

CIS 6930 Special Topics in Large Language Models

LLM Reasoning

Outline

- Introduction to LLM Reasoning
- Methods for LLM Reasoning
 - 1. Chain-of-Thought Prompting / Decoding
 - 2. Least-to-Most
 - 3. Self-Consistency
 - 4. Self-Improve
 - 5. Analogical Prompting
 - 6. Step-back Abstraction Prompting
 - 7. Small LM Reasoning
- Analysis of LLM Reasoning

Introduction to LLM Reasoning

What is Reasoning

Using **facts** and **logic** to arrive at an answer

Deductive Reasoning: Use logic to go from premise to firm conclusion

Premise: All mammals have kidneys

Premise: All whales are mammals

Conclusion: All whales have kidneys

Inductive Reasoning: From observation, predict a likely conclusion

Observation: When we see a creature with wings, it is usually a bird

Observation: We see a creature with wings.

Conclusion: The creature is likely to be a bird

Abductive Reasoning: From observation, predict the most likely explanation

Observation: The car cannot start and there is a puddle of liquid under the engine.

Likely Explanation: The car has a leak in the radiator

Formal vs Informal Reasoning

Formal Reasoning: Follows formal rules of logic along with axiomatic knowledge to derive conclusions.

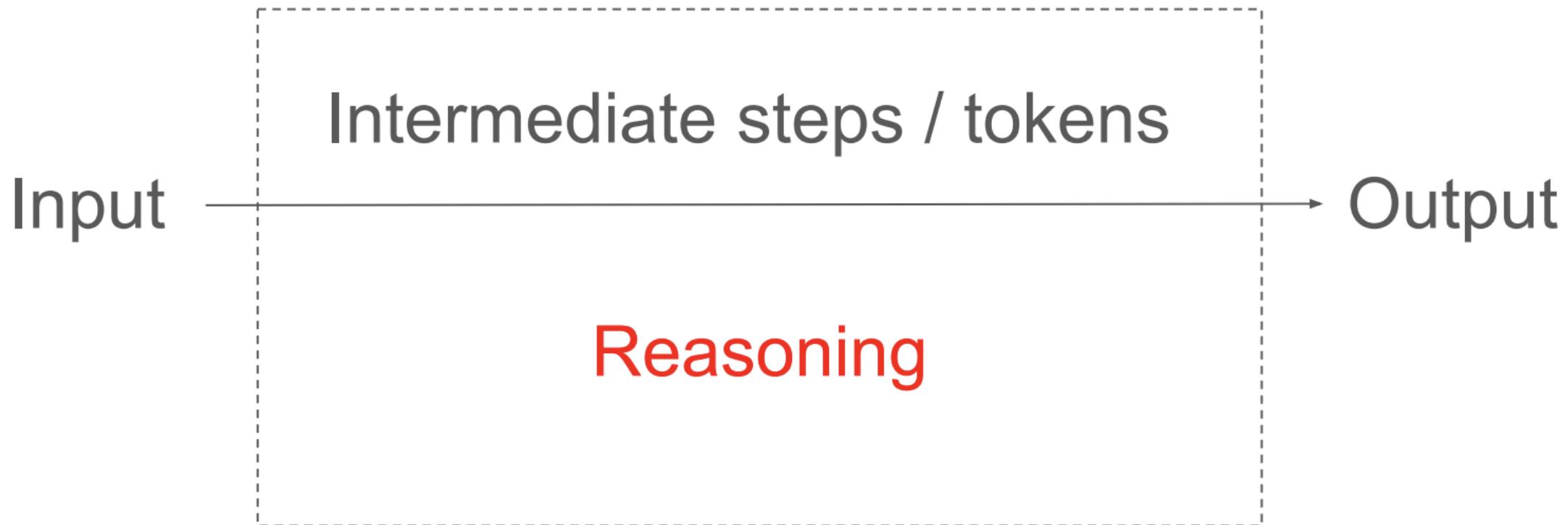
Informal Reasoning: Uses intuition, experience, common sense to arrive at answers.

For most of this lecture, by “reasoning” we mean informal deductive reasoning, often involving multiple steps

LLM Reasoning

- Large Language models are really good at predicting plausible continuations of text (text generation via language modeling), that respect constraints in the input (conditioned generation), and align well with human preferences (post-training lecture)
- Question: Can current LLMs reason?

LLM Reasoning



Ling et al. Program Induction by Rationale Generation: Learning to Solve and Explain Algebraic Word Problems. ACL 2017

Chen et al. Compositional generalization via neural-symbolic stack machines. NeurIPS 2020.

LLM Reasoning

What is the output when concatenating the last letter of each word in “artificial intelligence”?

No reasoning

The answer is “le”.

Reasoning

The last letter of “artificial” is “l”. The last letter of “intelligence” is “e”. Concatenating “l” and “e” leads to “le”. So the answer is “le”.



Why “Intermediate Tokens” / “Reasoning” Matters?

- For any problems solvable by boolean circuits of size T , **constant-size transformers** can solve it by generating $O(T)$ intermediate tokens
- If directly generating final answers, either requires a huge depth or cannot solve at all

Zhiyuan Li, Hong Liu, Denny Zhou, and Tengyu Ma. [Chain of Thought Empowers Transformers to Solve Inherently Serial Problems](#). ICLR 2024.

