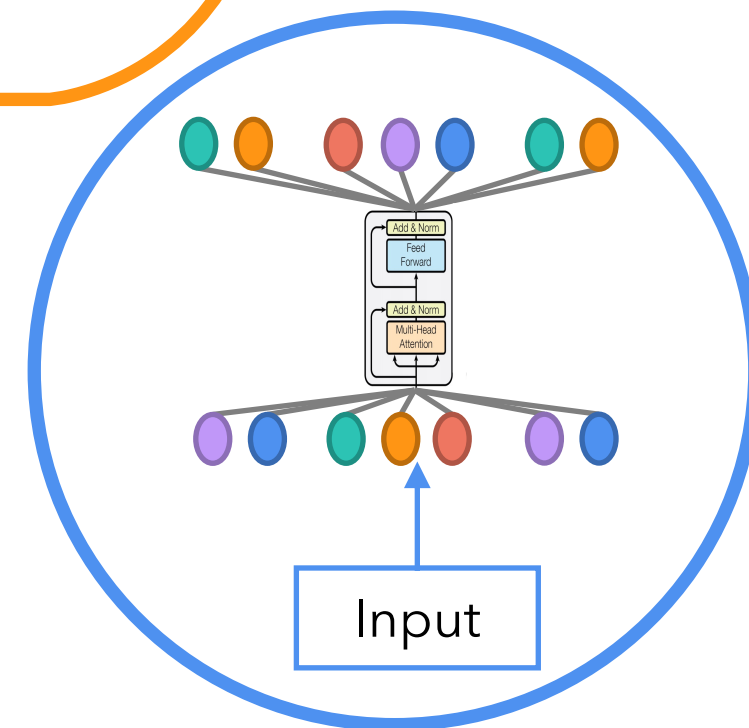
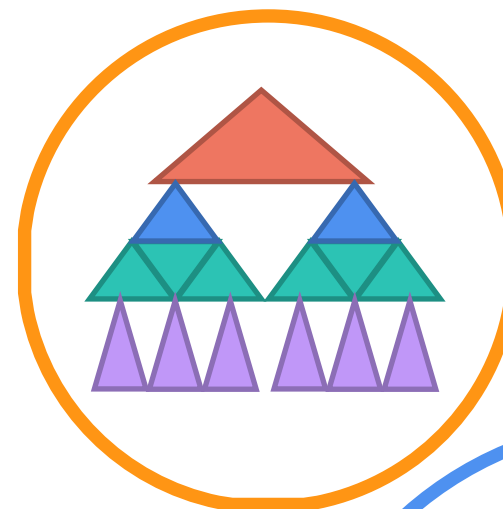


[1] Sequential Monte Carlo Steering of Large Language Models using Probabilistic Programs , 2023

[2] NeuroLogic Decoding: (Un)supervised Neural Text Generation with Predicate Logic Constraints , 2021

# Challenges with adding constraints to the prompt

1. How to **represent constraints** in the prompt
  - Constraints are *symbolic*
2. Limited **context length**





# Prompting with Constraints in Conjunctive Normal Form (CNF)

- Constraints = contains key words or their inflections
  - Can be represented in CNF!
- Convert CNF constraint to text

**CNF**

## Conjunctive Normal Form

Write a sentence using the words (*word1a* or *word1b* or ... ) and (*word2a* or *word2b* or ... ) and (*word3a* or *word3b* or ... )

# Addressing context length through abstraction

- Abstract constraint representation
  - Reduces length of constraint supplied to the prompt