Chain-of-Thought

Chain-of-Thought Prompting

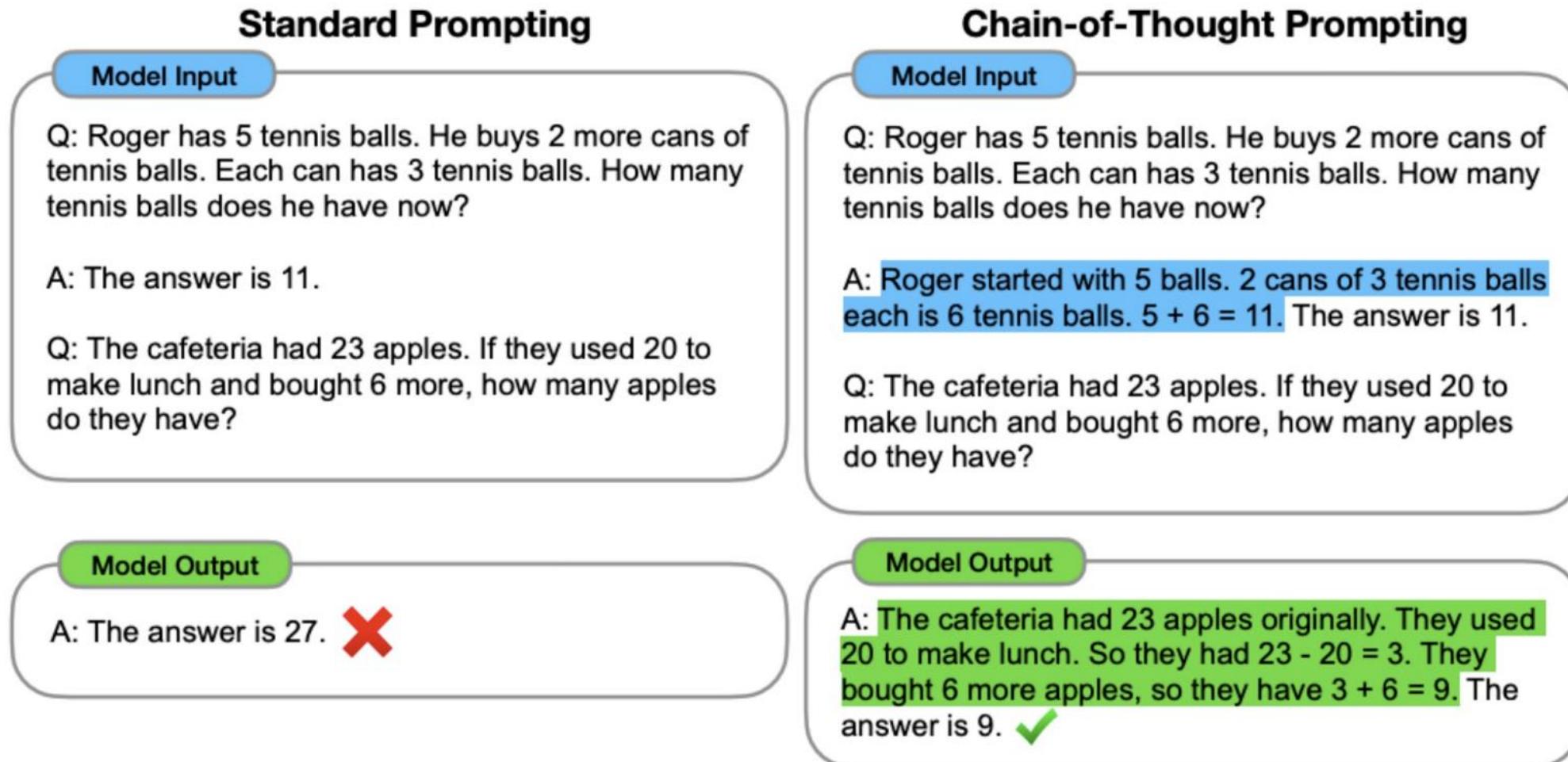


Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

Zero-shot Chain-of-Thought Prompting

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The answer is 8. X

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are $16 / 2 = 8$ golf balls. Half of the golf balls are blue. So there are $8 / 2 = 4$ blue golf balls. The answer is 4. ✓

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: The answer (arabic numerals) is

(Output) 8 X

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: **Let's think step by step.**

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

Pros and Cons of Chain-of-Thought Prompting

Pros: Simple and works

Cons:

CoT prompting needs task-specific examples

“Let’s think step by step” is generic, but performs much worse than few-shot

Pros and Cons of Chain-of-Thought Prompting

Prompting approaches are actually weird

When asking someone a question —

will you first show similar problems/solutions before asking?

or, at the end of your question, will you have to say “let’s think step by step”?

Of course not!

Chain-of-Thought Decoding

Pretrained LLMs cannot reason without further
prompting, engineering or finetuning

WRONG

Pretrained LLMs are ready to reason

All we need is decoding

Chain-of-Thought Decoding

No reasoning? Check more generation candidates!

I have 3 apples. My dad has 2 more apples than me. How many apples do we have in total?

5 apples. (Greedy Decoding)

I have 3 apples, my dad has 2 more apples than me, so he has 5 apples. $3+5=8$.

We have 8 apples in total.

You have 3 apples, your dad has 2 more apples than you, so he has 5 apples. $3+5=8$.

The answer is 5.

Chain-of-Thought Decoding

How to select the best response? By length?

I have 3 apples. My dad has 2 more apples than me. How many apples do we have in total?

5 apples. (Greedy Decoding)

I have 3 apples, my dad has 2 more apples than me, so he has 5 apples. $3+5=8$.

We have 8 apples in total.

You have 3 apples, your dad has 2 more apples than you, so he has 5 apples. $3+5=8$.

The answer is 5.

Chain-of-Thought Decoding

Select responses with high confidence on answers!

I have 3 apples. My dad has 2 more apples than me. How many apples do we have in total?

5 apples. (Greedy Decoding)

I have 3 apples, my dad has 2 more apples than me, so he has 5 apples. $3+5=$

We have 8 apples in total.

You have 3 apples, your dad has 2 more apples than you, so he has 5 apples. $3+5=$

The answer is 5.

Way higher confidence on reasoning-based answers!

Chain-of-Thought Decoding

Chain-of-Thought Decoding

1. Go beyond greedy decoding by checking more generation candidates
 2. Choose candidates which have the highest confidence on the final answer
-

Recap

Chain-of-Thought Methods:

- Chain-of-Thought Prompting
 1. Chain-of-Thought Prompting – reasoning process of few examples
 2. Zero-shot Chain-of-Thought Prompting – “let’s think step by step”
- Chain-of-Thought Decoding

Sample multiple candidates and choose the one with the highest confidence on the final answer

Least-to-Most Prompting

Problem decomposition with Least-to-Most Prompting

Stage 1: Decompose Question into Subquestions

Q: It takes Amy 4 minutes to climb to the top of a slide. It takes her 1 minute to slide down. The water slide closes in 15 minutes. How many times can she slide before it closes?

Language Model

A: To solve "How many times can she slide before it closes?", we need to first solve: "How long does each trip take?"

Stage 2: Sequentially Solve Subquestions

Subquestion 1

It takes Amy 4 minutes to climb to the top of a slide. It takes her 1 minute to slide down. The slide closes in 15 minutes.

Q: How long does each trip take?

Language Model

A: It takes Amy 4 minutes to climb and 1 minute to slide down. $4 + 1 = 5$. So each trip takes 5 minutes.

Append model answer to Subquestion 1

Subquestion 2

It takes Amy 4 minutes to climb to the top of a slide. It takes her 1 minute to slide down. The slide closes in 15 minutes.

Q: How long does each trip take?

A: It takes Amy 4 minutes to climb and 1 minute to slide down. $4 + 1 = 5$. So each trip takes 5 minutes.

Q: How many times can she slide before it closes?

Language Model

A: The water slide closes in 15 minutes. Each trip takes 5 minutes. So Amy can slide $15 \div 5 = 3$ times before it closes.

Problem decomposition with Least-to-Most Prompting

Q: Elsa has 5 apples. Anna has 2 more apples than Elsa. How many apples do they have together?

A: Anna has 2 more apples than Elsa, so Anna has $2 + 5 = 7$ apples. Elsa and Anna have $5 + 7 = 12$ apples together. The answer is 12.

Standard CoT prompting

Q: Elsa has 5 apples. Anna has 2 more apples than Elsa. How many apples do they have together?

A: Let's break down this problem: 1. How many apples does Anna have? 2. How many apples do Elsa and Anna have together?

1. Anna has 2 more apples than Elsa. So Anna has $2 + 5 = 7$ apples.

2. Elsa and Anna have $5 + 7 = 12$ apples together.

Least-to-most prompting

Q: {question}

A: Let's break down this problem:

The answer is:

Problem decomposition with Least-to-Most Prompting

Accuracy by Steps (GSM8K)	All	2 Steps	3 Steps	4 steps	≥ 5 steps
Least-to-Most	62.39	74.53	68.91	59.73	45.23
Chain-of-Thought	60.87	76.68	67.29	59.39	39.07

Generalizes to more #steps than in-context example!

Prompting method	Accuracy
Zero-Shot	16.38
Standard prompting	17.06 ³
Chain-of-Thought (original)	61.18
Chain-of-Thought (1-shot)	60.88
Least-to-Most (1-shot)	62.39
Chain-of-Thought (best)	68.61 ³
Least-to-Most (best)	68.01

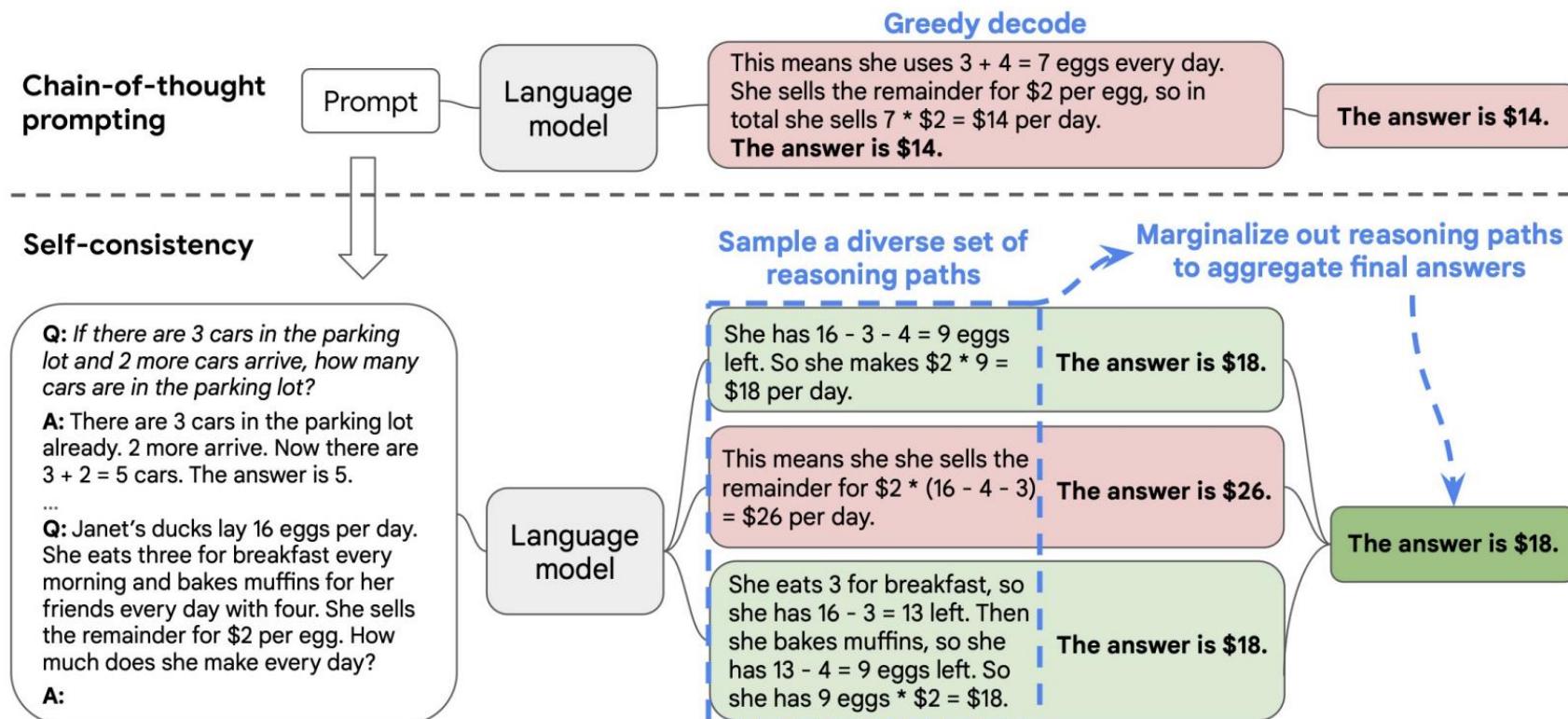
But with enough prompt engineering, CoT \approx Least-to-Most

Self-Consistency

Self-Consistency

CoT with "Self-consistency": Replace greedy decoding with an ensemble of samples...

Main idea: correct reasoning processes have greater agreement than incorrect processes.



Self-Consistency

Self-Consistency

1. Generate multiple responses by randomly sampling
 2. Choose the answer that appears most frequently
-

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, Denny Zhou.
[Self-Consistency Improves Chain of Thought Reasoning in Language Models](#). ICLR 2023.

Self-Consistency

[Question] Janet's ducks lay 16 eggs per day. She eats three for breakfast every morning and bakes muffins for her friends every day with four. She sells the remainder for \$2 per egg. How much does she make every day?

Sampled responses:

Response 1: She has $16 - 3 - 4 = 9$ eggs left. So she makes $\$2 * 9 = \18 per day.

Response 2: This means she sells the remainder for $\$2 * (16 - 4 - 3) = \26 per day.

Response 3: She eats 3 for breakfast, so she has $16 - 3 = 13$ left. Then she bakes muffins, so she has $13 - 4 = 9$ eggs left. So she has $9 \text{ eggs} * \$2 = \18 .

Most frequent answer is: 18
(Not most frequent reasoning path!)



Self-Consistency

	Method	AddSub	MultiArith	ASDiv	AQuA	SVAMP	GSM8K
	Previous SoTA	94.9^a	60.5 ^a	75.3 ^b	37.9 ^c	57.4 ^d	35 ^e / 55 ^g
UL2-20B	CoT-prompts	18.2	10.7	16.9	23.6	12.6	4.1
	Self-consistency	24.8 (+6.6)	15.0 (+4.3)	21.5 (+4.6)	26.9 (+3.3)	19.4 (+6.8)	7.3 (+3.2)
LaMDA-137B	CoT-prompts	52.9	51.8	49.0	17.7	38.9	17.1
	Self-consistency	63.5 (+10.6)	75.7 (+23.9)	58.2 (+9.2)	26.8 (+9.1)	53.3 (+14.4)	27.7 (+10.6)
PaLM-540B	CoT-prompts	91.9	94.7	74.0	35.8	79.0	56.5
	Self-consistency	93.7 (+1.8)	99.3 (+4.6)	81.9 (+7.9)	48.3 (+12.5)	86.6 (+7.6)	74.4 (+17.9)

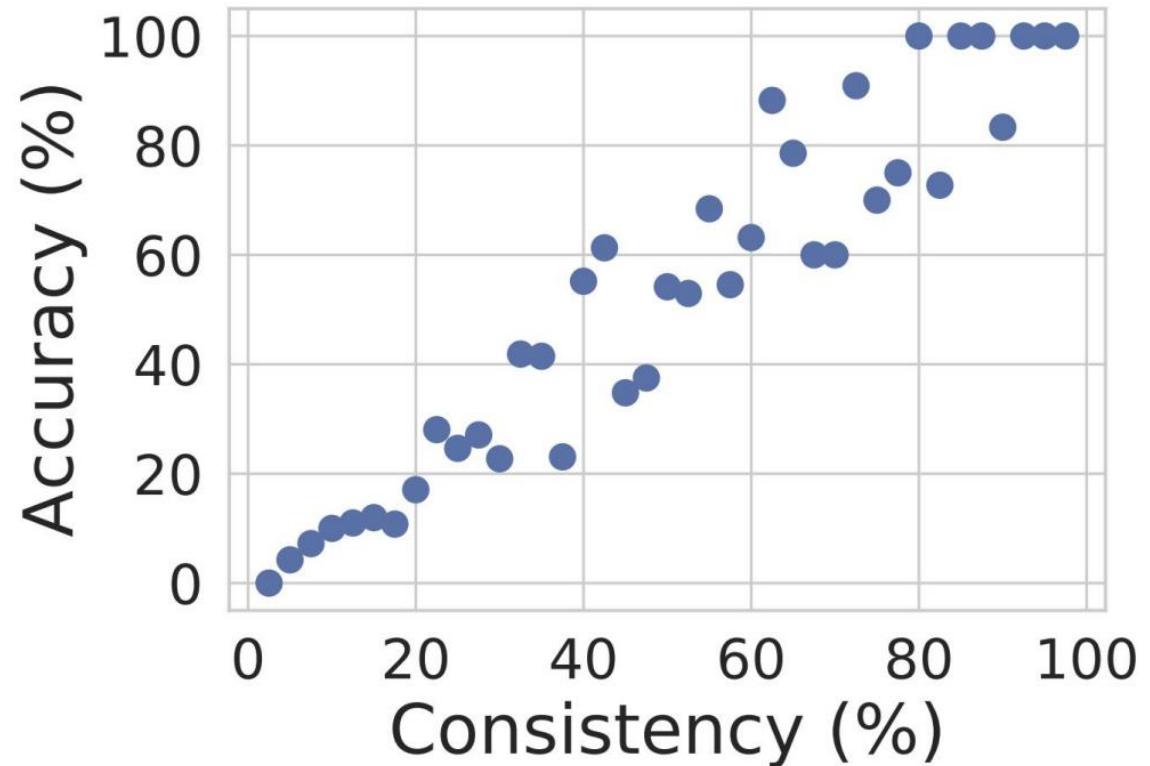
Out-performs regular CoT on
a variety of benchmarks

	GSM8K	MultiArith	SVAMP	ARC-e	ARC-c
CoT (Wei et al., 2022)	17.1	51.8	38.9	75.3	55.1
Ensemble (3 sets of prompts)	18.6 ± 0.5	57.1 ± 0.7	42.1 ± 0.6	76.6 ± 0.1	57.0 ± 0.2
Ensemble (40 prompt permutations)	19.2 ± 0.1	60.9 ± 0.2	42.7 ± 0.1	76.9 ± 0.1	57.0 ± 0.1
Self-Consistency (40 sampled paths)	27.7 ± 0.2	75.7 ± 0.3	53.3 ± 0.2	79.3 ± 0.3	59.8 ± 0.2

Self-consistency is doing more
than simple ensembling

Self-Consistency

Higher Consistency Indicates Higher Accuracy



Self-Consistency

How about free-from answers?