**ABS**

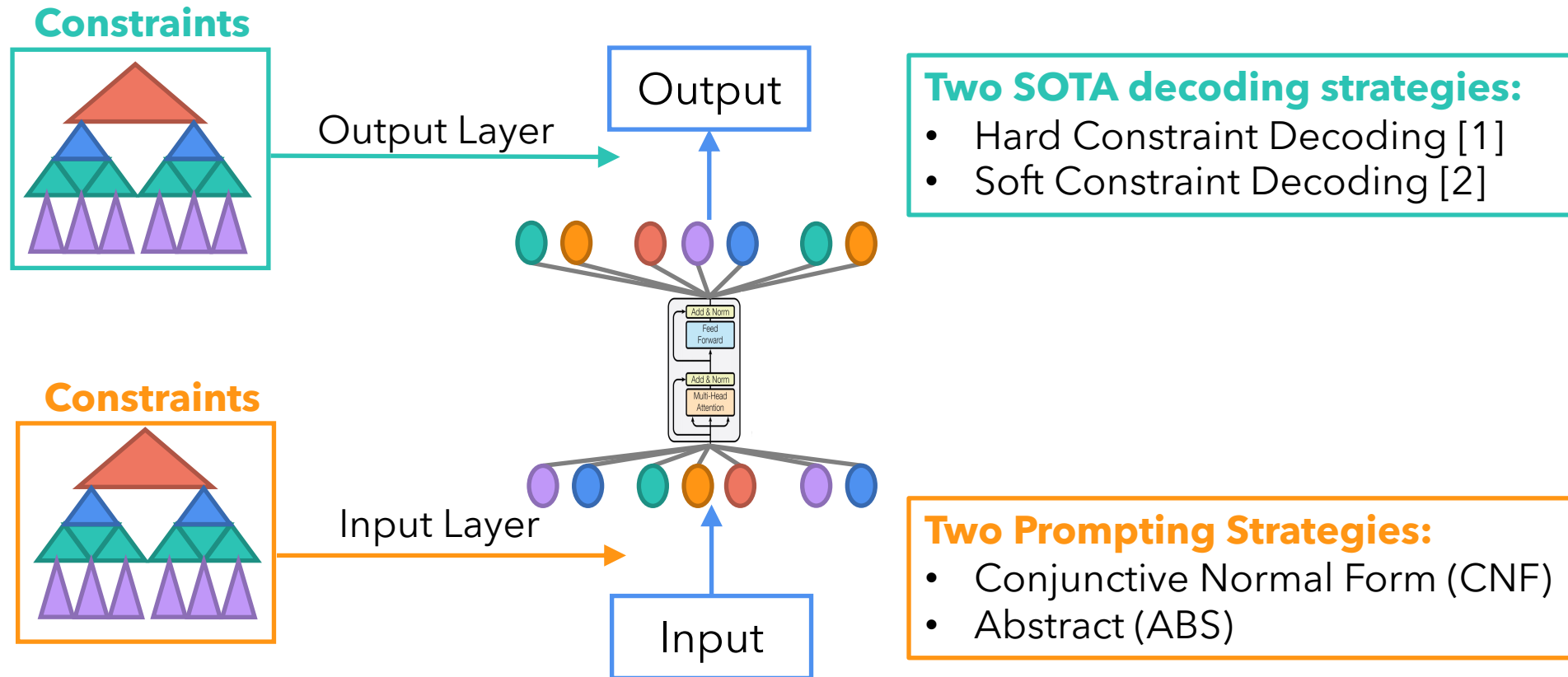
## **Abstract**

Given a set of words  $x$ , write a sentence using all words in  $x$  or inflections of  $x$