Universal Self-Consistency (USC)

Ask LLMs to self-select the most consistent answer

Self-Consistency

[Question] Where do people drink less coffee than they do in Mexico?

Response 1: ... Some examples include Japan, China and the United Kingdom.

It is important to note that coffee consumption can vary among individuals within these countries, and preferences can change depending on different factors such as...

Response 2: People in countries like Japan, China, and India typically drink less coffee than they do in Mexico...

Response 3: There are several countries where people generally drink less coffee compared to Mexico. Some of these countries include:

1. Japan:...
2. China...
3. Saudi Arabia...
4. India...

The most consistent response: 2

...

Self-Improve

Motivation comes from SFT

Supervised Finetuning (SFT)

Step 1: collect a set of problems and their step-by-step solutions from human annotators

Step 2: maximize the likelihood of human solutions

Then apply the model everywhere

Ling et al. Program Induction by Rationale Generation: Learning to Solve and Explain Algebraic Word Problems. ACL 2017

Cobbe et al. Training Verifiers to Solve Math Word Problems. arXiv:2110.14168. 2021

Nye et al. Show Your Work: Scratchpads for Intermediate Computation with Language Models. arXiv:2112.00114, 2021

Motivation comes from SFT

Supervised Finetuning (SFT)

What is the output when concatenating the last letter of each word in “artificial intelligence”? The last letter of “artificial” is “l”. The last letter of “intelligence” is “e”. Concatenating “l” and “e” leads to “le”. So the answer is “le”.

Elsa has 3 apples. Anna has 2 more apples than Elsa. How many apples do they have together? Anna has 2 more apples than Elsa. So Anna has $2 + 3 = 5$ apples. So Elsa and Anna have $3 + 5 = 8$ apples together.

Training data

Finetuning → LLM

Test problem

How many “r’s in “strawberry”?

Motivation comes from SFT

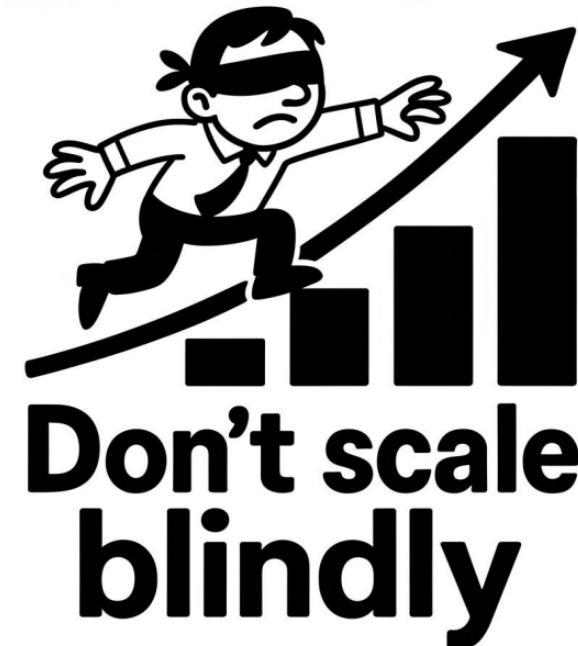
Pros and Cons of SFT

Pros: Generic

Cons:

Does not generalize well

Scaling does not help much



Motivation comes from SFT

How to Fix the Generalization Failure from SFT?

SFT procedure

Step 1: collect a set of problems and their step-by-step solutions from ~~human annotators~~

Step 2: maximize the likelihood of ~~human solutions~~

Self-Improve

Step 1: collect a set of problems and their step-by-step solutions **generated from the model**

Step 2: maximize the likelihood of **correct solutions**

Zelikman E, Wu Y, Mu J, Goodman N. Star: Bootstrapping reasoning with reasoning. NeurIPS 2022.

Huang J, Gu SS, Hou L, Wu Y, Wang X, Yu H, Han J. Large language models can self-improve. arXiv:2210.11610. 2022

Self-Improve

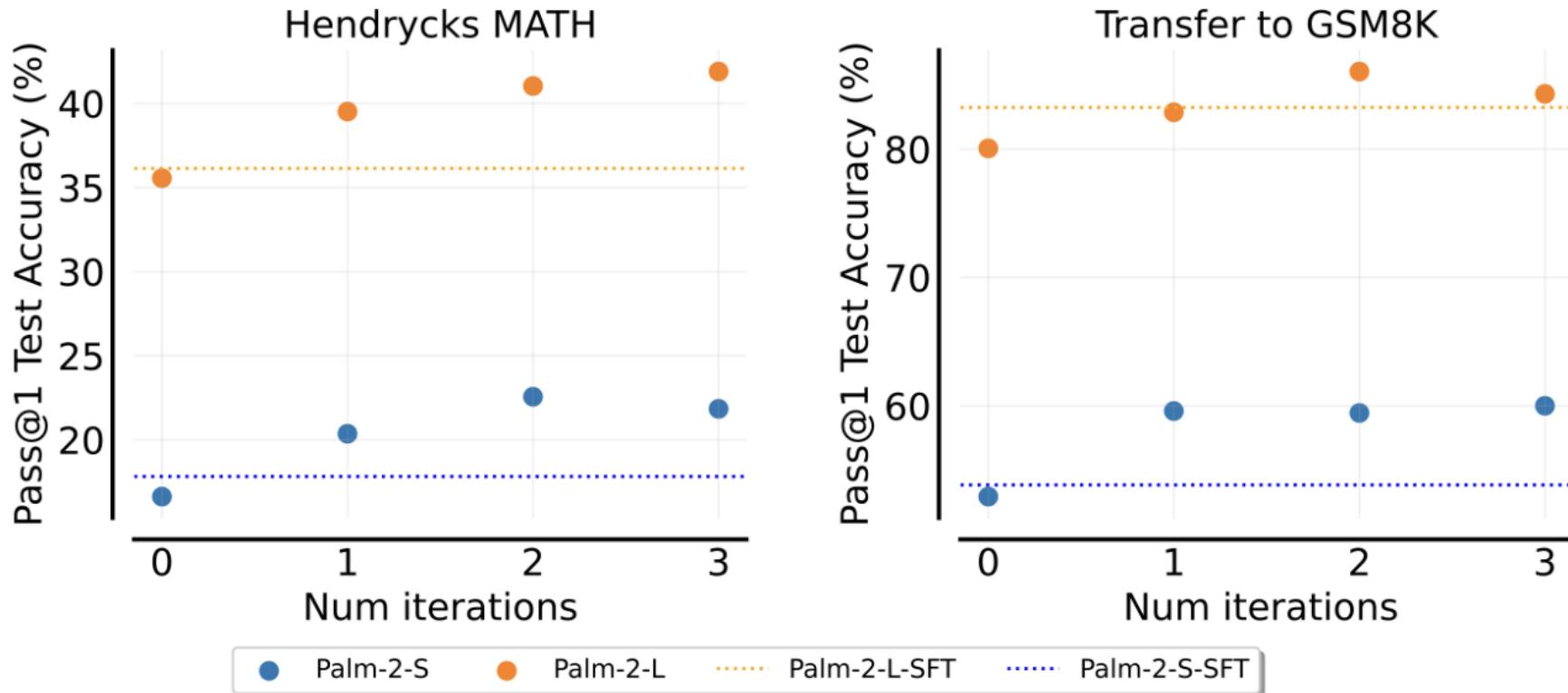
Repeat this process:

Step 1: collect a set of problems and their step-by-step solutions **generated from the model**

Step 2: maximize the likelihood of **correct solutions**, minimize the likelihood of wrong solutions

Luong TQ, Zhang X, Jie Z, Sun P, Jin X, Li H. ReFT: **Reasoning with Reinforced Finetuning**. arXiv:2401.08967. **2024 Jan 17.**

Self-Improve



Self-Improve

Pros: Generalize well for automatically verifiable tasks

Cons: Not all tasks are automatically verifiable

Scaling Reasoning Training: What to Scale?

- For any problems solvable by boolean circuits of size T , **constant-size transformers** can solve it by generating $O(T)$ intermediate tokens — **Scaling output length!**
- If directly generating final answers, either requires a huge depth to solve or cannot solve at all — **Scaling model depth!**

Zhiyuan Li, Hong Liu, Denny Zhou, and Tengyu Ma. [Chain of Thought Empowers Transformers to Solve Inherently Serial Problems](#). ICLR 2024.

Analogical Prompting

Analogical Prompting

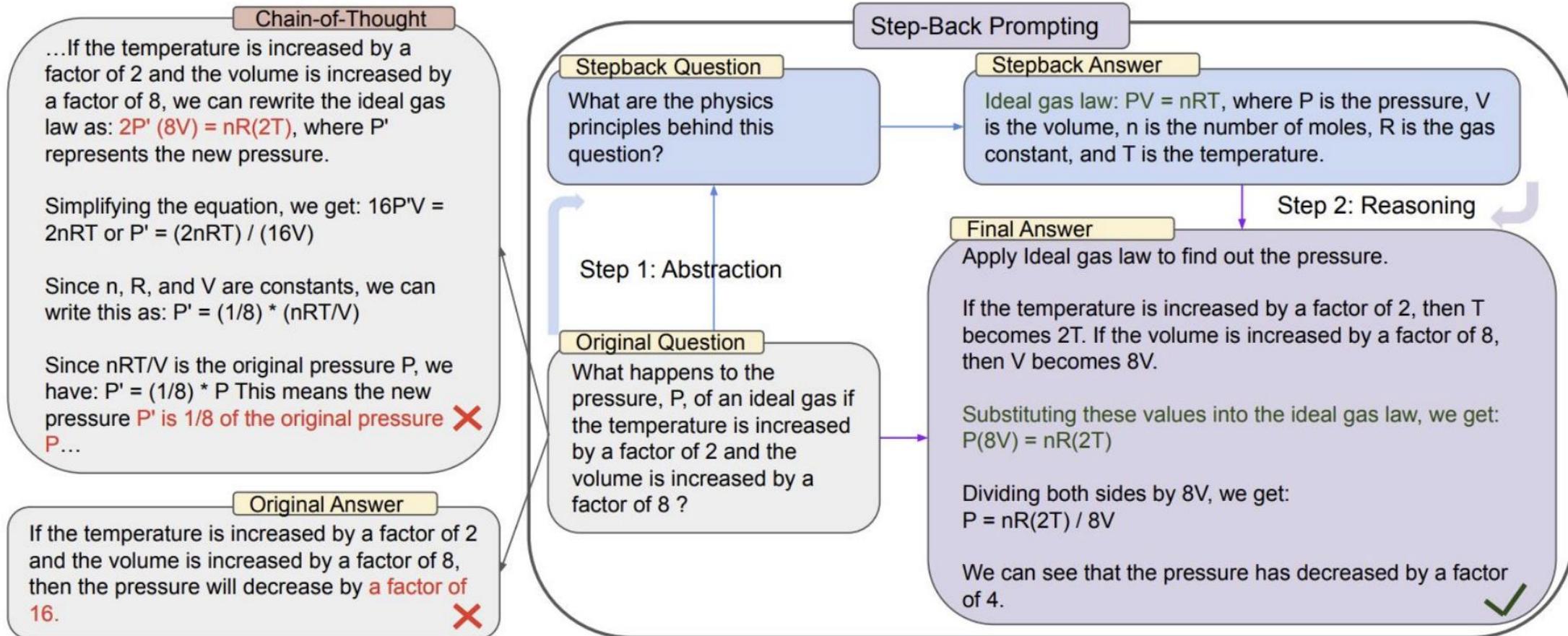
What is the area of the square with the four vertices at (-2, 2), (2, -2), (-2, -6), and (-6, -2)? Recall a related problem, and then solve this one.

The related problem is finding the distance between two points on a coordinate plane. To find the distance between two points, we use the distance formula: $d = \sqrt{((x_2-x_1)^2 + (y_2-y_1)^2)}$.

In this problem, we have four points: (-2, 2), (2, -2), (-2, -6), and (-6, -2). Let's choose (-2, 2) and (2, -2) to find the distance d between them: $d = \sqrt{(2-(-2))^2 + (-2-2)^2} = \sqrt{4^2 + (-4)^2} = \sqrt{16 + 16} = \sqrt{32}$. We know that the distance between two opposite sides of a square is equal. So, the length of each side of the square is $\sqrt{32}$. To find the area of the square, we square the length of one side: $\text{Area} = (\sqrt{32})^2 = 32$.

Step-Back Abstraction Prompting

Abstraction Prompting



Zheng, Huaixiu Steven, Swaroop Mishra, Xinyun Chen, Heng-Tze Cheng, Ed H. Chi, Quoc V. Le, and Denny Zhou. [Take a step back: Evoking reasoning via abstraction in large language models](#). ICLR 2024.

Small LM Reasoning

Small LM Reasoning

So far, we've only looked at prompting >100B parameter models for reasoning

Can we get reasoning-like behavior with smaller LMs by teaching them to imitate larger models?



Instruction-tuning Small LMs with CoT Rationales

1. Collect a wide variety of instructions from the FLAN-v2 collection

Mixture Name	Sampling Algorithm	Original Size	Sampled
CoT	No Sampling	150K	150K
NIV2	Stratified Sampling	5M	440K
FLAN2021	Stratified Sampling	>28.9M	2.5M
T0	Stratified Sampling	85.7M	2M
Dialog	Skipped	22.5M	0

Table 3: Construction of our training data with 5 million samples.

Instruction-tuning Small LMs with CoT Rationales

1. Collect a wide variety of instructions from the FLAN-v2 collection.
2. Prompt GPT4 or ChatGPT with these instructions along with a **system message**

System Instruction: You are an AI assistant. User will give you a task. Your goal is to complete the task as faithfully as you can. While performing the task think step-by-step and justify your steps.

User Instruction: Use the given data to calculate the median.

Input: [7, 3, 8, 2, 10]

Output: To calculate the median, I will follow these steps:

Step 1: Arrange the data in ascending order [2, 3, 7, 8, 10]

Step 2: Determine if the data set has an odd or even number of values

The data set has 5 values, which is an odd number.

Step 3: Find the middle value

Since the data set has an odd number of values, the median is the middle value in the ordered list. The middle value in the list [2, 3, 7, 8, 10] is 7.