# Prompting improves time and quality but cannot guarantee satisfaction



# Empirical study: prompting and decoding with constraints



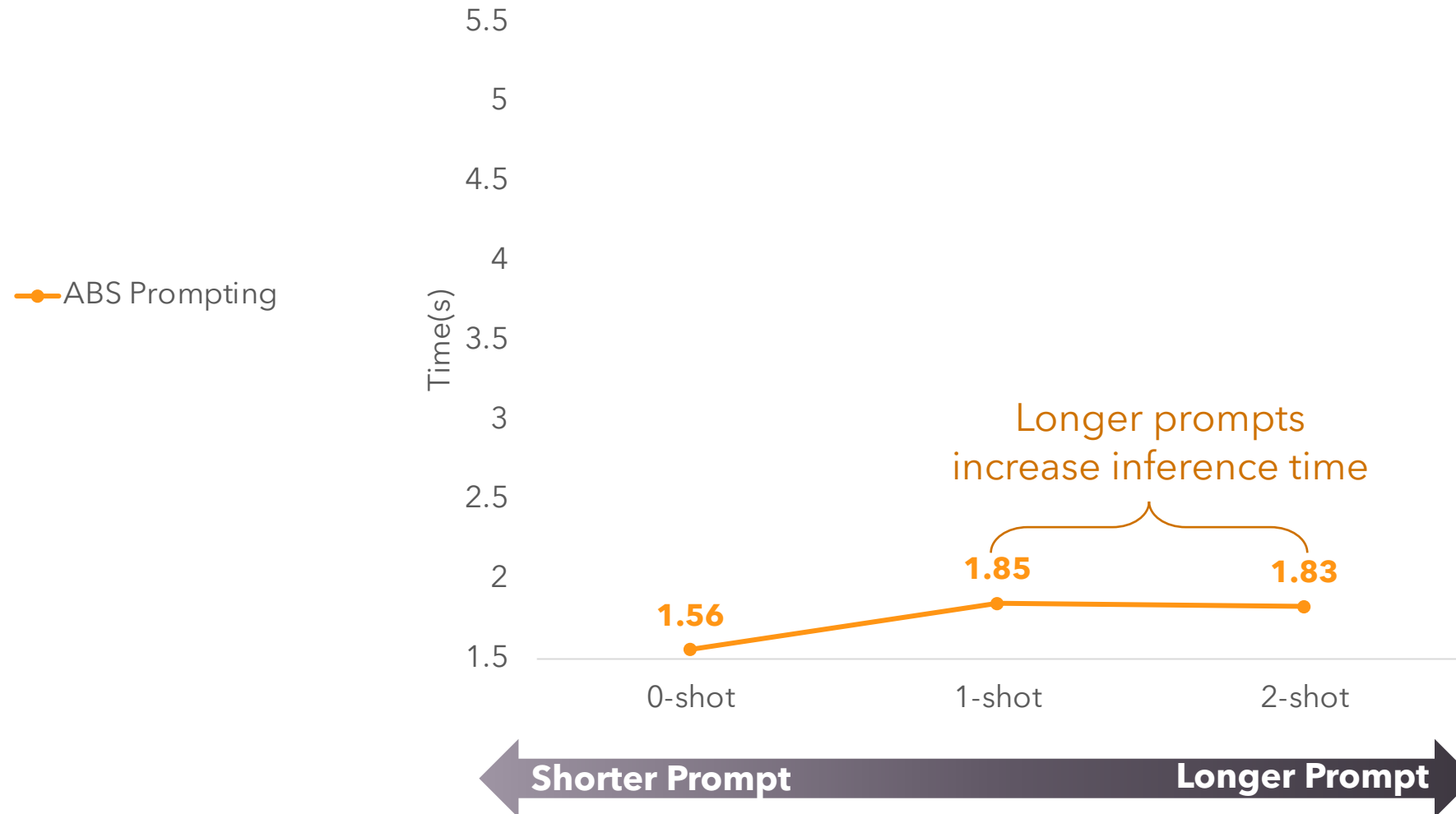
- CommonGen dataset [3] constraints = contains key words or their inflections

[1] Sequential Monte Carlo Steering of Large Language Models using Probabilistic Programs , 2023