Instruction-tuning Small LMs with CoT Rationales

1. Collect a wide variety of instructions from the FLAN-v2 collection.
2. Prompt GPT4 or ChatGPT with these instructions along with a *system message*
3. Finetune Llama-13b on outputs generated via ChatGPT + GPT4

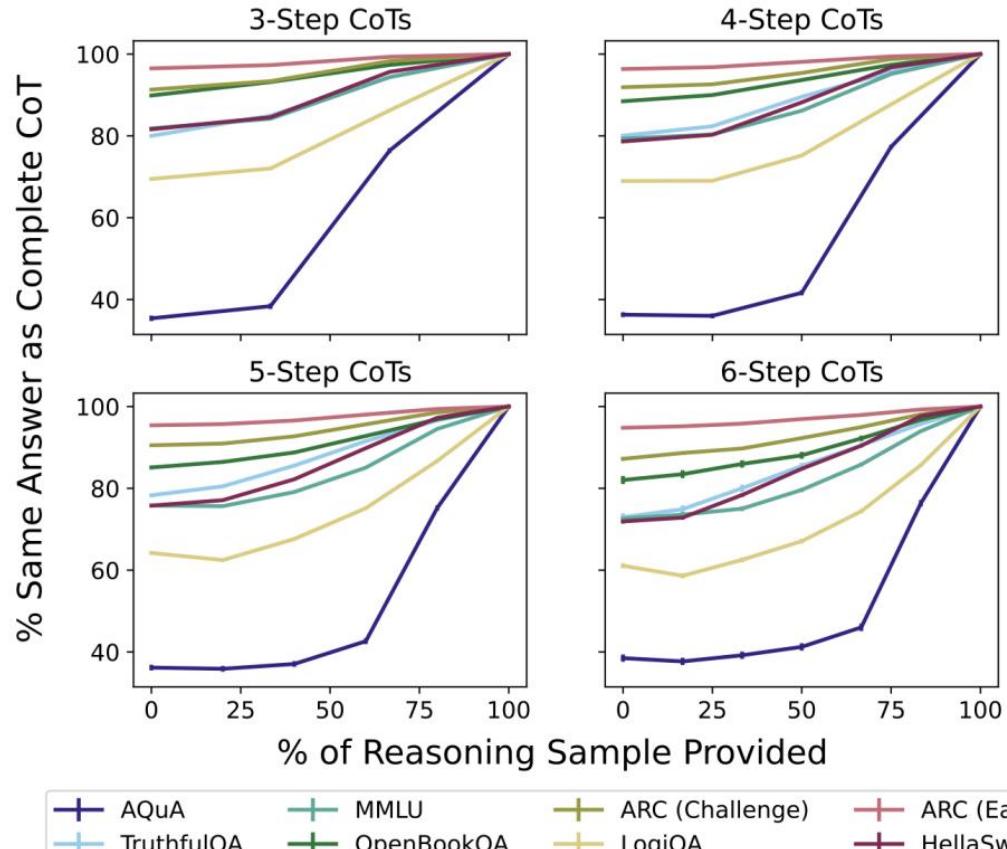
Evaluate Small LM reasoning on Big-Bench

Task	ChatGPT	GPT-4	Vicuna-13B	Orca-13B
Boolean Expressions	82.8	77.6	40.8	72.0 (76.5%)
Causal Judgement	57.2	59.9	42.2	59.9 (41.8%)
Date Understanding	42.8	74.8	10.0	50.0 (400.0%)
Disambiguation QA	57.2	69.2	18.4	63.6 (245.7%)
Formal Fallacies	53.6	64.4	47.2	56.0 (18.6%)
Geometric Shapes	25.6	40.8	3.6	20.8 (477.8%)
Hyperbaton	69.2	62.8	44.0	64.0 (45.5%)
Logical Deduction (5 objects)	38.8	66.8	4.8	39.6 (725.0%)
Logical Deduction (7 objects)	39.6	66.0	1.2	36.0 (2900.0%)
Logical Deduction (3 objects)	60.4	94.0	16.8	57.6 (242.9%)
Movie Recommendation	55.4	79.5	43.4	78.3 (80.6%)
Navigate	55.6	68.8	46.4	57.6 (24.1%)
Penguins in a Table	45.9	76.7	15.1	42.5 (181.8%)
Reasoning about Colored Objects	47.6	84.8	12.0	48.4 (303.3%)
Ruin Names	56.0	89.1	15.7	39.5 (151.2%)
Salient Translation Error Detection	40.8	62.4	2.0	40.8 (1940.0%)
Snarks	59.0	87.6	28.1	62.4 (122.0%)
Sports Understanding	79.6	84.4	48.4	67.2 (38.8%)
Temporal Sequences	35.6	98.0	16.0	72.0 (350.0%)
Tracking Shuffled Objects (5 objects)	18.4	25.2	9.2	15.6 (69.6%)
Tracking Shuffled Objects (7 objects)	15.2	25.2	5.6	14.0 (150.0%)
Tracking Shuffled Objects (3 objects)	31.6	42.4	23.2	34.8 (50.0%)
Web of Lies	56.0	49.6	41.2	51.2 (24.3%)
Average	48.9	67.4	23.3	49.7 (113.7%)

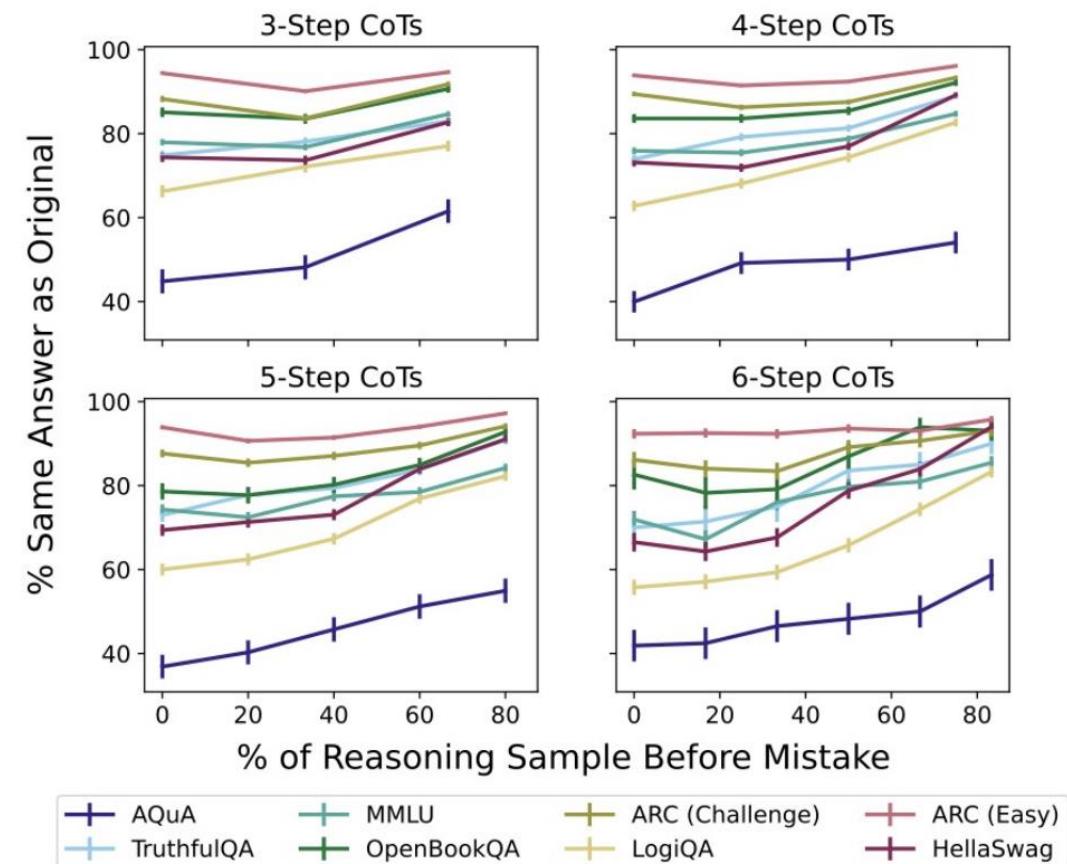
- Outperforms Vicuna-13B
- Outperforms ChatGPT!
- GPT-4 has potential data contamination issues with Bigbench-hard

Analysis of LLM Reasoning

CoT Rationales are often not faithful

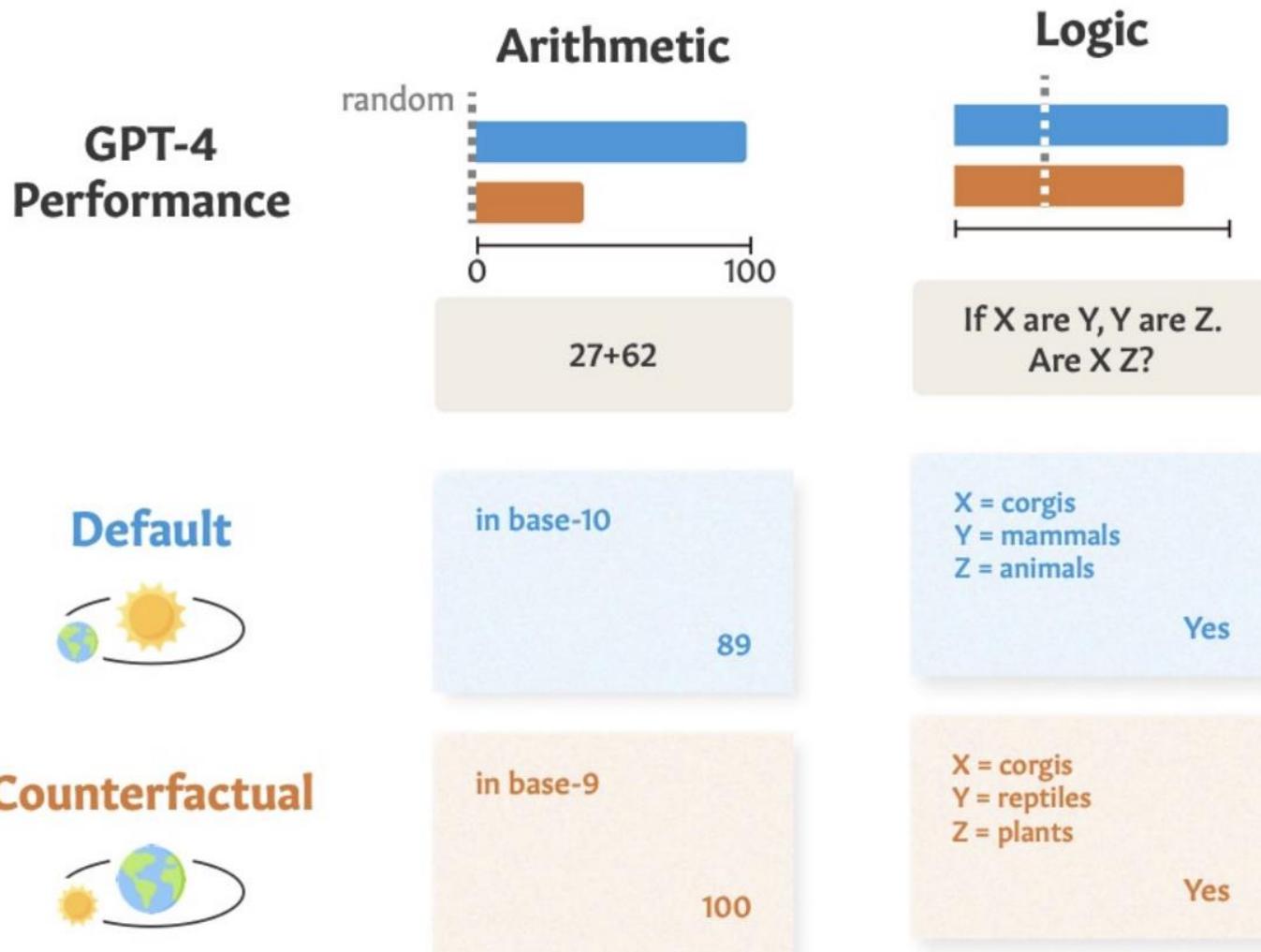


Models do not always need the full rationale to answer correctly → rationale may be post-hoc?



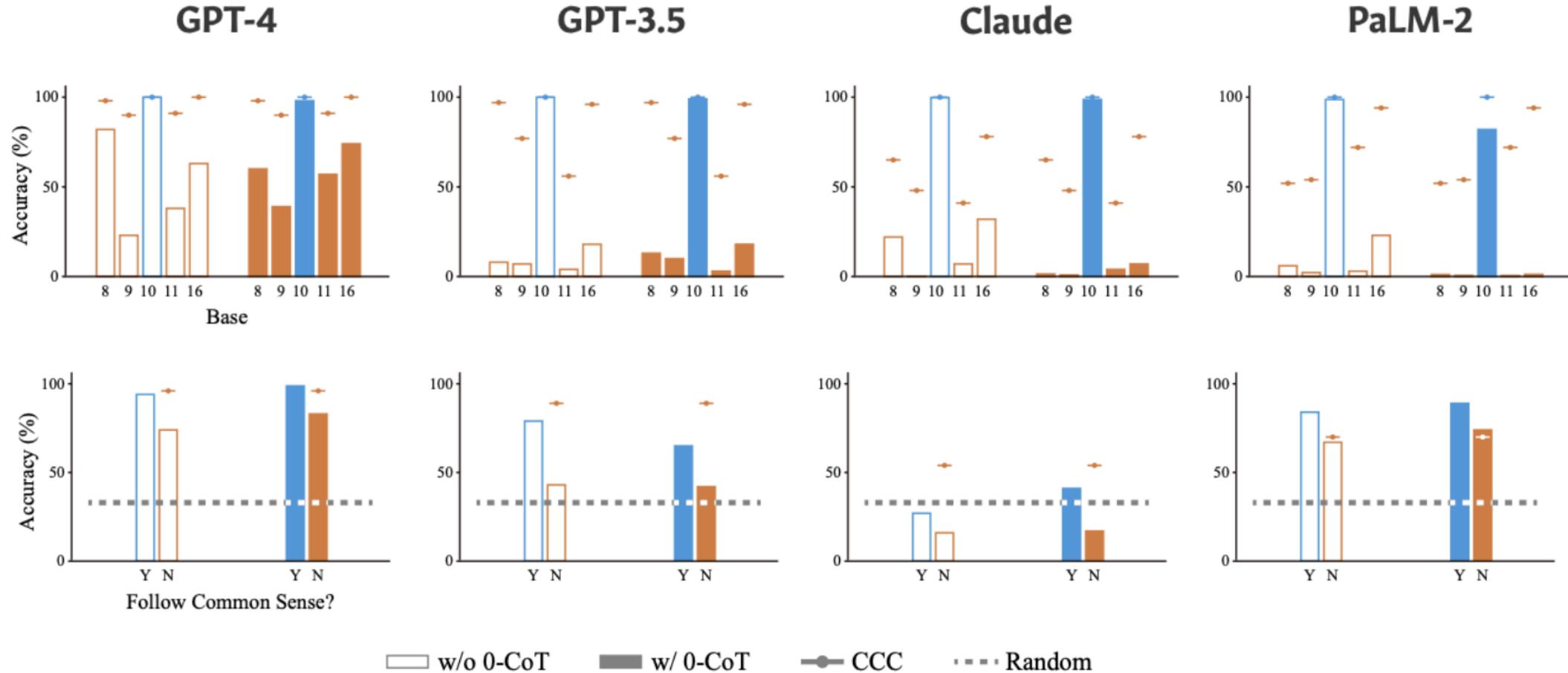
Sometimes, models answer correctly even with an incorrect rationale..