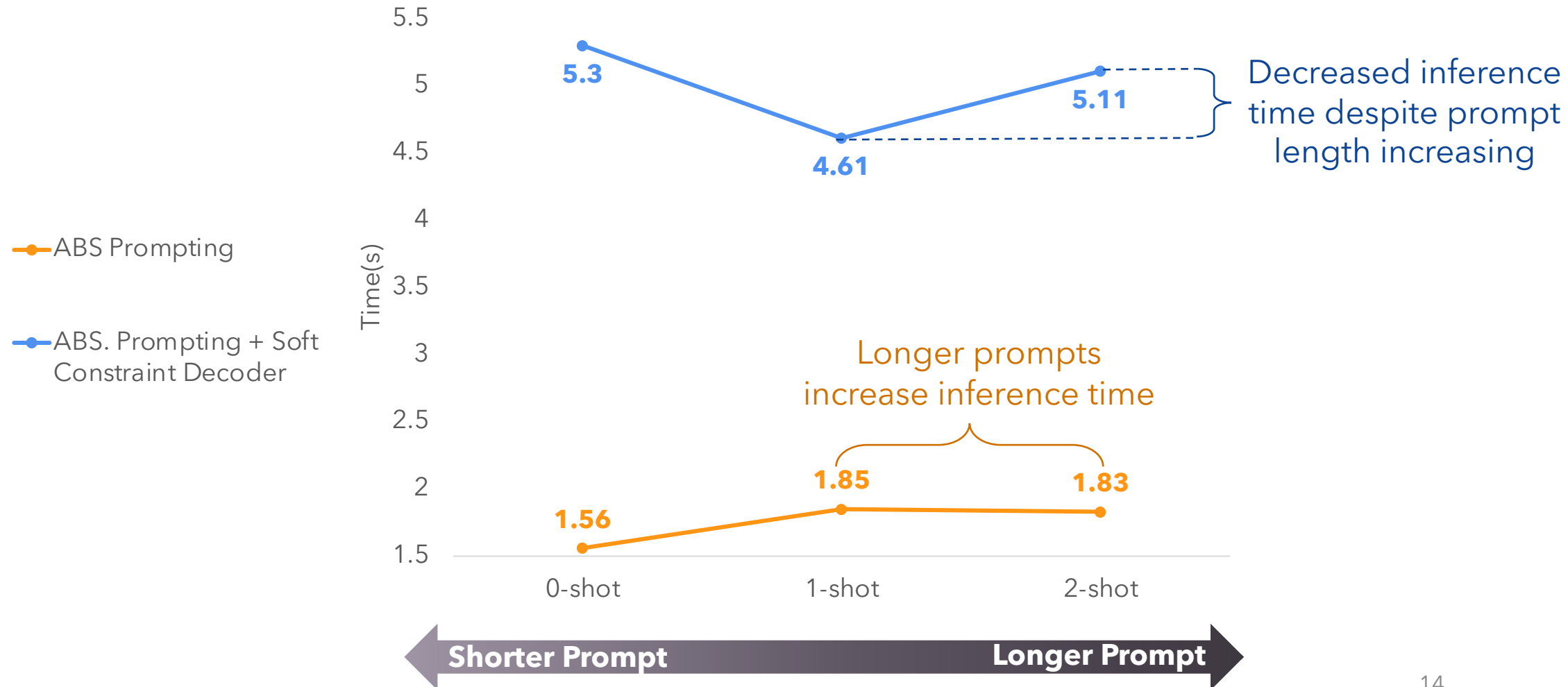
[2] NeuroLogic Decoding: (Un)supervised Neural Text Generation with Predicate Logic Constraints , 2021