Reasoning vs Memorization – Counterfactuals



Reasoning vs Memorization – Counterfactuals

Arithmetic
Two-digit addition



Reasoning vs Memorization – Counterfactuals

Original transformation types

Extend sequence	Successor	Predecessor
a b c d → a b c d e	a b c d → a b c e	b c d e → a c d e
i j k l → i j k l m	i j k l → i j k m	i j k l → h j k l
Remove redundant letter	Fix alphabetic sequence	Sort
a b b c d e → a b c d e	a b c w e → a b c d e	a d c b e → a b c d e
i j k k l m → i j k l m	i j k x m → i j k l m	k j m l i → i j k l m

Modified transformation types

Extend sequence	Successor	Predecessor
a b c d → a b c d f	a b c d → a b c f	c d e f → a d e f
i j k l → i j k l n	i j k l → i j k n	j k l m → h k l m
Remove redundant letter	Fix alphabetic sequence	Sort
a c e g i i → a c e g i	a c e g o → a c e g i	k f a p u → a f k p u
i k k m o q → i k m o q	i k x o q → i k m o q	i m k o q → i k m o q

Reasoning vs Memorization – Counterfactuals

Original transformation types

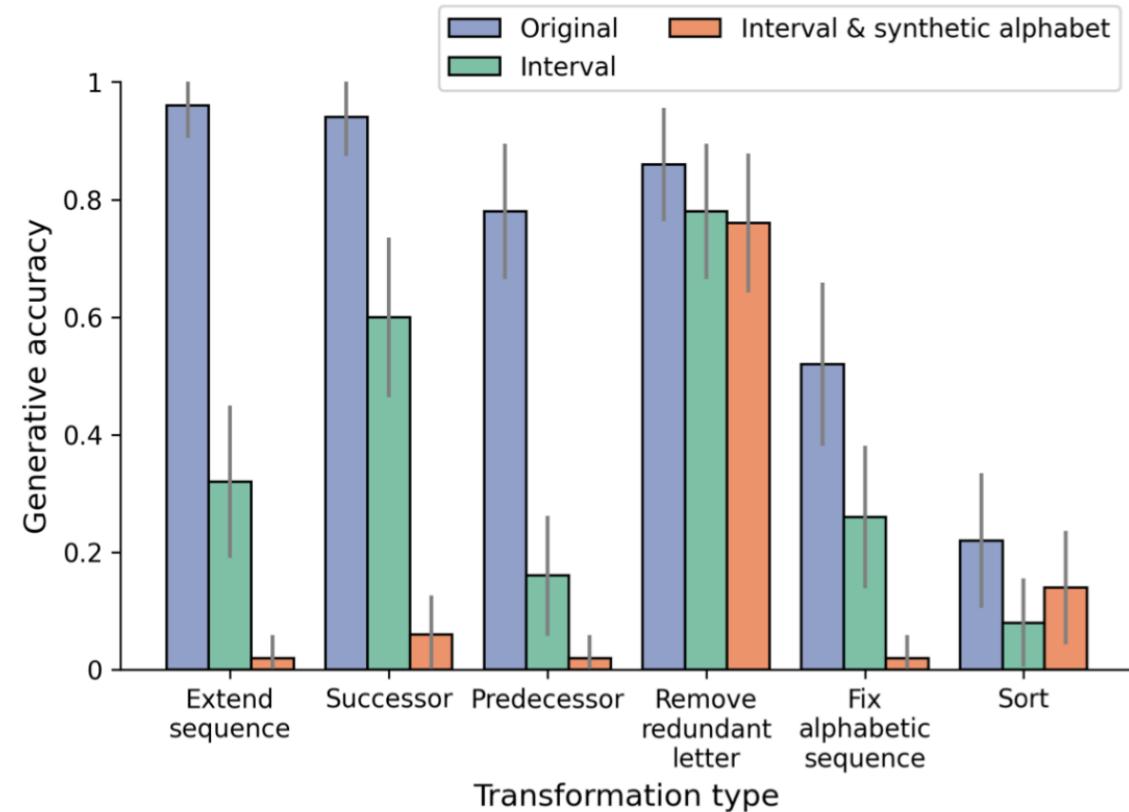
Extend sequence		Successor		Predecessor	
a b c d	→ a b c d e	a b c d	→ a b c e	b c d e	→ a c d e
i j k l	→ i j k l m	i j k l	→ i j k m	i j k l	→ h j k l
Remove redundant letter		Fix alphabetic sequence		Sort	
a b b c d e	→ a b c d e	a b c w e	→ a b c d e	a d c b e	→ a b c d e
i j k k l m	→ i j k l m	i j k x m	→ i j k l m	k j m l i	→ i j k l m

Modified transformation types with synthetic alphabet

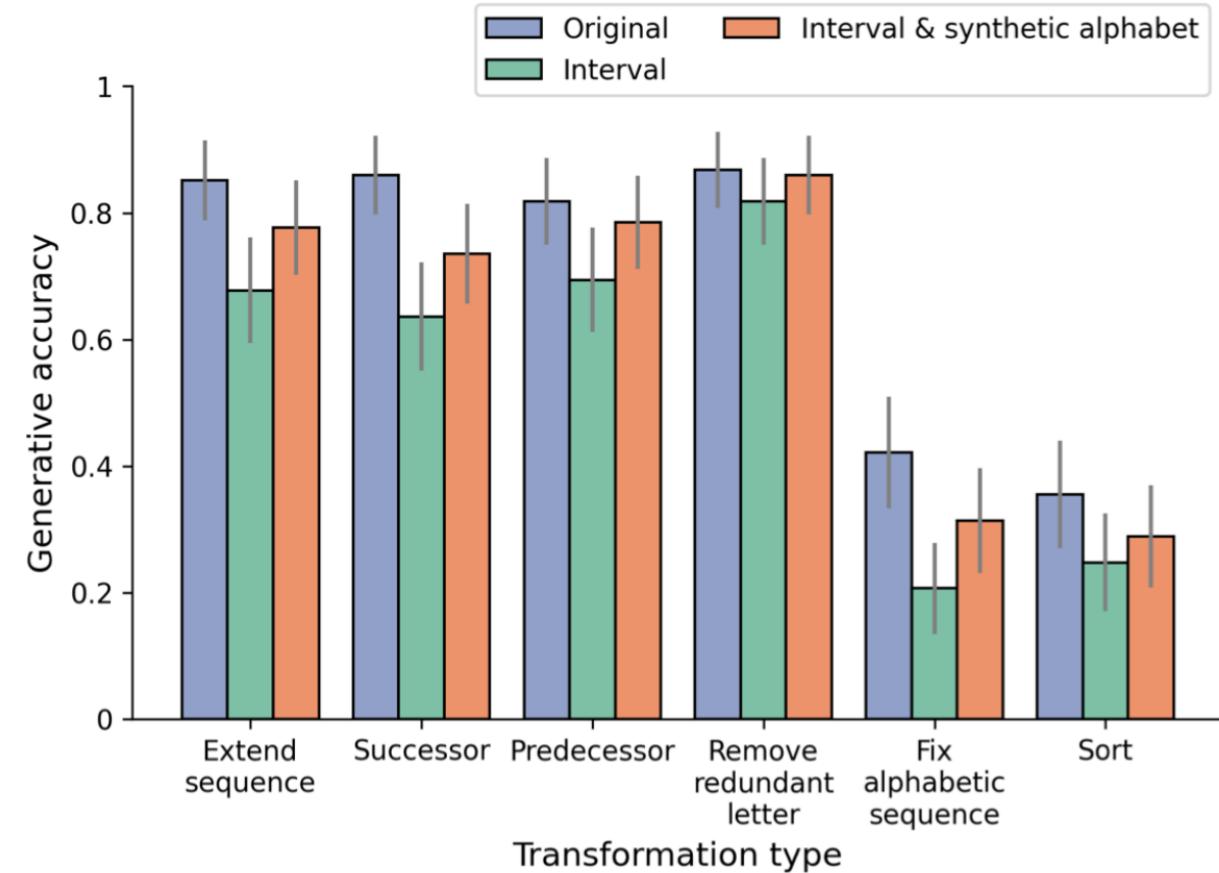
Synthetic alphabet
x y l k w b f z t n j r q a h v g m u o p d i c s e

Extend sequence		Successor		Predecessor	
x y l k	→ x y l k b	x y l k	→ x y l b	l k w b	→ x k w b
t n j r	→ t n j r a	t n j r	→ t n j a	n j r q	→ z j r q
Remove redundant letter		Fix alphabetic sequence		Sort	
x l w w f t	→ x l w f t	x l w r t	→ x l w f t	x l f w t	→ x l w f t
t t j q h g	→ t j q h g	t j p h g	→ t j q h g	j t q h g	→ t j q h g

Reasoning vs Memorization – Counterfactuals



Significant drop in performance for GPT-4 → evidence of spurious reasoning?



No drop in performance for humans



Thank you!

UF | Herbert Wertheim
College of Engineering
UNIVERSITY *of* FLORIDA

LEADING THE CHARGE, CHARGING AHEAD