[3] CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning, 2019

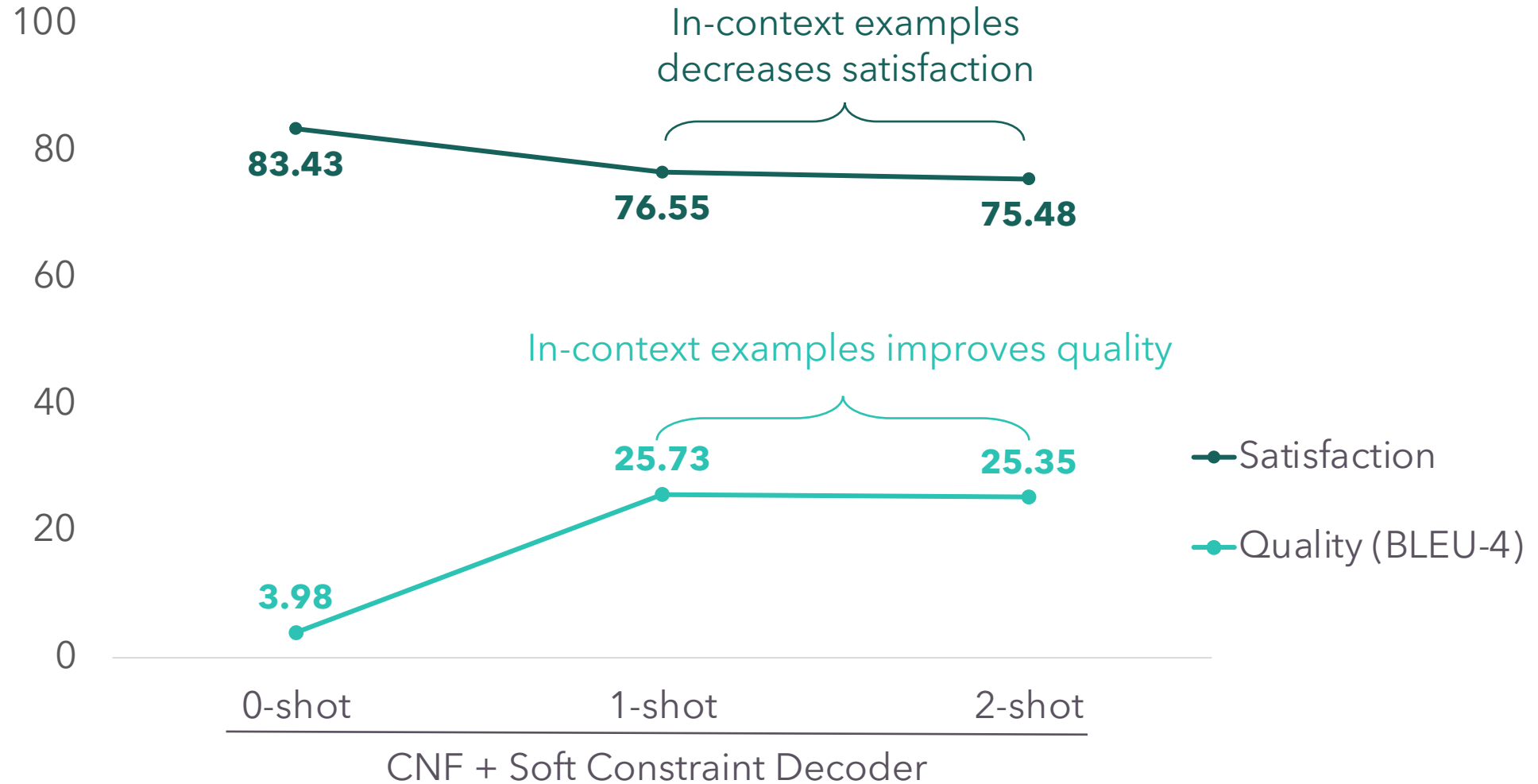
# Prompting reduces the search space for (soft style) decoders using search strategies



# Prompting reduces the search space for (soft style) decoders using search strategies



# Irrelevant context hurts satisfaction



# Key Takeaways

1

Decoding with constraints *improves* satisfaction, but **hurts** quality and time

2

Prompting with constraints *improves* quality and time, but **hurts** satisfaction

3

Prompting **reduces search space** for decoders using search strategies

4

In-context examples *improves* quality, but may **decrease** satisfaction

5

Challenging to optimize over *all metrics* at the same time

Thank you!